Machine learning SIDIS reconstruction at the EIC

Connor Pecar and Anselm Vossen

Duke University

I. INTRODUCTION

Deep-inelastic scattering (DIS) of electrons off ions is at the forefront of experimental efforts to probe the internal structure of nucleons and nuclei and will be a primary focus of study at the Electron-Ion Collider. In semi-inclusive DIS, selected particles produced by the fragmentation of the struck quark are observed in co-incidence with the scattered electron, $e(k) + N(P) \rightarrow e(k') + h(p) + X$, resulting in observables which provide access to a convolution of parton distribution functions (PDFs), describing the momentum of partons within the nucleon, and fragmentation functions (FFs), describing the probability of producing a final state particle with some momentum from the struck quark in the factorization approach. [1]

The kinematic variables describing the DIS process, with center of mass energy squared $s = (P + k)^2$ can be defined in terms of the virtual photon four-momentum q as [2],

$$Q^2 = -q^2, \ y = \frac{P \cdot q}{P \cdot k}, \ x = \frac{Q^2}{sy} \tag{1}$$

II. SIDIS KINEMATIC RECONSTRUCTION

In semi-inclusive DIS, observables are extracted in the nucleon center of mass frame, with the SIDIS cross-section a function of the inclusive DIS variables as well as $(p_{h\perp}, z, \phi_h)$. The relevant transverse momentum is defined with respect to the virtual photon axis, and the single-hadron azimuthal angle ϕ_h is defined between the lepton scattering plane and hadron production plane (figure 1). z is defined as $z = \frac{p_h \cdot P}{q \cdot P}$ The calculation of
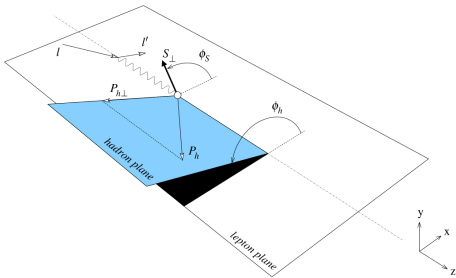

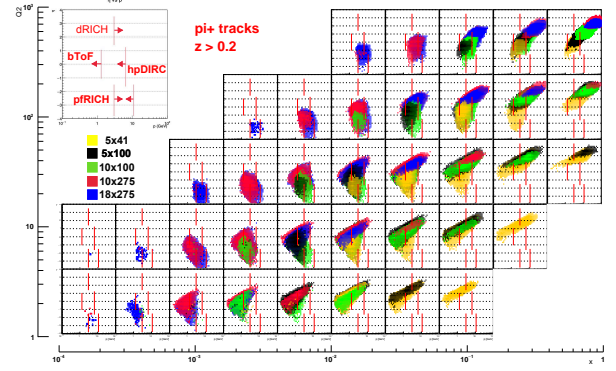FIG. 1: Definition of semi-inclusive kinematic variables in target rest frame, from [1].

FIG. 2: Momentum vs pseudorapidity in $x - Q^2$ bins for positive pions with $z > 0.2$ and requiring $y > 0.05$, generated with Pythia-8. Different histogram colors represent different $\sqrt{s}$ values. Red bands represent 3-σ PID coverage based on ATHENA proposal projections.

SIDIS kinematics therefore requires precise reconstruction of the four-momenta of the selected hadron and the exchanged virtual-photon.

A. Electron method

Extraction of SIDIS observables and multiplicities at the EIC presents a new challenge, as fully multi-dimensional SIDIS studies have so far only been carried out in lower energy fixed target experiments. In fixed target SIDIS studies, q has been determined using only the scattered electron, $q = k - k'$. However, studies done for the EIC yellow report and EIC detector proposals have found that a significant contribution to uncertainty in SIDIS kinematics is poor reconstruction of the virtual photon four-momentum when using only the scattered electron. In particular, the electron method fails in such regions of kinematic phase space at the EIC such as at low y ($y < 0.05$), where the energy loss of the electron is small and not well resolved. This is a significant issue for the study of TMD effects at e-p colliders, as at low-Q^2 and large-x spin-orbit correlations are expected to be most significant and higher twist effects are observable. Additionally, the low-y region will be critical for overlapping the phase space covered by the EIC and SIDIS studies carried out at other facilities such as Jefferson Lab.

B. Hadronic final state methods

Fast simulation studies for the EIC yellow report [3] and ATHENA (A Totally Hermetic Electron Nucleon Apparatus) proposal [4] have demonstrated that DIS reconstruction methods developed at past e-p colliders [2] can be used to improve the reconstruction of inclusive DIS variables at the EIC. The DIS reconstruction methods developed at HERA utilized combinations of measured quantities from the scattered electron and the hadronic final state (HFS). The use of the HFS allowed these additional methods, such as the double angle (DA) and Σ-methods [2], to improve inclusive DIS kinematic reconstruction for various regions of the HERA kinematic space, as well as to make the reconstruction robust with respect to QED radiative effects [5, 6]. For the studies planned at the EIC, methods utilizing the HFS must be extended to the reconstruction of SIDIS kinematics.

The authors of this contribution conducted first studies of SIDIS kinematic reconstruction for the EIC and demonstrated methods in which the hadronic final state can be used to improve the reconstruction of the virtual photon four momentum. This was carried out in the EIC yellow report and ATHENA proposal [3, 4] by first obtaining the transverse component of q from the recoil of the HFS transverse to the beamline through a sum of the momenta of HFS particles. Following the determination of this transverse recoil, the remaining two components of q can be constrained by the system of equations including q from the definitions of Q^2 and y,

$$q_x = \sum_{i}^{N_{HFS}} p_{x,i}, q_y = \sum_{i}^{N_{HFS}} p_{y,i} \qquad (2)$$

$$q_z, q_t \leftarrow \begin{cases} Q^2 = -(q_x^2 + q_y^2 + q_z^2 - q_t^2) \\ y = \frac{P_x q_x + P_y q_y + P_z q_z - P_t q_t}{P \cdot k} \end{cases} \qquad (3)$$

In the EIC yellow report and ATHENA proposal [3, 4], this procedure was carried out using various inclusive DIS reconstruction methods developed at HERA[2], in fast simulations showing improvements over the electron method in some regions of the DIS kinematic space. As methods such as the Jacquet-Blondel (JB) method [2] use only the hadronic final state information, this also allows for the determination of q from the HFS alone. Results using this approach are shown in the next section compared to ML and electron methods, with resolution using this method expected to improve with further developed full simulations based on fast simulation results.

III. MACHINE LEARNING KINEMATIC RECONSTRUCTION

A. Network architecture

Multiple studies have been conducted demonstrating an improved resolution of inclusive DIS variables Q^2, y, x
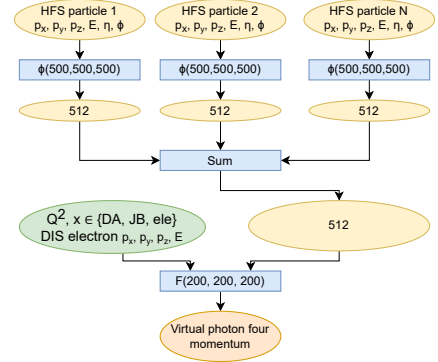

FIG. 3: Network diagram of Particle Flow Networks [9] with global event features. Features of each HFS particle supplied individually to layers Φ, then summed over to form a latent space representation. Latent space features and global features of event (green), including reconstructed inclusive DIS variables and DIS electron momentum, supplied to layer F which produces final output.

through deep learning approaches [7, 8], but these have not yet been extended to reconstruction of semi-inclusive DIS kinematics. In this study, we demonstrate that machine learning models which learn from the full HFS and scattered electron can be used to improve on current reconstruction methods to provide reliable reconstruction of the virtual photon axis across all of the DIS kinematic coverage at the EIC.

This approach to semi-inclusive DIS reconstruction is centered on the use of deep neural networks to better leverage the full hadronic final state at the level of reconstructed tracks. While previous applications of deep learning to inclusive DIS reconstruction directly regressed the kinematic variables of interest [7, 8], this study aims to improve kinematics by directly regressing the virtual photon four-momentum in the lab frame.

Improvements to the HFS reconstruction are carried out through the use of Particle Flow Networks [9]. Particle Flow Networks are an application of the deep sets neural network architecture, which learns a function of an unordered set of objects rather than from a fixed size input. The network consists of fully connected linear neural network layers which take as input the features of each particle individually, the outputs of which are summed over all particles to create a latent space representation of the event. The latent space variables and supplied global features of the event are then passed to another set of dense layers which produce the final output of the network [9]. Particle flow networks have seen particular success in tasks such as jet classification at the LHC. Particle flow networks implemented in Keras [10] are included in the EnergyFlow python package. [9]
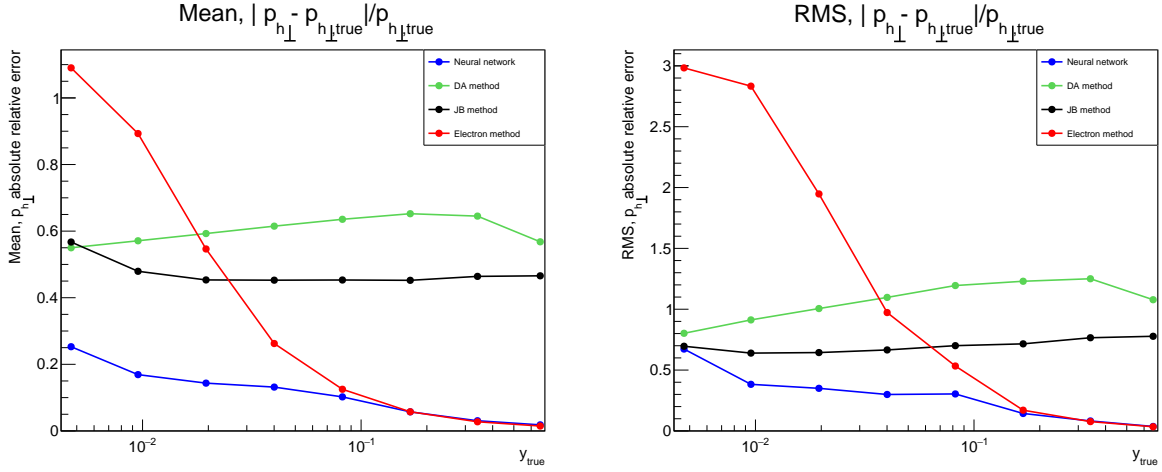
FIG. 4: SIDIS $p_{h\perp}$ resolution mean (left) and RMS (right) as a function of y_{true} for positive pions with $z > 0.2$, $p_{h\perp} > 0.1 GeV$. HFS methods surpass electron method for very low y, while PFN equals or outperforms electron method for all y.
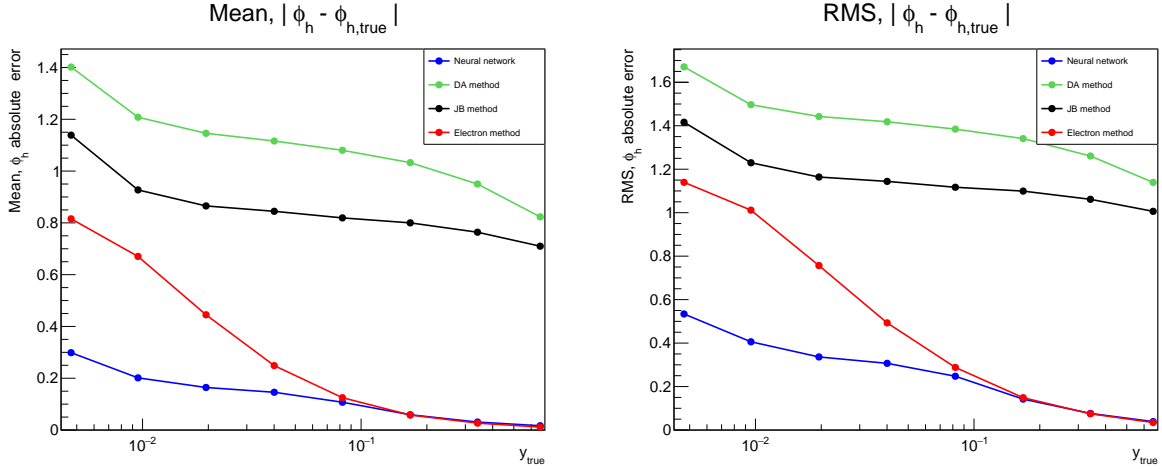
FIG. 5: SIDIS ϕ_h resolution mean (left) and RMS (right) as a function of y_{true} for positive pions with $z > 0.2$, $p_{h\perp} > 0.1 GeV$.

B. Variables and dataset

The features of the hadronic final state reconstructed particles provided to the particle flow network include the four-momentum of each particle, as well as the lab frame azimuthal angle and pseudorapidity to provide direct information on angular acceptance in addition to momentum in each direction.

The global features used for training include the four-momentum of the scattered electron and the DIS variables x and Q^2 from the electron, DA, and JB methods. By supplying the full electron four-momentum following the single-particle layers Φ, the model is intended to learn corrections to the electron method based on the hadronic final state latent space variables. When a greater amount of fully simulated EIC simulated data is available, the DIS methods could also be replaced by the output of the deep learning models for inclusive DIS variables de-scribed previously.

The particle flow network was trained to predict the full four-momentum q in the lab frame. The particle flow network, implemented in Keras [10] and available in the EnergyFlow python package, is used with per-particle dense layer units $\phi = (500, 500, 500)$, $l = 512$, and final dense layer units $F = (200, 200, 200)$. Both the layers making up ϕ and F employ a relu activation function, with the final output layer having linear activation.

The dataset used for the training and testing of this model was the ATHENA full simulation developed for the ATHENA detector proposal for the first interaction region at the EIC. ATHENA was developed with the objective of meeting the resolution and physics goals laid out in the EIC yellow report. The ATHENA full simulation was implemented in DD4hep, Geant4, and Juggler [11–13]. At the time of the detector proposal, PID algorithms were not fully implemented, meaning PID infor-
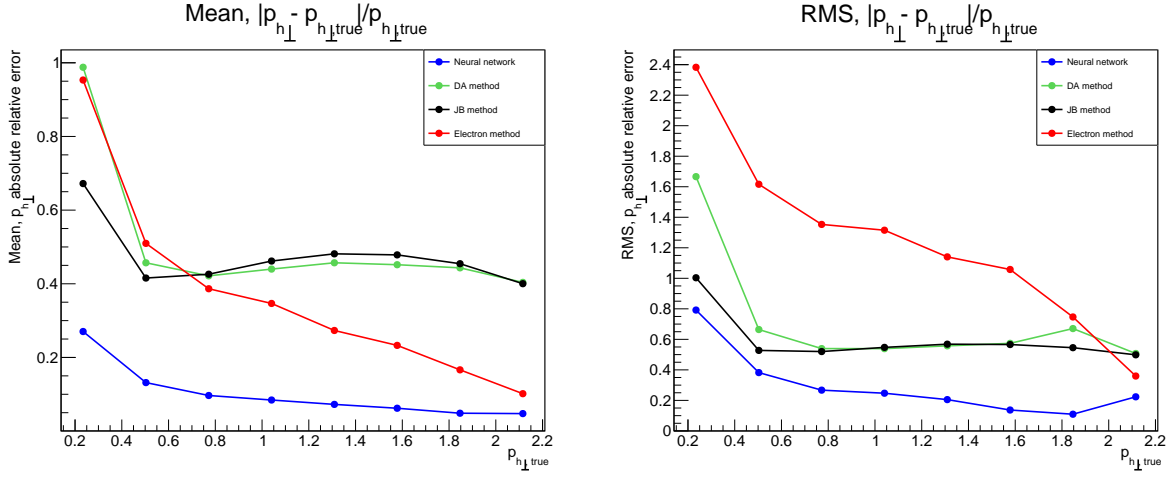
FIG. 6: SIDIS $p_{h\perp}$ resolution mean (left) and RMS (right) as a function of $p_{h\perp,true}$ for positive pions with $z > 0.2$.
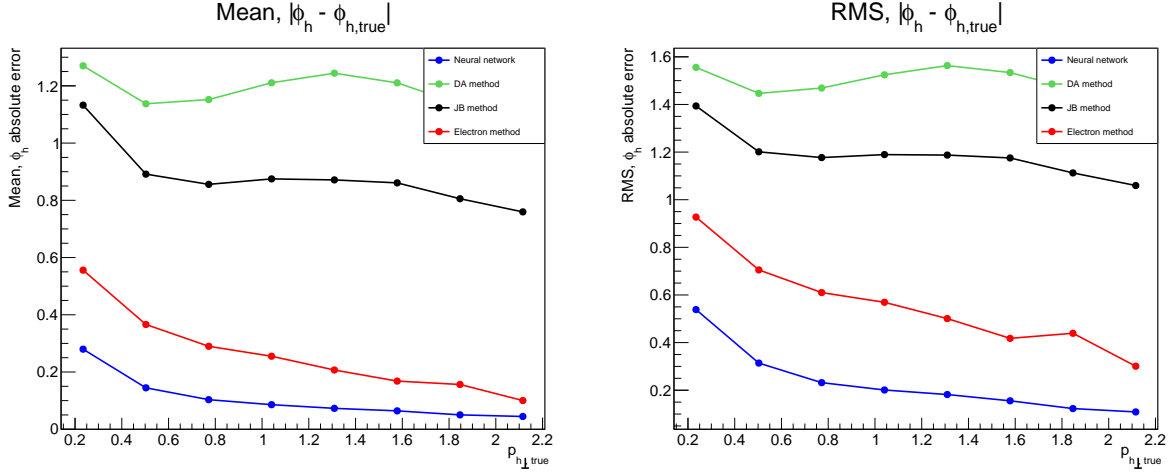

FIG. 7: SIDIS ϕ_h resolution mean (left) and RMS (right) as a function of $p_{h\perp,true}$ for positive pions with $z > 0.2$.

mation was not included in this model. Additionally, the scattered electron was taken as always correctly identified by matching the scattered electron with the MC truth information.

The simulated event sample used for model training and testing was a neutral current DIS sample generated using Pythia-8 [14], with additional beam smearing and crossing angle effects implemented. 3 million events with $Q^2 > 1 \text{ GeV}^2$ and 2 million events with $Q^2 > 10 \text{ GeV}^2$ were used for training with 1 million $Q^2 > 1 \text{ GeV}^2$ set aside for model validation.

IV. RESULTS

As a function of y (figures 4 and 5), using the virtual photon four-momentum as predicted by the neural network model results in significantly improved reconstruction of $p_{h\perp}$, ϕ_h and z for low-y, when compared to both the electron method and methods utilizing informa-

tion from the hadronic final state. The neural network reconstruction of q results in a distribution of the SIDIS variables which is both better centered around the true value, and with a significantly smaller RMS where the electron method begins to fail at low-y. At large-y, the neural network achieves performance only slightly surpassing that of the electron method, which is expected based on the projected energy and tracking resolution for the scattered electron with ATHENA.

As a function of $p_{h\perp,true}$, we also observe a significant improvement in kinematic reconstruction for both transverse momentum and for the semi-inclusive azimuthal angle. As the electron method begins to degrade for lower values of $p_{h\perp}$, the neural network reconstruction of q results in stable performance to the lowest values of $p_{h\perp}$ in the dataset.

V. SUMMARY

The EIC will provide the first opportunity for semi-inclusive DIS measurements in an e-A collider context, giving access to new kinematic regions in which to precisely explore the 3-dimensional spin structure of nucleons. The development of reliable kinematic reconstruction methods will be critical to enabling precision extraction of SIDIS observables, especially at low-y. This can be achieved through the use of information from the hadronic final state alongside the scattered electron. As demonstrated in this contribution, machine learning, here using particle flow networks, can combine the information from the scattered electron and full HFS to provide reliable SIDIS kinematic reconstruction across the DIS variable space. Further steps in this work will include the consideration of QED radiative effects on SIDIS reconstruction, as well as possible extension to other neural network architectures exploiting correlations between particles. Additionally, this approach will continue to be studied and validated as more detailed full detector simulations are developed for the EIC.

[1] A. Bacchetta, U. D'Alesio, M. Diehl, and C. Miller, Single-spin asymmetries: The trento conventions, Physical Review D **70**, 10.1103/physrevd.70.117504 (2004).

[2] J. Blümlein, The theory of deeply inelastic scattering, Progress in Particle and Nuclear Physics **69**, 28–84 (2013).

[3] R. A. Khalek, A. Accardi, *et al.*, Science requirements and detector concepts for the electron-ion collider: Eic yellow report (2021), arXiv:2103.05419.

[4] ATHENA Collaboration, Athena detector proposal for the eic, submitted for publication May, 2022.

[5] U. Bassler and G. Bernardi, On the kinematic reconstruction of deep inelastic scattering at HERA: The Sigma method, Nucl. Instrum. Meth. A **361**, 197 (1995), arXiv:hep-ex/9412004.

[6] H. Abramowicz and A. C. Caldwell, Hera collider physics, Rev. Mod. Phys. **71**, 1275 (1999).

[7] 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, Spectrome-

ters, Detectors and Associated Equipment **1025**, 166164 (2022).

[8] M. Diefenthaler, A. Farhat, A. Verbytskyi, and Y. Xu, Deeply learning deep inelastic scattering kinematics (2021), arXiv:2108.11638.

[9] P. T. Komiske, E. M. Metodiev, and J. Thaler, Energy Flow Networks: Deep Sets for Particle Jets, JHEP **01**, 121, arXiv:1810.05165 [hep-ph].

[10] F. Chollet *et al.*, Keras, `https://keras.io` (2015).

[11] M. Frank, F. Gaede, M. Petric, and A. Sailer, Aidasoft/dd4hep (2018), webpage: http://dd4hep.cern.ch/.

[12] J. Allison, K. Amako, J. Apostolakis, *et al.*, Recent developments in geant4, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **835**, 186 (2016).

[13] Juggler event processor, https://eicweb.phy.anl.gov/eic/juggler.

[14] C. Bierlich, S. Chakraborty, N. Desai, *et al.*, A comprehensive guide to the physics and usage of pythia 8.3 (2022), https://arxiv.org/abs/2203.11601.