Prepared for submission to JINST

Reconstructing the kinematics of semi-inclusive deep inelastic scattering at the EIC utilizing the hadronic final state and machine learning

Connor Pecar and Anselm Vossen

Duke University, Department of Physics

E-mail: connor.pecar@duke.edu

Abstract: Semi-inclusive deep inelastic scattering, $e(k) + P(p) \rightarrow e'(k') + h(P_h) + X$, is critical to mapping the three-dimensional momentum structure of the nucleon and is a major focus of the Electron-Ion Collider experimental project [1]. The Electron-Ion Collider will feature high luminosity, precise charged hadron tracking and particle identification, and polarized electron and ion beams providing the ability to study many semi-inclusive deep inelastic scattering observables across the wide kinematic space available at an e-A collider. In this study, full simulations of the EIC project detector are used to demonstrate methods utilizing the hadronic final state (particles produced from hadronization of struck quark) to determine the four-momentum of the exchanged virtual photon in deep-inelastic scattering events, around which the relevant angles and transverse momenta of semi-inclusive DIS are defined. These approaches to reconstruction include a full summation of the hadronic final state, extending on methods developed at HERA for inclusive-DIS kinematic reconstruction, as well as the first machine learning approach to semi-inclusive DIS kinematic reconstruction. The performance of these methods are compared to the reconstruction using only the scattered lepton, with improved kinematic resolution demonstrated in much of the kinematic space at the Electron-Ion Collider.

Contents

1 Introduction

Deep inelastic scattering (DIS) is used to probe the partonic substructure of nucleons and is a foundation of the experimental program at the future Electron-Ion Collider (EIC) [1]. In this electroweak process, a lepton scatters with a single parton inside the nucleon through the exchange of a virtual photon or W/Z-boson. The differential cross-section of inclusive-DIS, in which the scattered lepton is measured, is proportional to Parton Distribution Functions (PDFs), giving the probability of finding a parton with fractional momentum x. The point-like lepton probe makes this a clean process, with inclusive-DIS, $e(k) + P(p) \to e'(k') + X$, kinematics defined as

$$\begin{aligned} q = k - k',\ Q^2 = -q^2, \\ x = \frac{Q^2}{2P.q},\ y = \frac{P.q}{P.k}, \end{aligned} \tag{1.1}$$

where q is the four-momentum of the exchanged virtual photon. Q^2 is interpreted as the resolution of the probe, and x is the fraction of the longitudinal momentum of the nucleon carried by the struck parton. y can be interpreted as the "inelasticity" of the hard scattering. With $s = (P + k)^2$, these quantities can be related by $Q^2 = xys$.

Final state hadrons resulting from the hadronization of the struck parton can be measured in addition to the scattered lepton, $e(k) + P(p) \to e'(k') + h(P_h) + X$, known as semi-inclusive DIS (SIDIS). The cross-section is then proportional to combinations of PDFs with fragmentation functions (FFs) [2], with FFs describing the probability of a struck parton forming a specific hadron [3]. The additional scale provided by the transverse momentum of the hadron with respect to the virtual photon provides sensitivity to the transverse momentum dependence of PDFs and FFs [2], making semi-inclusive measurements a valuable tool for probing the three-dimensional

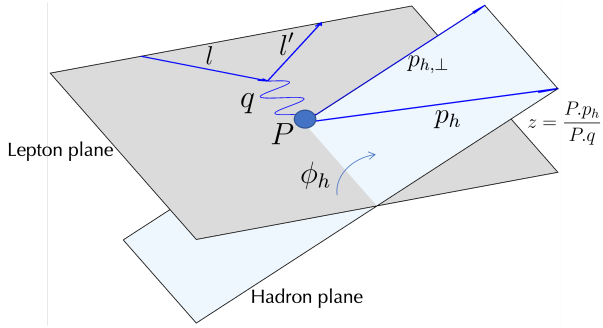

Figure 1. SIDIS kinematics in nucleon target rest frame.

structure of nucleons in momentum space. Additionally, measurements of the semi-inclusive production of different hadrons such as pions and kaons can give access to the flavor makeup of sea quarks in the nucleon [4]. SIDIS spin-asymmetry measurements at experiments such as HERMES [5], COMPASS [6], and CLAS [7] have provided critical experimental information on the spin structure of the nucleon, and SIDIS measurements have been historically carried out at the first e-p collider HERA [8, 9].

1.1 Semi-inclusive DIS kinematics

The mapping of TMD-PDFs and TMD-FFs through SIDIS measurements requires a precise determination of both the inclusive DIS kinematics as well as kinematic variables related to the selected final state hadron. These kinematics are defined in the rest frame of the nucleon target [10] (figure 1). In single hadron SIDIS, the transverse momentum scale of the process is defined as the momentum of the hadron transverse to the virtual photon axis, $p_{h,\perp}$, which enters into both the PDFs and FFs. The relevant azimuthal angle of the final state hadron is taken between the hadron plane, spanning $\vec{q}$ and $\vec{p_h}$, and the lepton plane, spanning $\vec{q}$ and $\vec{k}$. These kinematic quantities are depicted in figure 1, with the convnetions defined in reference [10]. Finally, an additional scaling variable z enters into the fragmentation functions, defined as $z = P.P_h/P.q$, the fraction of the energy of the struck quark carried by the measured hadron.

2 SIDIS reconstruction methods

As the SIDIS kinematic quantities are defined around the four-momentum of the virtual photon, q must be well-constrained throughout the entire kinematic space available at the EIC. In past SIDIS studies carried out at fixed-target experiments and HERA, q is determined using only the scattered lepton, $q = k - k'$. Studies of SIDIS kinematic reconstruction in EIC simulation for the EIC yellow report [1] found the kinematic resolution achieved using this method to be reliable at large values of y, but with performance degrading rapidly for values of $y < 0.1$, as y is proportional to the energy loss of the lepton.

The low-y kinematic space is critical for the measurement of a variety of SIDIS observables at the EIC [11]. Due to depolarization factors which are dependent on y, the low-y region provides the greatest sensitivity to observables with an unpolarized lepton beam, such as target spin asymmetries [11]. Additionally, the large-x and low-Q^2 region accessible at low-y provides critical overlap with the kinematic space covered by fixed-target SIDIS measurements.

In this study, we demonstrate two methods which utilize information from the hadronic final state (all final state particles produced from the struck parton) in addition to the scattered electron to determine the four-momentum of the virtual photon, and thus improve the resolution of SIDIS kinematics at low-y.

2.1 Hadronic final state methods

Due to conservation of momentum and energy, the hadronic final state (HFS) of the DIS interaction, or all particles in the hadronization jet resulting from the struck quark, will also contain enough information to fully constrain q. Methods to reconstruct the inclusive DIS kinematics x, Q^2, and y using the HFS were developed for the analysis of HERA data [12]. These included methods utilizing only the HFS which are required for the analysis of charged-current DIS, such as the Jacquet-Blondel method, or hybrid HFS-scattered electron methods, such as the double angle method [12]. HERA simulation studies found these additional methods to provide advantages in precision in some regions of the kinematic space, and in some cases benefits when considering effects of QED radiation, such as the Σ-method [13]. In this section, we introduce two methods further utilizing the HFS to improve the reconstruction of SIDIS kinematics, making these methods the first use of the full HFS for kinematic reconstruction in SIDIS.

2.1.1 Constraining q with transverse recoil, y, and Q^2

As defined in equation 1.1, the virtual photon four-momentum is used to define the inclusive DIS variables x, y, Q^2 in their Lorentz invariant form. Given that the alternative methods of reconstructing DIS variables developed at HERA [12] outperform the electron method in some regions of the kinematic space, the determination of the inclusive-DIS variables can be leveraged to place constraints on the reconstruction of the virtual photon four-momentum q.

Using the inclusive DIS variables as computed from a hybrid HFS-electron method, such as the double-angle method [12], and with q_x, q_y taken from the transverse recoil of either the scattered electron or the hadronic final state, the remaining two components of q can be constrained from the following system of equations:

$$\left.\begin{aligned} (q_x, q_y) &= \text{HFS } \vec{p_T} \mid\mid \text{electron } \vec{p_T}, \\ Q^2 &= -q^2, \\ y &= \frac{P.q}{P.k} \end{aligned}\right\} (q_x, q_y, q_z, q_t). \tag{2.1}$$

In a kinematic region in which the hybrid DIS method has better overall accuracy than the electron method, this can result in an improved determination of q_z, q_E, as demonstrated by us for the first time in simulation studies for the EIC yellow report [1].

2.1.2 Machine learning approach

To produce a further optimized combination of the information from the scattered electron and HFS for the reconstruction of q, a machine learning (ML) approach was developed using the full final state information of simulated DIS events. Studies have been conducted on the use of machine learning for directly reconstructing inclusive DIS kinematics in the context of e-p colliders, using simulated HERA data to demonstrate the ability of machine learning approaches to outperform traditional reconstruction methods as well as minimize the impact of radiative corrections [14, 15]. Additionally, recent work utilizing Bayesian Neural Networks [16] and kinematic fitting based on Bayesian inference [17] have demonstrated methods which both outperform traditional methods of reconstruction inclusive DIS kinematics and providing event-level uncertainty quantification. Our application of machine learning to reconstructing q was first demonstrated on the full simulation developed for the ATHENA detector proposal [18, 19].

In this study, a neural network is trained to determine the four-momentum of the virtual photon q, rather than directly reconstructing the SIDIS kinematics. The network architecture utilized for this application is Particle Flow Networks (PFN) [20]. This architecture takes as input the features of an unordered and variable size set of particles. The features p_i of each particle are passed individually through a first set of fully connected layers Φ, after which they are summed over to form a latent space which is passed to a final set of fully connected layers F,

$$\mathrm{PFN} = F(\sum_{i=1}^{M} \Phi_{(}p_i))$$

[20]. Global features of the event, not associated with any particular particle, can be concatenated with the latent space variables formed after summing over the outputs of Φ.

The values of p_x, p_y, p_z, E from the set of all HFS particles was input to the layers Φ, with the size of the HFS varying. The global features of the event were taken as the components of the virtual photon four momentum as determined using the electron method, $q = k - k'$, as well as the value of $-log_{10}(x)$ and $log_{10}(Q^2)$as reconstructed from the electron, double-angle, and Σ methods (methods defined in [12]).

The Particle Flow Network architecture was used through the Energyflow python package [20]. Each hidden layer was followed by the 'relu' activation function, with linear final output. A mean-squared-error loss function was chosen and two models were trained: one to reconstruct the lab-frame q_x and q_y, and one to reconstruct q_z, q_E. The number of hidden layers and units per layer for each network was optimized using a grid search by selecting the parameters resulting in the minimal validation loss after 50 epochs. The network trained to reconstruct q_x and q_y had 3 Φ layers of 350 units and 3 F layers of 350 units. The network trained to reconstruct q_z and q_E had 3 Φ layers of 350 units and 3 F layers of 200 units.

3 ePIC Simulation Dataset

In this analysis, the above described reconstruction methods are demonstrated again on the further developed full simulation of the ePIC detector. The ePIC detector will be the first EIC detector [21], located at interaction point-6 with data taking expected to begin in the early 2030s. ePIC is planned to utilize a 1.7T solenoidal magnet and Si MAPS tracking system, providing precise momentum reconstruction of charged particles with a minimum transverse momentum of 0.1 GeV/c. Precise electromagnetic calorimetry will be carried out with a PbWO4 backwards ECal, and an imaging barrel ECal. Wide particle ID coverage will be provided by various DIRC and RICH detectors in each region of the detector.

The dataset used in this analysis is the July 2023 full simulation of the ePIC detector. The July 2023 ePIC simulation campaign contains full simulation and reconstruction of the ePIC calorimetry and tracking systems. The detector geometry [22] is implemented in dd4hep [23] with reconstruction carried out with EICrecon [24].

The events used for this analyis were neutral-current DIS events generated using pythia-8 [25] including beam momentum smearing and a beam crossing-angle of 25 mrad. The energy configuration used in this study was an electron beam energy of 18GeV and proton beam energy of 275GeV. QED radiative corrections are not applied in this dataset. Standard event-level DIS cuts were placed of $Q^2 > 1\text{GeV}^2$ and HFS invariant mass $W > 3$ GeV, and it was required that more than one particle be reconstructed in the HFS.

The reconstructed particle information used for this analysis of kinematic reconstruction was primarily sourced from the ePIC tracking systems. The scattered electron was determined by taking the matching reconstructed charged track to the MC-truth scattered electron ID, as a more realistic DIS-electron finder is still under development. The HFS four-momenta were taken from all other reconstructed charged tracks with a minimum lab frame transverse momentum of 0.1 GeV/c, with any calorimeter clusters not associated with a track taken as additional neutral HFS particles.

The two PFN models were each trained on 1.6 million events with 1 million events set aside for final validation.

4 Kinematic reconstruction results

On the 1 million event validation dataset, SIDIS kinematics for positive pions were computed using q as reconstructed from the electron, hybrid HFS-electron, and the ML method. The HFS-electron hybrid method in this study used the transverse recoil as measured from the scattered electron and the double-angle method to compute y and Q^2. Typical cuts on the true SIDIS kinematics were placed requiring $p_{h,\perp} > 0.1$ GeV/c and $z > 0.2$.

In figure 2, the mean and RMS of the error or relative error in $z, p_{h,\perp}, \phi_h$ is plotted for values of y_{true}, which best captures the behavior of the electron reconstruction method. At large-y, the electron energy loss is larger and the electron method performs well, as expected. The PFN reconstruction at large-y succesfully matches the performance of the electron method for reconstruction of z and $p_{h,\perp}$, with slightly worse resolution of ϕ_h.

For decreasing y, the electron method performance degrades rapidly for all SIDIS kinematic quantities. In contrast, the methods utilizing the HFS maintain a narrow distribution better centered around zero even at very low y-values of 0.005. The ML method in-particular outperforms or equals the electron method for z and $p_{h,\perp}$ in all y-bins, and for ϕ_h in all but the highest y-bin. The decrease in performance as a function of y is significantly more gradual with the ML approach. As the network is directly given information on q as reconstructed from the electron method, it appears that the network is able to learn additional corrections to the electron method through correlations with the full hadronic final state.

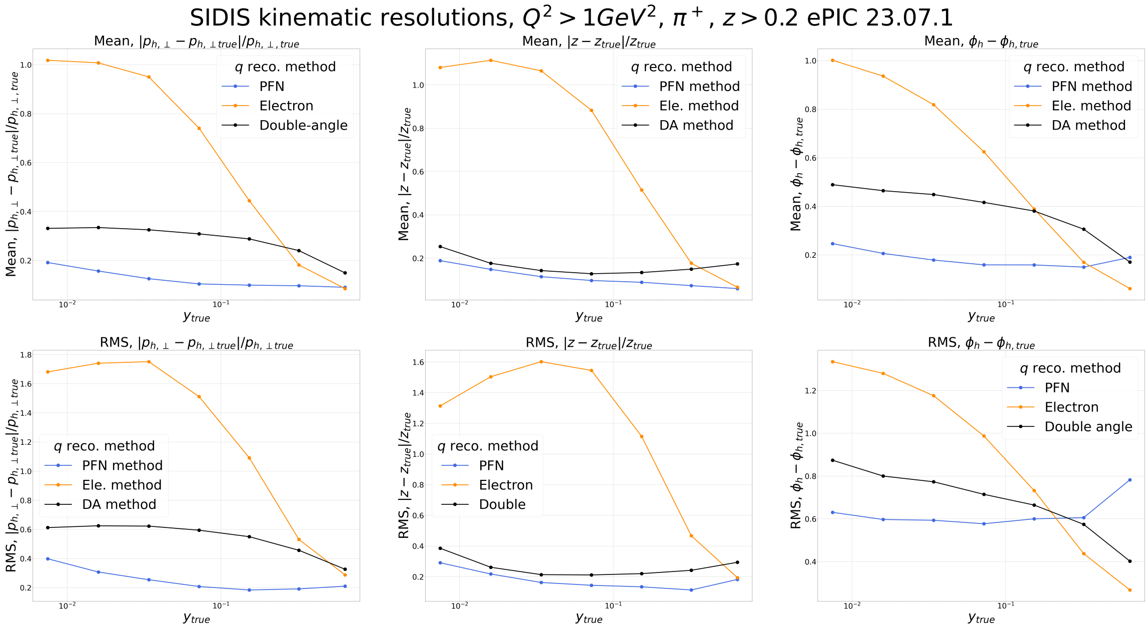

Figure 2. Absolute relative error in $p_{h,\perp}$ and z and error in ϕ_h with validation dataset in bins of y_{true} ranging from y=0.005 to y=1. A cut is placed requiring the absolute relative error be less than 1000% to understand the core of the distribution, which removes about 0.5% of events for the electron method. The results using each of the electron method, DA+HFS method, and PFN trained as described are shown. Top row displays the mean of these distributions as a function of y, and bottom row displays the RMS of these distributions.

5 Conclusion

Using the July 2023 full simulation of the ePIC detector, we have demonstrated two methods which improve on the reconstruction of semi-inclusive DIS kinematics by combining information from the hadronic final state and the scattered electron. Both HFS methods demonstrated maintain stable reconstruction at low-y where the electron method fails, and the use of Particle Flow Networks to reconstruct the virtual photon four-momentum outperforms the electron method for almost all values of y. This first use of a machine learning approach to the reconstruction of SIDIS kinematics will provide a better understanding of the ultimate kinematic resolution that will be achieved with the ePIC detector. As the ePIC simulation is developed further, the impacts of QED radiative corrections and a realistic DIS-electron finder must be investigated.

Acknowledgments

We would like to thank the ePIC collaboration for the use of their simulation campaign in this study.

References

[1] R. Abdul Khalek, A. Accardi, J. Adam, D. Adamiak, W. Akers, M. Albaladejo et al., *Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report*, Nuclear

Physics A **1026** (2022) 122447.

[2] A. Bacchetta, M. Diehl, K. Goeke, A. Metz, P.J. Mulders and M. Schlegel, *Semi-inclusive deep inelastic scattering at small transverse momentum*, *Journal of High Energy Physics* **2007** (2007) 093.

[3] A. Metz and A. Vossen, *Parton fragmentation functions*, *Progress in Particle and Nuclear Physics* **91** (2016) 136.

[4] E.C. Aschenauer, I. Borsa, R. Sassot and C. Van Hulse, *Semi-inclusive deep-inelastic scattering, parton distributions, and fragmentation functions at a future electron-ion collider*, *Physical Review D* **99** (2019) 094004.

[5] A. Airapetian, N. Akopov, Z. Akopov, M. Amarian, A. Andrus, E.C. Aschenauer et al., *Single-Spin Asymmetries in Semi-Inclusive Deep-Inelastic Scattering on a Transversely Polarized Hydrogen Target*, *Physical Review Letters* **94** (2005) 012002.

[6] V.Y. Alexakhin, Y. Alexandrov, G.D. Alexeev, A. Amoroso, B. Badełek, F. Balestra et al., *First Measurement of the Transverse Spin Asymmetries of the Deuteron in Semi-inclusive Deep Inelastic Scattering*, *Physical Review Letters* **94** (2005) 202002.

[7] H. Avakian, P. Bosted, V.D. Burkert, L. Elouadrhiri, K.P. Adhikari, M. Aghasyan et al., *Measurement of Single- and Double-Spin Asymmetries in Deep Inelastic Pion Electroproduction with a Longitudinally Polarized Target*, *Physical Review Letters* **105** (2010) 262002.

[8] C. Alexa, V. Andreev, A. Baghdasaryan, S. Baghdasaryan, W. Bartel, K. Begzsuren et al., *Measurement of charged particle spectra in deep-inelastic ep scattering at HERA*, *The European Physical Journal C* **73** (2013) 2406.

[9] M. Derrick, D. Krakauer, S. Magill, D. Mikunas, B. Musgrave, J. Repond et al., *Inclusive charged particle distributions in deep inelastic scattering events at HERA*, *Zeitschrift für Physik C Particles and Fields* **70** (1996) 1.

[10] A. Bacchetta, U. D'Alesio, M. Diehl and C.A. Miller, *Single-spin asymmetries: The Trento conventions*, *Physical Review D* **70** (2004) 117504.

[11] V.D. Burkert, L. Elouadrhiri, A. Afanasev, J. Arrington, M. Contalbrigo, W. Cosyn et al., *Precision studies of QCD in the low energy domain of the EIC*, *Progress in Particle and Nuclear Physics* **131** (2023) 104032.

[12] J. Blümlein, *The theory of deeply inelastic scattering*, *Progress in Particle and Nuclear Physics* **69** (2013) 28.

[13] U. Bassler and G. Bernardi, *On the kinematic reconstruction of deep inelastic scattering at HERA*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **361** (1995) 197.

[14] M. Diefenthaler, A. Farhat, A. Verbytskyi and Y. Xu, *Deeply learning deep inelastic scattering kinematics*, *The European Physical Journal C* **82** (2022) 1064.

[15] M. Arratia, D. Britzger, O. Long and B. Nachman, *Reconstructing the kinematics of deep inelastic scattering with deep learning*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1025** (2022) 166164.

[16] C. Fanelli and J. Giroux, *ELUQuant: event-level uncertainty quantification in deep inelastic scattering*, *Mach. Learn. Sci. Tech.* **5** (2024) 015017.

[17] R. Aggarwal and A. Caldwell, *Kinematic fitting of neutral current events in deep inelastic ep collisions.*, *Journal of Instrumentation* **17** (2022) P09035.

[18] C. Pecar and A. Vossen, *Reconstruction of event kinematics in semi-inclusive deep-inelastic scattering using the hadronic final state and Machine Learning*, 2209.14489.

[19] J. Adam, L. Adamczyk, N. Agrawal, C. Aidala, W. Akers, M. Alekseev et al., *ATHENA detector proposal — a totally hermetic electron nucleus apparatus proposed for IP6 at the Electron-Ion Collider*, *Journal of Instrumentation* **17** (2022) P10019.

[20] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy flow networks: deep sets for particle jets*, *Journal of High Energy Physics* **2019** (2019) 121.

[21] “ePIC Collaboration wiki.” https://wiki.bnl.gov/EPIC.

[22] “ePIC detector geometry.” https://github.com/eic/epic.

[23] M. Frank, F. Gaede, M. Petric and A. Sailer, *Aidasoft/dd4hep*, Oct., 2018. 10.5281/zenodo.592244.

[24] “EICrecon github.” https://github.com/eic/EICrecon.

[25] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, Mar., 2022. 10.48550/arXiv.2203.11601.