

## Extraction of the Siverson function with deep neural networks

I. P. Fernando<sup>\*</sup> and D. Keller<sup>†</sup>

*Department of Physics, University of Virginia, Charlottesville, Virginia 22904, USA*

 (Received 10 March 2023; accepted 9 August 2023; published 8 September 2023)

Deep neural networks (DNNs) are a powerful and flexible tool for information extraction and modeling. In this study, we use DNNs to extract the Siverson functions by globally fitting semi-inclusive deep inelastic scattering (SIDIS) data. To make predictions of this transverse momentum-dependent distribution, we construct a minimally biased model using data from COMPASS and HERMES. The resulting Siverson function model, constructed using SIDIS data, is also used to make predictions for Drell-Yan kinematics specific to the valence and sea quarks, with careful consideration given to experimental errors, data sparsity, and complexity of phase space.

DOI: [10.1103/PhysRevD.108.054007](https://doi.org/10.1103/PhysRevD.108.054007)

### I. INTRODUCTION

The development of modern information extraction techniques has far outpaced experimental progress. Even with decades of experimental results, there is limited data on transverse momentum dependent parton distribution functions (TMDs) for global analyses that can be applied to a 3D phenomenological interpretation. Theoretical efforts are providing an ever-evolving framework and toolbox to interpret the data, leading to new areas of research that optimize information extraction unique to spin physics and the internal structure of hadrons. Artificial intelligence (AI) is accelerating data-driven research with the superior capacity of deep neural networks (DNNs) to be used as function approximators. DNNs can learn to approximate any relationships contained within data, provided there are no limits to the number of neurons, layers, and computing power. With enough data and training, such approximations can be made to nearly any desired degree of accuracy.

The Siverson function is the most studied of the eight leading-twist TMD distributions that pertain to polarized nucleons. The Siverson distribution is naively time-reversal odd and is expected to be process dependent, which leads to a distribution equal in magnitude but opposite in sign in semi-inclusive deep inelastic scattering (SIDIS) compared with the Drell-Yan (DY) process [1,2]. The Siverson distribution, as measured from the transverse single-spin asymmetries, provides information on the correlation between

the nucleon's spin and the angular distribution of outgoing hadrons in SIDIS or the dimuons in DY. The quark Siverson function,  $\Delta^N f_{q/N^\uparrow}$ , describes the number density in momentum space of unpolarized quarks inside the transversely polarized target nucleon with nuclear spin ( $N^\uparrow$ ). A nonzero Siverson function indicates a contribution of quark orbital angular momentum to the target's spin.

TMD extraction and modeling with sensitivity to TMD evolution and factorization is critical and provides predictive power in the collinear limit characterizing the non-perturbative effects. Representing evolution also accounts for the momentum-dependent QCD interactions between partons inside the hadron, which affect their distribution and can significantly impact observables. Some theoretical approaches use the parton model approximation without TMD evolution and demonstrate good agreement with experimental results. At the time of their origin, these calculations assumed that evolution effects in asymmetries are suppressed, as asymmetries are ratios of cross sections where evolution and higher-order effects should cancel out [3–5]. The more modern picture [6] implies this is likely an oversimplification. There is now strong evidence that the main effect of evolution takes place at lower energies ( $< 10$  GeV) [6,7], so it is precisely this domain that offers the best opportunity to capture these important QCD relations in TMD distributions. On the other hand, studies that incorporated TMD evolution, as seen in [8,9], encountered difficulties and did not achieve better agreement with the Drell-Yan data compared with earlier analyses [4,5,10]. This situation poses a challenge in confirming the validity of any theoretical implementation with respect to the TMD factorization theorem or modern utilities required to preserve it, particularly the scale-dependent universal non-perturbative evolution kernel, the so-called Collins-Soper (CS) kernel. However, the work in [7,11] demonstrated the conditional universality of the Siverson function using a

<sup>\*</sup>ishara@virginia.edu

<sup>†</sup>dustin@virginia.edu

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

simultaneous fit to SIDIS, Drell-Yan, and W/Z boson production data, including TMD evolution and the universal nonperturbative CS kernel extracted in [12] from unpolarized measurements. In contrast, the working principle of DNNs as a means of extraction can facilitate the necessary flexibility to capture not only the TMD and Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution features but also the full complexity of QCD. In this regard, a carefully designed DNN model accompanied by a systematic method of extraction can incorporate not only the evolution features but other components not yet quantifiable in the formalism. The inherent capacity for DNNs to reduce dimensionality and organize intricate correlations in the data can be used to help determine the correct theoretical implementation. The challenge, of course, is that once captured in a DNN model, disentangling these features into formal expressions consistent with factorization is nontrivial, and we expect it to be the focus of much future work.

The approach outlined in this article highlights the inherent flexibility of employing specialized DNNs in combination with candidate multiplicative expressions holding some type of separate parametrization which may carry incorrect Ansätze or assumptions (biases). This can be done while still rendering a high-quality numerical model encoding all information available in the data. These types of biases can be managed by training the model with the biased term included. The DNN term then becomes uniquely parametrized to account for the biases. Consequently, the resulting combination of terms in the model remains largely unbiased, even though each individual term may not be independently meaningful. The schema can not only facilitate the exploration of the existing formalism found in the literature but also provide a means to derive and test novel phenomenology using a technique that can directly manage and study the biases in individual terms. This will be discussed in more detail in Sec. VI.

The phenomenology used to interpret and analyze experimental data relies on TMD factorization [13–22], proven for single-spin asymmetries described in terms of convolutions of TMDs. The Siverson function has been previously extracted from SIDIS data by several groups, with generally consistent results [4,23–27]. All previous phenomenological fits of the Siverson function (and other TMDs) require an Ansatz characterizing the shape of the distribution combined with an assumed form of the Bjorken- $x$  dependence. This can lead to ambiguity in determining both the quantitative results from the fit and the qualitative features of the momentum distributions and their associated dynamics. The function form of the Bjorken- $x$  dependence is usually offered only as a placeholder and is assumed to at least contain the appropriate ingredients to facilitate the extraction. This is undoubtedly a considerable oversimplification but one that has permitted significant progress. In the following analysis, we perform a global fit with the goal of illustrating a method that can also

permit significant progress even with data limitations, and with assumptions, and Ansätze in place. We focus on testing the extraction ability of a DNN representing a single term to maximize information extraction and minimize both the fit error and the analytical ambiguity associated with the interpretation of a generic Bjorken- $x$  dependence,  $\mathcal{N}_q(x)$ . There are, of course, broader implications of how these tools can be used to accelerate the field, some of which we will also touch on through our extraction of the Siverson function.

The exceptional capacity of DNNs to be ideal for function approximation is rigorously provable through the universal approximation theorem [28,29]. This is the advantage of DNNs over other machine learning approaches. In this regard, even the mere existence of a function implies that DNNs can be used to represent it and work with it without actually knowing the function form. With such a high-level abstraction, one can make use of the available data and make assessments not otherwise possible, even given an arbitrary degree of complexity. The complexity can be contained in the data relationships as well as in the experimental uncertainties. In order to make optimal use of experimental uncertainties, a detailed analysis must be provided on their estimated scale and correlations.

DNNs are also Turing complete, implying the potential to simulate the computational aspects of any real-world general-purpose computing process. The implications are that there is potential for a type of generalizable framework that can be utilized and further developed over time without the knowledge of the exact rigorous details of the underlying mechanisms. Provided appropriately detailed global models, higher-level symbolic regressions can then be performed to infer the strict mathematical form. However, such an approach requires access to a significant amount of experimental data that hold the necessary information. Without the necessary amount of quality data, no matter the number of nodes or sophistication of architecture, DNNs are limited in what useful information they can extract, along with any other technique. Even with this constraint, DNNs can make considerable advancements with the use of sparse data with large experimental uncertainties. Generally, fitting with large errors or performing computational tasks with inherent fuzzy logic are tasks that are difficult to make optimal use of modern computational resources. DNNs are uniquely suited for such challenges.

In the remainder of this paper, a first-level DNN extraction of  $\mathcal{N}_q(x)$  is performed to deduce the Siverson function from a global analysis using HERMES and COMPASS data. This investigation is exploratory, with the intention of developing tools and techniques that minimize error and maximize utility, which we hope to expand upon in further work. The choice in this paper is to focus on the Siverson function, rather than the unpolarized TMD  $f_{q/N}(x, k_\perp)$ , and to more clearly demonstrate the power of the method even given the limitations in data. Our global fit of  $f_{q/N}(x, k_\perp)$  will be presented in later work. The examination of  $\mathcal{N}_q$  in relation to

$x$  is prioritized over other possibilities only to illustrate how generating functions can be utilized to measure and analyze systematics arising from the extraction produced. As we demonstrate in this work, this technique enables a systematic enhancement of information extraction.

In Sec. II, we present the formalism of the Sivers function and the kinematics for both SIDIS and DY. In Sec. III, a discussion of the fitting techniques of  $\mathcal{N}_q(x)$  is presented with a focus on the methodology of the DNN approach. Section IV explains in detail the extraction technique, starting with model testing. We perform a baseline fit using the classical MINUIT  $\chi^2$  minimization algorithm and then perform the DNN fit, demonstrating with pseudodata the fidelity of the procedure. We then walk through the final DNN fit to experimental data for the polarized proton and the deuteron separately. The results of the fits are presented in Sec. V. Some further analysis and systematic studies are reported in Sec. VI. Then in Sec. VII we present the implication of the results in terms of the 3D tomography of the proton. In Sec. VIII, we perform a preliminary analysis on TMD evolution and finally, in Sec. IX, some concluding remarks are provided.

## II. KINEMATICS AND FORMALISM

With the spin of the proton perpendicular to the transverse plane, the Sivers function is expected to reflect an anisotropy of quark momentum distributions for the up and down quarks, indicating that their motion is in opposite directions [1,2]. This is manifestly due to quark orbital angular momentum (OAM). The most interesting and relevant aspects of the OAM, such as magnitude and partonic distribution shape as a function of the proton's state, cannot be determined by the Sivers effect alone. However, systematic studies can be performed to investigate the full 3D momentum distribution of the quarks in a transversely polarized proton, which can be used in concert with other information to exploit multidimensional partonic degrees of freedom using a variety of hard processes. Here, we focus specifically on SIDIS and DY, but it should be noted that there is significant potential for broader model development that can come from combining all available data from multiple processes with additional constraints using the simultaneous DNN fitting approach presented here.

The Sivers function describes a difference in probabilities, which implies that it may not be positive definite. Making a comparison between the Sivers function from the DY process and the SIDIS process is still the focus of much experimental and theoretical effort. Under time reversal, the future-pointing Wilson lines are replaced by past-pointing Wilson lines that are appropriate for factorization in the DY process. This implies that the Sivers function is not uniquely defined and cannot exhibit process universality, as it depends on the contour of the Wilson line. This feature of the Sivers function is directly tied to the QCD interactions between the quarks (or gluons) active in the

process, resulting in a conditional universality, as shown in [30],

$$\Delta^N f_{q/N^\uparrow}(x, k_\perp)|_{\text{SIDIS}} = -\Delta^N f_{q/N^\uparrow}(x, k_\perp)|_{\text{DY}}. \quad (1)$$

This fundamental prediction still needs to be tested. Direct sign tests [4,10,31] can be performed, but experimental proof would require an analysis over a broad phase space of both SIDIS and DY, with consideration given to flavor and kinematic sensitivity for both valence and sea quarks. Our analysis will, in part, rely on this relationship rather than making direct tests of the validity of the sign change.

### A. SIDIS process

The SIDIS process involves scattering a lepton off of a polarized nucleon and measuring the scattered lepton and a fragmented hadron. In the nucleon-photon center of mass frame, the nucleon three-momentum  $\vec{p}$  is along the  $z$  axis, and its spin polarization  $\vec{S}_T$  is on the plane perpendicular (transverse) to the  $\hat{z}$  axis. In Fig. 1 the struck quark, virtual photon (with four-momentum  $\vec{q}$ ), and the lepton belong to a plane called the ‘‘lepton plane’’ (represented in yellow). The fragmented hadron with momentum  $\vec{p}_h$  and its projection onto the  $\hat{x} - \hat{y}$  (i.e.  $\vec{p}_{hT}$ ) belong to the so-called ‘‘hadron plane’’ (represented in transparent green). Thus, the transverse momentum  $\vec{k}_\perp$  of the struck quark and  $\vec{p}_{hT}$  are falling onto the transverse plane (represented in transparent blue) perpendicular to both the lepton plane and hadron plane. The azimuthal angle  $\phi_h$  of the produced hadron  $h$  is the angle between the lepton plane and the hadron plane [32].

To perform a global analysis using a generalized DNN approach, it is not necessary to postulate an expression or function form for the shape of the partonic distributions or any aspects of the nonperturbative contribution which encodes information about the intrinsic quark-gluon correlations within the nucleon. But to provide a more careful demonstration we do not generalize and focus directly on

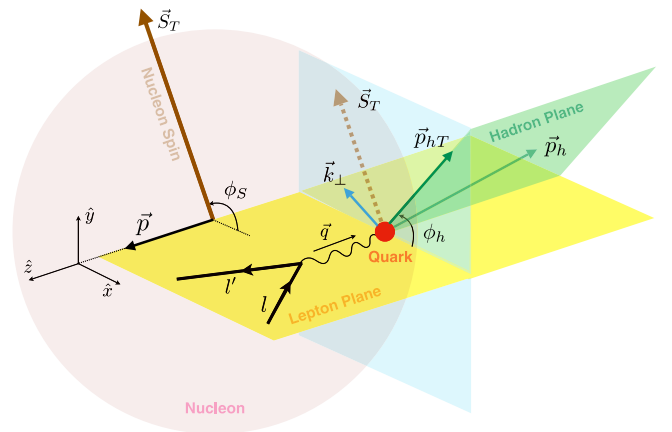


FIG. 1. Kinematics of the SIDIS process in the nucleon-photon center-of-mass frame.

$\mathcal{N}_q(x)$  to explore using quality metrics in combination with techniques to improve the overall extraction. It is useful to demonstrate this using a very standard approach that is already well-developed in the literature. With this in mind the TMDs (fragmentation functions)  $x$  and  $k_\perp$  ( $z$  and  $p_\perp$ ) dependence are decoupled. It is common to apply a Gaussian parametrization for the transverse momentum dependence, but this imposes a bias implying that the transverse momentum is nonperturbative and primarily driven by the intrinsic properties of the colliding hadrons rather than hard gluon radiation. Such biases can be managed by the technique itself which we expect to come clear through the examples to come. There are several options for the choice of formalism, and there is still much ongoing debate on how formalism should correctly be implemented for single-spin asymmetries to best respect factorization theory. Our choice of formalism is purely illustrative. There remains much ongoing debate about this, and we intend no favoritism in one strategy over another. The focus of the present work is the demonstration of how the methods and tools can be implemented for arbitrary formalism. There are however some very useful characteristics in one of the traditional descriptions of the TSSA [33] that offer a clear means of detailing the necessary steps. However, we stress that the type of parametrization used in [4,34] does not have the complete features of TMD evolution but is comparable to full TMD evolution at next-to-leading logarithmic accuracy.

The differential cross section for the SIDIS process depends on both collinear parton distribution functions (PDFs)  $f_{q/N}(x; Q^2)$  and fragmentation functions  $D_{h/q}(z; Q^2)$ , where  $q$  is the quark flavor,  $N$  represents the target nucleon,  $h$  is the hadron type produced by the process, and  $z$  is the momentum fraction of the final-state hadron with respect to the virtual photon. A simplified version of the SIDIS differential cross section can be written up to  $\mathcal{O}(k_\perp/Q)$  as [26,35]

$$\begin{aligned} \frac{d^5\sigma^{IN\rightarrow lhX}}{dx dQ^2 dz d^2p_\perp} &= \sum_q e_q^2 \int d^2\mathbf{k}_\perp \left( \frac{2\pi\alpha^2 \hat{s}^2 + \hat{u}^2}{x^2 s^2 Q^4} \right) \\ &\times \hat{f}_{q/N^\dagger}(x, k_\perp) D_{h/q}(z, p_\perp) + \mathcal{O}(k_\perp/Q), \end{aligned} \quad (2)$$

where  $\hat{s}, \hat{u}$  are partonic Mandelstam invariants, and  $\hat{f}_{q/p^\dagger}(x, k_\perp)$  is the unpolarized quark distribution,

$$\begin{aligned} \hat{f}_{q/N^\dagger}(x, k_\perp) &= f_{q/N}(x, k_\perp) + \frac{1}{2} \Delta^N f_{q/N^\dagger}(x, k_\perp) \vec{S}_T \cdot (\hat{p} \times \hat{k}_\perp) \\ &= f_{q/N}(x, k_\perp) - \frac{k_\perp}{m_p} f_{1T}^{\perp q}(x, k_\perp) \vec{S}_T \cdot (\hat{p} \times \hat{k}_\perp) \end{aligned} \quad (3)$$

with transverse momentum  $k_\perp$  inside a transversely polarized (with spin  $\vec{S}_T$ ) proton with three-momentum  $\vec{p}$ , where

$\Delta^N f_{q/p^\dagger}(x, k_\perp)$  denotes Sivvers functions that carry the nucleon's spin-polarization effects on the quarks which can be considered as a modulation to the unpolarized quark PDFs [4],

$$\Delta^N f_{q/N^\dagger}(x, k_\perp) = 2\mathcal{N}_q(x) h(k_\perp) f_{q/N}(x, k_\perp), \quad (4)$$

where

$$f_{q/N}(x, k_\perp) = f_q(x) \frac{1}{\pi \langle k_\perp^2 \rangle} e^{-k_\perp^2 / \langle k_\perp^2 \rangle}, \quad (5)$$

$$h(k_\perp) = \sqrt{2} e \frac{k_\perp}{m_1} e^{-k_\perp^2 / m_1^2}. \quad (6)$$

Here  $\mathcal{N}_q(x)$  is considered a factorized  $x$ -dependent function with a form that has yet to be formally established. In fact the form of  $f_{q/N}(x, k_\perp)$  and  $h(k_\perp)$  are also assumed, and these expressions should be considered embedded biases. For the best quality model of the Sivvers function, the form of  $f_{q/N}(x, k_\perp)$  could be determined and parametrized separately using unpolarized data; however it is instructive to preserve these original forms for this exercise. The parameter  $m_1$  allows the  $k_\perp$  Gaussian dependence of the Sivvers function to be different from that of the unpolarized TMDs [4].  $f_q(x; Q^2)$  is the collinear PDF for flavor  $q$  that is obtained from the CTEQ61 [36] grid through LHAPDF [37] initially to be consistent with [4] during method testing and later improved using NNPDF4.0 [38] for the real extraction, whereas the fragmentation functions for  $\pi^{\pm,0}$  are from [39], and for  $K^\pm$  are from [40] (DSS formalism), from recent global analyses of fragmentation functions at next-to-leading-order accuracy in QCD.

In terms of the cross section ratios, the Sivvers asymmetry in the SIDIS process can be written as

$$A_{UT}^{\sin(\phi_h - \phi_S)}(x, y, z, p_{hT}) = \frac{d\sigma^{\uparrow N \rightarrow h l X} - d\sigma^{\downarrow N \rightarrow h l X}}{d\sigma^{\uparrow N \rightarrow h l X} + d\sigma^{\downarrow N \rightarrow h l X}} \quad (7)$$

and can be parametrized [4] and further rearranged as

$$\begin{aligned} A_{UT}^{\sin(\phi_h - \phi_S)}(x, z, p_{hT}) &= \mathcal{A}_0(z, p_{hT}, m_1) \left( \frac{\sum_q \mathcal{N}_q(x) e_q^2 f_q(x) D_{h/q}(z)}{\sum_q e_q^2 f_q(x) D_{h/q}(z)} \right), \end{aligned} \quad (8)$$

where

$$\begin{aligned} \mathcal{A}_0(z, p_{hT}, m_1) &= \frac{\sqrt{2} e z p_{hT} [z^2 \langle k_\perp^2 \rangle + \langle p_\perp^2 \rangle] \langle k_S^2 \rangle^2}{m_1 [z^2 \langle k_S^2 \rangle + \langle p_\perp^2 \rangle]^2 \langle k_\perp^2 \rangle} \\ &\times \exp \left[ -\frac{p_{hT}^2 z^2 (\langle k_S^2 \rangle - \langle k_\perp^2 \rangle)}{(z^2 \langle k_S^2 \rangle + \langle p_\perp^2 \rangle) (z^2 \langle k_\perp^2 \rangle + \langle p_\perp^2 \rangle)} \right], \end{aligned} \quad (9)$$

$$\langle k_S^2 \rangle = \frac{m_1 \langle k_\perp^2 \rangle}{m_1^2 + \langle k_\perp^2 \rangle}, \quad (10)$$

and fragmentation functions  $D_{h/q}(z, p_\perp)$  (before  $p_\perp$  integration),

$$D_{h/q}(z, p_\perp) = D_{h/q}(z) \frac{1}{\pi \langle p_\perp^2 \rangle} \exp^{-p_\perp^2 / \langle p_\perp^2 \rangle}, \quad (11)$$

with  $\langle k_\perp^2 \rangle = 0.57 \pm 0.08 \text{ GeV}^2$  and  $\langle p_\perp^2 \rangle = 0.12 \pm 0.01 \text{ GeV}^2$  from the fits [41,42] to HERMES multiplicities [43]. Note that we use the shorthand notation for the PDFs and fragmentation functions as well as TMDs by omitting  $Q^2$  in the expressions for the sake of convenience as is done in the literature.

Through this azimuthal asymmetry, the SIDIS process provides information about the correlations between the transverse momentum of the partons leaving through the fragmented target and the spin of the target itself. In this regard, SIDIS allows one to study the structure of individual hadrons by selecting these decay fragments at the detection level. In general, SIDIS provides access to a wide range of TMDs, and allows for studying TMDs of hadrons carrying different flavors and polarizations.

For our present analysis, HERMES and COMPASS have the best-polarized proton target data for SIDIS, while COMPASS has the best-polarized neutron target data. In

the COMPASS data, the neutron target is actually a polarized deuteron, but the neutron carries over 90% of the deuteron polarization when polarized in solid-state form. The JLab data on polarized  $^3\text{He}$  is of a different class of experiments and will not be combined with the polarized deuteron data from COMPASS. It is worth noting that the uncertainties in the experimental data can greatly differ depending on the choice of polarized target.

## B. DY process

Consider the Drell-Yan process  $A^\uparrow B \rightarrow l^+ l^- X$ , where  $A^\uparrow$  is a transversely polarized target, and  $B$  is the hadron beam. In the hadronic c.m. frame (see Fig. 2), the 4-momentum  $q$  and the invariant mass squared ( $Q_M$ ) of the final-state dilepton pair, Feynman  $x$  ( $x_F$ ) and the Mandelstam variable  $s$  are related as

$$q = (q_0, q_T, q_L), \quad q^2 = Q_M, \quad x_F = \frac{2q_L}{\sqrt{s}},$$

$$s = (p_A + p_B)^2. \quad (12)$$

In the kinematical region of

$$q_T^2 \ll Q_M, \quad k_\perp \simeq q_T, \quad (13)$$

at order  $\mathcal{O}(k_\perp/Q_M)$ , and in the hadronic c.m. frame, the Sivers single spin asymmetry can be given as [3,33]

$$A_N^{\sin(\phi_\gamma - \phi_S)}(x_F, Q_M, q_T) = \frac{\int_0^{2\pi} d\phi_\gamma (d\sigma^{A^\uparrow B \rightarrow l^+ l^- X} - d\sigma^{A^\downarrow B \rightarrow l^+ l^- X}) \sin(\phi_\gamma - \phi_S)}{\frac{1}{2} \int_0^{2\pi} d\phi_\gamma (d\sigma^{A^\uparrow B \rightarrow l^+ l^- X} + d\sigma^{A^\downarrow B \rightarrow l^+ l^- X})}, \quad (14)$$

$$= \frac{\int_0^{2\pi} d\phi_\gamma (\sum_q \int d^2 k_{\perp 2} d^2 k_{\perp 1} \delta^2(k_{\perp 1} + k_{\perp 1} - q_T) \Delta^N f_{q/A^\uparrow}(x_1, k_{\perp 1}) f_{\bar{q}/B}(x_2, k_{\perp 2}) \hat{\sigma}_0^{q\bar{q}}) \sin(\phi_\gamma - \phi_S)}{\int_0^{2\pi} d\phi_\gamma (\sum_q \int d^2 k_{\perp 2} d^2 k_{\perp 1} \delta^2(k_{\perp 1} + k_{\perp 1} - q_T) f_{q/A}(x_1, k_{\perp 1}) f_{\bar{q}/B}(x_2, k_{\perp 2}) \hat{\sigma}_0^{q\bar{q}})}, \quad (15)$$

where

$$\hat{\sigma}_0^{q\bar{q}} = e_q^2 \frac{4\pi\alpha^2}{9Q_M}, \quad (16)$$

$$x_{1,2} = \frac{\pm x_F + \sqrt{x_F^2 + 4Q_M/s}}{2}. \quad (17)$$

Note that here we follow the same convention as in [3,25,44,45] for the azimuthal angle in the  $A - B$  center-of-mass frame with the hadron  $A^\uparrow$  moving along the positive  $z$  axis and hadron  $B$  along the negative  $z$  axis. Thus the mixed product  $\vec{S}_T \cdot (\hat{p} \times \hat{k}_{\perp i})$  upon integration in  $k_{\perp i}$  (where  $i = \{1, 2\}$ ) yields a  $\sin(\phi_\gamma - \phi_S) = \cos \phi_\gamma$  (when

$\phi_S = \pi/2$ ) dependence for the Sivers asymmetry, which implies an overall  $[-\sin^2(\phi_\gamma - \phi_S)]$  in Eq. (15). For the case in which polarized hadron  $A^\uparrow$  moves along the  $-\hat{z}$  axis (i.e. for the processes  $BA^\uparrow \rightarrow l^+ l^- X$ ), the corresponding overall factor is  $[\sin^2(\phi_\gamma - \phi_S)]$ . The analytical integration of the numerator and denominator of Eq. (15) can be written as

$$A_N^{\sin(\phi_\gamma - \phi_S)}(x_F, Q_M, q_T) = \frac{\int d\phi_\gamma \mathcal{C}(x_F, Q_M, q_T, \phi_\gamma) \sin(\phi_\gamma - \phi_S)}{\int d\phi_\gamma \mathcal{D}(x_F, Q_M, q_T)}, \quad (18)$$

where

$$\mathcal{C}(x_F, Q_M, q_T, \phi_\gamma) \equiv \frac{d^4\sigma^\uparrow}{dx_F dQ_M d^2q_T} - \frac{d^4\sigma^\downarrow}{dx_F dQ_M d^2q_T} \quad (19)$$

$$= \frac{4\pi\alpha^2}{9sQ_M} \left( \frac{q_T}{m_1} \sqrt{2e} \frac{\langle k_S^2 \rangle^2 \exp[-q_T^2 / (\langle k_S^2 \rangle + \langle k_{\perp 2}^2 \rangle)]}{\pi(\langle k_S^2 \rangle + \langle k_{\perp 2}^2 \rangle)^2 \langle k_{\perp 2}^2 \rangle} \right) \sin(\phi_\gamma - \phi_S) \sum_q \frac{e_q^2}{x_1 + x_2} \Delta^N f_{q/A^\uparrow}(x_1) f_{\bar{q}/B}(x_2), \quad (20)$$

and

$$\begin{aligned} \mathcal{D}(x_F, Q_M, q_T) &\equiv \frac{1}{2} \left[ \frac{d^4\sigma^\uparrow}{dx_F dQ_M d^2q_T} + \frac{d^4\sigma^\downarrow}{dx_F dQ_M d^2q_T} \right] \\ &= \frac{4\pi\alpha^2}{9sQ_M} \left( \frac{\exp[-q_T^2 / (\langle k_{\perp 1}^2 \rangle + \langle k_{\perp 2}^2 \rangle)]}{\pi(\langle k_{\perp 1}^2 \rangle + \langle k_{\perp 2}^2 \rangle)} \right) \sum_q \frac{e_q^2}{x_1 + x_2} f_{q/A}(x_1) f_{\bar{q}/B}(x_2), \end{aligned} \quad (21)$$

and it can be further simplified as

$$A_N^{\sin(\phi_\gamma - \phi_S)}(x_F, M, q_T) = \mathcal{B}_0(q_T, m_1) \frac{\sum_q \frac{e_q^2}{x_1 + x_2} \mathcal{N}_q(x_1) f_{q/A}(x_1) f_{\bar{q}/B}(x_2)}{\sum_q \frac{e_q^2}{x_1 + x_2} f_{q/A}(x_1) f_{\bar{q}/B}(x_2)}, \quad (22)$$

where

$$\mathcal{B}_0(q_T, m_1) = \frac{q_T \sqrt{2e}}{m_1} \frac{Y_1(q_T, k_S, k_{\perp 2})}{Y_2(q_T, k_{\perp 1}, k_{\perp 2})}, \quad (23)$$

and

$$Y_1(q_T, k_S, k_{\perp 2}) = \left( \frac{\langle k_S^2 \rangle^2}{\langle k_{\perp 2}^2 \rangle (\langle k_S^2 \rangle + \langle k_{\perp 2}^2 \rangle)} \right) \exp\left( \frac{-q_T^2}{\langle k_S^2 \rangle + \langle k_{\perp 2}^2 \rangle} \right), \quad (24)$$

$$Y_2(q_T, k_{\perp 1}, k_{\perp 2}) = \left( \frac{1}{\langle k_{\perp 1}^2 \rangle + \langle k_{\perp 2}^2 \rangle} \right) \exp\left( \frac{-q_T^2}{\langle k_{\perp 1}^2 \rangle + \langle k_{\perp 2}^2 \rangle} \right), \quad (25)$$

$$\frac{1}{\langle k_S^2 \rangle} = \frac{1}{m_1^2} + \frac{1}{\langle k_{\perp 1}^2 \rangle}, \quad (26)$$

with the assumption  $\langle k_{\perp 1}^2 \rangle = \langle k_{\perp 2}^2 \rangle = \langle k_{\perp}^2 \rangle = 0.25 \text{ GeV}^2$  as in [3].

Through this azimuthal asymmetry, the Drell-Yan process allows one to preferentially probe from the target and beam hadrons to create the quark antiquark annihilation process of interest resulting in a dimuon pair in the detector. SIDIS only permits the measurement of a convolution of the TMDs function with a fragmentation function, whereas Drell-Yan allows the direct measurement of the TMDs without the complications of fragmentation functions and final-state interactions. Coupled with its innate sensitivity

to sea quarks, Drell-Yan is a critical process for determining the TMDs of the sea quarks.

### III. FITTING $\mathcal{N}_q(x)$

To obtain accurate three-dimensional tomographic information on quarks and gluons inside the nucleon, it is critical to extract TMDs with minimal model dependence and little to no unknown biases. Fitting with statistical analysis tools such as MINUIT [46] relies on  $\chi^2$  minimization or log-likelihood functions to compute the best-fit parameter values and uncertainties, including correlations between parameters. This class of algorithms has well-established statistical methods that have been used for decades in various scientific fields, making them a reliable and trusted

tool. In frequentist statistics, the reliability of a  $\chi^2$  minimization fitting method can be evaluated through the concept of hypothesis testing. The fitting method minimizes the difference between the observed data and the expected theoretical model, expressed through a  $\chi^2$  statistic. The  $\chi^2$  statistic follows a known distribution, and the probability of obtaining a value as extreme as the observation can be directly calculated. The reliability of the  $\chi^2$  minimization fitting method depends greatly on the ability to accurately estimate the theoretical uncertainties and the degree to which the model approximates the observed data. When these conditions are met, the method can provide a reliable estimate of the parameters that describe the model and its uncertainties. However, chi-square fits can be sensitive to the choice of initial parameter values and may not always converge to the correct solution. Fitting with DNNs can provide considerable advantages and does not inherently sacrifice the statistical framework provided by chi-square fits, but it is worth touching on some key attributes needed in the method in order to best maintain quality statistical relevance and interpretation of resulting fits.

To preserve the statistical robustness and reliability of traditional  $\chi^2$  minimization fitting when using a DNN, it is important to carefully consider the data quality, model selection, validation, interpretation, and testing criteria. The quality of the data used to train the DNN should be as high as possible to ensure the DNN learns the correct relationships between inputs and outputs. This is crucial because the reliability of the DNN is only as good as the quality of the data it is trained on. Quantifying differences between the training data and the real data used in the fit can be challenging and can lead to unknown biases and systematic errors. In the method used here, Monte Carlo data, which has been tuned and matched to the experimental data, is utilized. This is done by successively extracting information from the experimental data to impose into the generated Monte Carlo data and then using the improved Monte Carlo data to further refine the extraction technique.

The choice of DNN architecture, activation functions, regularization techniques, and other hyperparameters should be carefully selected to minimize overfitting and maximize generalization performance. Cross-validation techniques can be used to tune these hyperparameters and ensure the best possible fit to the data. The quality of the fit should be quantified with a metric that is well-defined and can be interpreted statistically. This could still be the  $\chi^2$  statistic but may also be a variety of possible loss functions. The trained DNN should be validated on an independent test dataset to ensure that it generalizes well to new data and that it does not overfit the training data. This is critical because overfitting can lead to an unreliable and unstable model.

When interpreting the results of the DNN fit, it is important to carefully examine the relationships between

the inputs and outputs. To better understand how the DNN makes its predictions, techniques such as feature importance and attention mechanisms can be used. While DNNs can have a reliable statistical interpretation, it requires a more detailed analysis than traditional algorithms like MINUIT.

Directed testing of the model predictions and verification of reliability through multiple trials are crucial. Studies to test accuracy and precision are useful along with quantifying the robustness of the extraction method itself once a DNN architecture has been chosen. In this regard, it is important to prove that the method can be flexible as well as correct. It is most useful to have a model and method that yield consistently accurate results for a broad phase space.

The conventional chi-square minimization routines are limited in their flexibility and applicability to more complex problems because they assume a specific functional form of the relationship between inputs and outputs. In contrast, DNNs can learn complex and nonlinear relationships, making them suitable for tasks that require some degree of abstraction and where there is no known specific functional form of the relationship. DNNs are proven to be universal approximators and can handle large amounts of data while generalizing well to new data, which improves their accuracy and robustness. This is especially helpful in the present application, where DNNs can be used to build better models with new experimental data as they become available.

The ambiguity in the literature around  $\mathcal{N}_q(x)$  and the accompanying factorized terms in the Sivers function ( $h(k_\perp)$  and  $f_{q/N}(x, k_\perp; Q^2)$ ) makes it a good candidate for a DNN extraction. In most Sivers function extractions [5,8,23,25–27,35,47–51],  $\mathcal{N}_q(x)$  differs either by its parametrization or by the treatment of  $q$  in  $\mathcal{N}_q(x)$ . For example, in [7,11,26,27], all antiquarks ( $\bar{u}$ ,  $\bar{d}$ , and  $\bar{s}$ ) were treated the same (or combined) and referred to as the “unbroken sea.”

Our first step is a generalization of the MINUIT fit parametrization of  $\mathcal{N}_q(x)$  for all light-quark flavors. Section IV B summarizes the corresponding MINUIT fit results using *iminuit* (Python version of MINUIT) [46]. In these fits we use the same dataset as in [4], and obtained the fit parameters for  $\mathcal{N}_q(x)$  defined as

$$\mathcal{N}_q(x) = N_q x^{\alpha_q} (1-x)^{\beta_q} \frac{(\alpha_q + \beta_q)^{(\alpha_q + \beta_q)}}{\alpha_q^{\alpha_q} \beta_q^{\beta_q}}. \quad (27)$$

This expression is generalized for all light-quark flavors, where  $N_q$  is a scalar for quark flavor  $q$ . After our consistency check, this parametrization is used as a pseudodata generator to train and test the DNN model. We emphasize that our original MINUIT fit parametrization is used as a tool to demonstrate that the DNN model is capable of predicting (or confirming) the 19-parameter model used to generate pseudodata, as illustrated in

Section IV C. It should be clarified that we do not rely on the form of Eq. (27) in any way and any function that can be used to generate quality<sup>1</sup> pseudodata could be used. After building confidence in the DNN model from these preliminary tests, we move toward extracting the Sivvers functions from experimental data from the SIDIS process with a polarized-proton target (see Sec. IV D). Previous work on global fits to SIDIS data considered the data from polarized-proton targets, polarized-deuteron targets, and polarized <sup>3</sup>He gas targets as a combined dataset. This is usually motivated only by the overall lack of separate polarized target data. In [7], isospin symmetry was assumed for  $f_{1T,u}^\perp$  and  $f_{1T,d}^\perp$  for COMPASS2009 [52] and COMPASS2017 [53] datasets. However, these are very different polarized targets with very different results, so we explore if our method will be sensitive to flavor and target dependence by fitting each polarized target data independently. As the nuclear effects on the Sivvers functions are not very well understood, separately extracting the same observable with different types of polarized targets that contain different flavor dominance and different nuclear effects could provide valuable insight. Significant data would be required for each target type, so there are clear limitations at the moment, and our fits can only be considered preliminary in this regard.

The DNN's unique capacity to manage abstraction allows for the capture of additional complexity in a semi-model-independent way. To obtain the most information from the data and fit results we attempt to decompose some of this abstraction using the following conditions:

- (1) DNN fits to SIDIS data from proton and deuteron targets are performed independently to obtain separate models.
- (2) No kinematic cuts are applied to take full advantage of the available data and to allow the DNN to build implicit inclusion of the necessary corrections related to TMD factorization.
- (3) A Sivvers function for each light-quark flavor is obtained to ensure the  $SU(3)_{\text{flavor}}$  breaking effects in QCD are also contained.

The technique to achieve the extraction under the aforementioned conditions is somewhat novel, so explicit details are provided step by step for clarity in the next section.

#### IV. THE EXTRACTION TECHNIQUE

The factorized form,  $\mathcal{N}_q(x)h(k_\perp)f_{q/N}(x, k_\perp)$  assumes the validity of Eq. (4), leading to model dependencies associated with this assertion and the Gaussian interpretation for the  $k_\perp$  distributions of the partons. This introduces a bias to the analysis, but it is a known bias and can be

<sup>1</sup>The term quality here refers to how well it can represent the real experimental data.

managed. Within the construction of a DNN model in the fit function,

$$\mathcal{N}_q^{\text{DNN}}(x)h(k_\perp)f_{q/N}(x, k_\perp), \quad (28)$$

there could be considerable variation in definitions (or assumptions) of  $h(k_\perp)$  or  $f_{q/N}(x, k_\perp)$  still leading to a quality fit of  $\mathcal{N}_q^{\text{DNN}}(x)$  to the data due to its considerable flexibility and high number of parameters with respect to either of the other two terms in the function. This makes the final  $\mathcal{N}_q^{\text{DNN}}(x)$  highly dependent on the choice of  $h(k_\perp)$  and  $f_{q/N}(x, k_\perp)$  so the resulting  $\mathcal{N}_q^{\text{DNN}}(x)$  model must always be used with the same definitions of  $h(k_\perp)$  and  $f_{q/N}(x, k_\perp)$  it was trained with. However, no particular definition of  $h(k_\perp)$  or  $f_{q/N}(x, k_\perp)$  is required to make the combination of  $\mathcal{N}_q^{\text{DNN}}(x)h(k_\perp)f_{q/N}(x, k_\perp)$  nonbiased and perform well in the fit. From this standpoint, we do not consider this to be a fully model-independent extraction, but every attempt is made to make it minimally model dependent while still preserving the original assumptions relevant to the phenomenology for transparency of the schema. The DNN treatment of  $\mathcal{N}_q(x)$  enables the flexibility of handling all the light-quark flavors  $q = \{u, \bar{u}, d, \bar{d}, s, \bar{s}\}$  independently. The method presented is intended to be somewhat analogous to the approach of treating  $\mathcal{N}_q(x)$  as a fit function to be parametrized for the sake of illustrating the advantages. It is natural to extend this study toward a symbolic regression of  $\mathcal{N}_q(x)$  or the full expression of the Sivvers function, but that is not the scope of the present work.

We use a relationship for  $\mathcal{N}_q(x)$ , similar to the seminal work [4], as a tool to generate our pseudodata for testing accuracy and reproducibility only. This definition is not used in any aspect of the final DNN fit results. The generic feedforward DNN structure for  $\mathcal{N}_q(x)$  that we use in this work is represented in Fig. 3. As we consider the

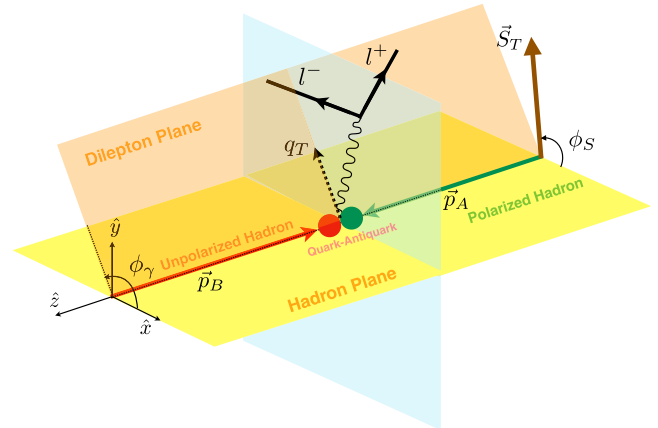


FIG. 2. Kinematics of the DY process in the hadronic center-of-mass frame.



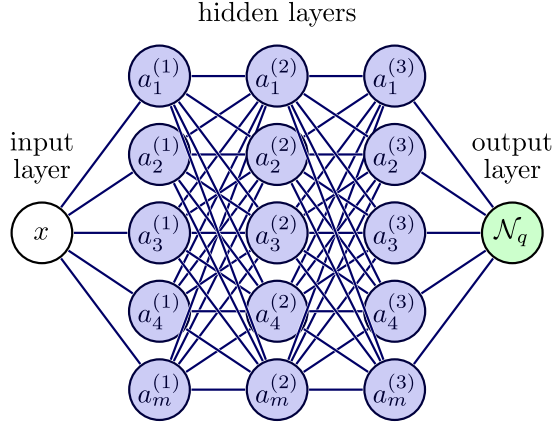


FIG. 3. A generic representation of the DNN architecture for  $\mathcal{N}_q(x)$ , where  $q = \{u, d, s, \bar{u}, \bar{d}, \bar{s}\}$ , and  $a_m^{(n)}$  represent the node  $m$  in the hidden layer  $n$ . The figure represents only up to  $n = 3$  for demonstration purposes.

$SU(3)_{\text{flavor}}$  symmetry breaking in QCD, we have  $\mathcal{N}_u(x)$ ,  $\mathcal{N}_{\bar{u}}(x)$ ,  $\mathcal{N}_d(x)$ ,  $\mathcal{N}_{\bar{d}}(x)$ ,  $\mathcal{N}_s(x)$ , and  $\mathcal{N}_{\bar{s}}(x)$  to handle the six light-quark flavors independently. Bjorken  $x$  is the only input as the initial layer of each  $\mathcal{N}_q(x)$ , and the final layer is a single-node output. The  $m_1$  in  $\mathcal{A}_0(z, phT, m_1)$  as defined in Eq. (9) is treated as a free parameter, with the initialization obtained from our first chi-square minimization fit (discussed in Sec. IV B), and then allowed to vary throughout the DNN training process with SIDIS data, as shown in Fig. 4. The DNN model results are then used to infer the projections for both SIDIS kinematics and DY kinematics (see Fig. 5).

There are many open-source software libraries for machine learning and artificial intelligence. For this particular extraction, we use TensorFlow [54]. A deep feedforward architecture is used with the hidden layers expanded with multiple numbers of nodes which are initialized randomly with Gaussian sampling of weights around *zero* with a standard deviation of 0.1. The degree of potential nonlinearity is introduced into the network by the choice of the activation function. The selection of the activation function can have a substantial impact on DNN performance and training dynamics. We chose the *ReLU6* activation function. This activation is a variant of the

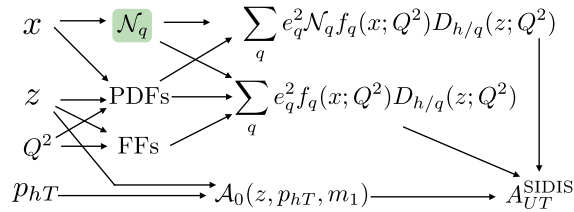


FIG. 4. The block diagram for SIDIS Sivers asymmetry [see Eq. (8)]:  $\mathcal{N}_q$  denote the DNN models for each quark flavor  $q = \{u, d, s, \bar{u}, \bar{d}, \bar{s}\}$ , and  $h = \{\pi^+, \pi^-, \pi^0, K^+, K^-\}$ .

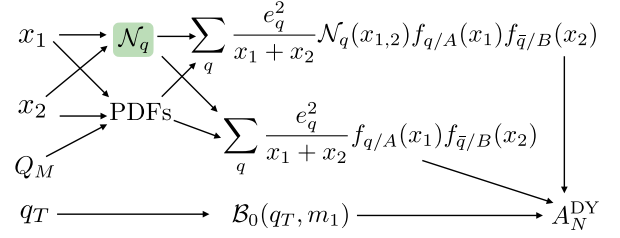


FIG. 5. The block diagram for DY Sivers asymmetries [see Eq. (22)].  $\mathcal{N}_q$  denotes the DNN models for each quark flavor  $q = \{u, d, s, \bar{u}, \bar{d}, \bar{s}\}$  obtained by training to SIDIS Sivers asymmetries.

rectified linear unit (*ReLU*) function. The *ReLU6* activation function has been shown to empirically perform better under low-precision conditions by encouraging the model to learn sparse features earlier, which is beneficial for learning complex patterns and relationships from the experimental data. We also use *least absolute shrinkage and selection operator regression* which is a regularization technique used to prevent overfitting and improve the model's performance and generalization ability, while also encouraging sparsity and feature selection. We also use L1 regularization. L1 regularization encourages sparsity in the activation by adding a penalty term to the loss function that is proportional to the absolute value of the weights [55]. By adding this regularization term, the most important inputs are weighted the most, so that noisy or redundant information is discarded. The strength of the regularization is controlled by the magnitude of the regularization coefficient, which is set to  $10^{-12}$ . Additionally, we use a dynamically decreasing learning rate. The learning rate is automatically reduced by 10% if the *training loss* has not decreased within the last 200 epochs<sup>2</sup> (i.e. *patience* = 200). The optimizer used was *Adam* while the loss function used was *mean squared error*. During the hyperparameter optimization process, there were slight deviations in the number of layers, nodes per layer, initial learning rate, batch size, and the number of epochs, but the basics of the scheme just described remain consistent for all DNNs used.

Data shuffling is used to randomly organize the data into batches that pass through the training process. When training, it is crucial to expose the DNN to a diverse range of training samples. Shuffling the training data ensures that the DNN does not learn patterns based on the sequential order of data. Instead, it promotes a more randomized and representative distribution of the data, preventing the model from becoming biased toward any specific subset or order. This randomization helps the model generalize better and makes it more robust to variations in the input data, ultimately improving its overall performance and accuracy. Batch size is a critical component of successful training as

<sup>2</sup>An epoch is a complete cycle of the passing of training data through the algorithm.

well. A batch size too large for a small dataset, even well-shuffled, can also lead to a biased model. In what follows, we train two dedicated DNN models: one for the SIDIS data with a polarized proton target and one for the data with a polarized deuteron (neutron) target. Since the number of available data points for the deuteron target is significantly smaller than the amount of data for the proton target, care is taken to find the optimum configuration of shuffling and batch size for the case of deuteron whereas for the proton the shuffling and batch size's impact are nearly negligible.

Our strategy is to first perform an exercise using only pseudodata to verify the extraction method that will ultimately be used on the real experimental data. First, we devise a *generating function* for the SIDIS Siverson asymmetry data using a conventional  $\chi^2$ -minimization routine (MINUIT in this case), without following the popular assumption of “unbroken sea” [4] in order to generalize the treatment of quarks and antiquarks. We perform a series of conventional MINUIT fits stepwise to obtain the final 19 parameters for the case of broken  $SU(3)_{\text{flavor}}$  symmetry in QCD. Then, we produce pseudodata (or replicas) for the SIDIS asymmetry by sampling from the mock experimental errors using the *generating function* with kinematics and binning in  $x$ ,  $z$  and  $p_{hT}$  as in the experimental data. Then a DNN model is constructed with all hyperparameters tuned in order to achieve the highest possible accuracy and precision. Here our nomenclature becomes quite specific, and we refer to the resulting distribution of DNN fits as a *DNN model*. The first model obtained with the method for a particular set of data is referred to as the *First Iteration*. At this stage, we use the distribution of fit results to obtain the mean and the error band from the initial DNN model to reparametrize the *generating function* so that it produces more realistic pseudodata. The DNN fits are performed again improving the quality (both accuracy and precision) of the resulting fits to result in a *Second Iteration* DNN model. One can repeat the number of iterations until the resulting model is no longer improving within the experimental uncertainties. In this way, the DNN model approaches the best approximation of the Siverson functions in comparison to the *true* values put into the *generating function*.

After confirming that the method works well, an extraction of the Siverson function using the SIDIS experimental data is performed. We treat the data for a polarized proton target and deuteron target separately for two reasons. First, fitting these together would introduce bias that would need to be managed directly. This is the case even assuming isospin symmetry in the  $u$  and  $d$  quarks' Siverson functions. Second, our approach leaves open the possibility to explore the nuclear dependence of the Siverson functions. The construction of the DNN models for proton and deuteron is analogous. To perform the fit another *First Iteration*, as previously described, is performed by developing and tuning a DNN model using the data from the real

experiment rather than pseudodata from the generating function. In the proceeding iterations, we use the generating function in order to tune the hyperparameters to achieve the highest possible quality of fit in comparison with the results from the *First Iteration*. Once a tuned model is obtained, we perform an extended study for evaluating the algorithmic uncertainty<sup>3</sup> as well as the systematic uncertainty of the DNN extraction method.

To elaborate on the pedagogy of this method, we organize the remainder of this section into the following subsections: (IV A) Data selection, (IV B) MINUIT fits for the case of  $SU(3)_{\text{flavor}}$ , (IV C) DNN model training with pseudodata, and (IV D) DNN model training with real experimental data.

### A. Data selection

Generally, TMD factorization is considered valid only under the kinematics restriction of  $\Lambda \ll Q$ , and  $p_{hT} \ll zQ$ . The first of these requirements is important because it is only in this limit one can apply massless-hadron kinematics resulting in near-negligible contamination from power corrections where  $Q > 2$  GeV is common. The second of these requirements provides the kinematic range in which a traditional factorization scheme will work. Additionally, the transverse mass scale should generally be preserved around  $(m/Q/z)^2$ . We do not strictly adhere to any of these guidelines but instead, study the consequence of using progressively less restrictive cuts. For global data selection, we focus our attention on the fixed target SIDIS and DY data. For the proton DNN fits, kinematically 3D binned data from HERMES2020 [56] are left out for validation and only 1D binned HERMES2020 data are used for the training. For the neutron DNN fits, the polarized  $^3\text{He}$  data from Jefferson Lab [57,58] are left out and used to test the new projections of the DNN model trained on the deuteron COMPASS data only.

Table I summarizes the kinematic coverage, the number of data points, and reaction types of the datasets that are considered in this work. In addition to the SIDIS datasets that are used in the fits, the polarized DY dataset from the COMPASS experiment is also listed in Table I as we demonstrate the predictive capability of the DNN model by comparing the projections with the real data points. The DY projections are made using the trained SIDIS DNN model assuming a sign change expected from *conditional* universality. For the case of training the DNN model related to the proton target we use HERMES2009 [59], COMPASS2015 [60] and HERMES2020 [56] data points associated with 1D kinematic binning, leaving the HERMES2020 [56] data associated with the 3D kinematic binning to compare with the projections from the trained model. The COMPASS2009 [52] dataset with a polarized-deuteron

<sup>3</sup>Algorithmic uncertainty is the degree of increase in the distribution of the resulting fits that is not directly from propagated experimental error.

TABLE I. The SIDIS and DY datasets considered in the fits include the DY data, which are used to demonstrate the predictive capability of the DNN model. Specifically, we make projections using the trained SIDIS DNN model, assuming a sign change to predict the real experimental DY data points. For the HERMES2020 dataset, data are available with both 1D and 3D kinematic bins. The 3D bin numbers are indicated in bold font.

Dataset	Kinematic coverage	Reaction	Data points
HERMES2009 (SIDIS) [59]	$0.023 < x < 0.4$	$p^\uparrow + \gamma^* \rightarrow \pi^+$	21
	$0.2 < z < 0.7$	$p^\uparrow + \gamma^* \rightarrow \pi^-$	21
	$0.1 < p_{hT} < 0.9$	$p^\uparrow + \gamma^* \rightarrow \pi^0$	21
	$Q^2 > 1 \text{ GeV}^2$	$p^\uparrow + \gamma^* \rightarrow K^+$	21
		$p^\uparrow + \gamma^* \rightarrow K^-$	21
HERMES2020 (SIDIS) [56]	$0.023 < x < 0.6$	$p^\uparrow + \gamma^* \rightarrow \pi^+$	27, <b>64</b>
	$0.2 < z < 0.7$	$p^\uparrow + \gamma^* \rightarrow \pi^-$	27, <b>64</b>
	$0.1 < p_{hT} < 0.9$	$p^\uparrow + \gamma^* \rightarrow \pi^0$	27
	$Q^2 > 1 \text{ GeV}^2$	$p^\uparrow + \gamma^* \rightarrow K^+$	27, <b>64</b>
		$p^\uparrow + \gamma^* \rightarrow K^-$	27, <b>64</b>
COMPASS2015 (SIDIS) [60]	$0.006 < x < 0.28$	$p^\uparrow + \gamma^* \rightarrow \pi^+$	26
	$0.2 < z < 0.8$	$p^\uparrow + \gamma^* \rightarrow \pi^-$	26
	$0.15 < p_{hT} < 1.5$	$p^\uparrow + \gamma^* \rightarrow K^+$	26
	$Q^2 > 1 \text{ GeV}^2$	$p^\uparrow + \gamma^* \rightarrow K^-$	26
COMPASS2009 (SIDIS) [52]	$0.006 < x < 0.28$	$d^\uparrow + \gamma^* \rightarrow \pi^+$	26
	$0.2 < z < 0.8$	$d^\uparrow + \gamma^* \rightarrow \pi^-$	26
	$0.15 < p_{hT} < 1.5$	$d^\uparrow + \gamma^* \rightarrow K^+$	26
	$Q^2 > 1 \text{ GeV}^2$	$d^\uparrow + \gamma^* \rightarrow K^-$	26
JLAB 2011, 2014 (SIDIS) [57,58]	$0.156 < x < 0.396$	${}^3\text{He}^\uparrow + \gamma^* \rightarrow \pi^+$	4
	$0.50 < z < 0.58$	${}^3\text{He}^\uparrow + \gamma^* \rightarrow \pi^-$	4
	$0.24 < p_{hT} < 0.43$	${}^3\text{He}^\uparrow + \gamma^* \rightarrow K^+$	4
	$1.3 < Q^2 < 2.7$	${}^3\text{He}^\uparrow + \gamma^* \rightarrow K^-$	1
COMPASS2017 (DY) [53]	$0.1 < x_N < 0.25$	$p^\uparrow + \pi^- \rightarrow l^+ l^- X$	15
	$0.3 < x_\pi < 0.7$		
	$4.3 < Q_M < 8.5$		
	$0.6 < q_T < 1.9$		

target is used for the neutron Sivers extraction as a separate DNN model.

For the initial  $\chi^2$ -minimization fit with MINUIT the same datasets are used as in [4] for consistency which is HERMES2009 [59], COMPASS2009 [52], and COMPASS2015 [60]. This fit is described in the next subsection. The error in Fit 1 of Table II is quite large, and many of the parameters are consistent with zero. This highlights the challenges of reproducing this type of analytical fit with a standard  $\chi^2$  minimization fit. The last column, Fit 5, is a fit from our generating function to the proton-DNN model which provides the most accurate parametrization of  $\mathcal{N}_q$  to date. Again, we point out we do not use this parametrization for any physics extraction but only to generate pseudodata.

### B. MINUIT fits for $SU(3)_{\text{flavor}}$

The analysis begins with a  $\chi^2$ -minimization fit with MINUIT similar to the approach in [4] except we expand

the number of parameters to treat each of the light-quark flavors separately. The results of the MINUIT fits are shown in Table II. Fit 1 is from the original fit results from Anselmino *et al.* directly from [4]. Here the  $\mathcal{N}_q(x)$  for the  $u$  and  $d$  quark used is Eq. (27) but  $\mathcal{N}_{\bar{q}}(x) = N_{\bar{q}}$  for antiquarks. In this fit, there are three parameters,  $\alpha_q$  and  $\beta_q$  and  $N_q$  for each quark flavor, and for each antiquark it is just  $N_{\bar{q}}$  plus  $m_1$ . This results in a nine-parameter fit. Fit 2 is a test to reproduce the same parametrization as in Fit 1. We note that in Fit 2 none of the nine parameters are fixed or have bounds imposed. Both of these first columns only consider  $u$  and  $d$  quarks and antiquarks. The Fit 1 parameters were used as the initial values to perform Fit 2. The difference in these two sets of fit parameters demonstrates the challenge of systematic consistency with this method though some parameters match reasonably well. For Fit 3 we use the same convention but add in the strange quark so there is an additional four parameters,  $N_s$ ,  $\alpha_s$ ,  $\beta_s$ , and  $N_{\bar{s}}$ , which leads to a 13-parameter fit. In order to initialize the 13 parameters in

TABLE II. Collection of MINUIT fit results. Fit 1 is from Anselmino *et al.* [4], Fit 2: refit as similar to [4], Fit 3: fit results including strange quarks, Fit 4: fit results with the same treatment for all three light-quark flavors.

Parameter	Fit 1	Fit 2	Fit 3	Fit 4	Fit 5
$m_1$	0.8	3.9	7.0	7	3.63
$\delta m_1$	$\pm 0.9$	$\pm 0.3$	$\pm 0.6$	$\pm 4$	$\pm 0.03$
$N_u$	0.18	0.48	0.89	0.89	1.08
$\delta N_u$	$\pm 0.04$	$\pm 0.03$	$\pm 0.05$	$\pm 0.06$	$\pm 0.02$
$\alpha_u$	1.0	2.41	2.78	2.75	2.85
$\delta \alpha_u$	$\pm 0.6$	$\pm 0.16$	$\pm 0.17$	$\pm 0.11$	$\pm 0.02$
$\beta_u$	6.6	15.0	19.4	20	12.4
$\delta \beta_u$	$\pm 5.2$	$\pm 1.4$	$\pm 1.6$	$\pm 2$	$\pm 0.1$
$N_{\bar{u}}$	-0.01	-0.032	-0.07	-0.12	11.8
$\delta N_{\bar{u}}$	$\pm 0.03$	$\pm 0.017$	$\pm 0.06$	$\pm 0.60$	$\pm 0.3$
$\alpha_{\bar{u}}$	...	...	...	0.4	1.90
$\delta \alpha_{\bar{u}}$	...	...	...	$\pm 0.5$	$\pm 0.03$
$\beta_{\bar{u}}$	...	...	...	20	1.28
$\delta \beta_{\bar{u}}$	...	...	...	$\pm 16$	$\pm 0.07$
$N_d$	-0.52	-1.25	-2.33	-2.4	-2.81
$\delta N_d$	$\pm 0.20$	$\pm 0.19$	$\pm 0.31$	$\pm 0.4$	$\pm 0.07$
$\alpha_d$	1.9	1.5	2.5	2.7	1.21
$\delta \alpha_d$	$\pm 1.5$	$\pm 0.4$	$\pm 0.4$	$\pm 0.6$	$\pm 0.02$
$\beta_d$	10	7.0	15.8	17	3.58
$\delta \beta_d$	$\pm 11$	$\pm 2.6$	$\pm 3.2$	$\pm 4$	$\pm 0.11$
$N_{\bar{d}}$	-0.06	-0.05	-0.29	-0.7	-32.9
$\delta N_{\bar{d}}$	$\pm 0.06$	$\pm 0.11$	$\pm 0.27$	$\pm 0.5$	$\pm 0.7$
$\alpha_{\bar{d}}$	...	...	...	1.5	5.2
$\delta \alpha_{\bar{d}}$	...	...	...	$\pm 0.6$	$\pm 0.7$
$\beta_{\bar{d}}$	...	...	...	20	12.9
$\delta \beta_{\bar{d}}$	...	...	...	$\pm 17$	$\pm 0.3$
$N_s$	...	...	-14	-20	-15.2
$\delta N_s$	...	...	$\pm 10$	$\pm 40$	$\pm 0.8$
$\alpha_s$	...	...	4.9	4.7	2.9
$\delta \alpha_s$	...	...	$\pm 3.3$	$\pm 3.0$	$\pm 0.1$
$\beta_s$	...	...	3	2.3	3.2
$\delta \beta_s$	...	...	$\pm 4$	$\pm 3.1$	$\pm 0.3$
$N_{\bar{s}}$	...	...	-0.1	20	-19.90
$\delta N_{\bar{s}}$	...	...	$\pm 0.2$	$\pm 5$	$\pm 0.09$
$\alpha_{\bar{s}}$	...	...	...	9.5	7.35
$\delta \alpha_{\bar{s}}$	...	...	...	$\pm 1.4$	$\pm 0.05$
$\beta_{\bar{s}}$	...	...	...	20	19.80
$\delta \beta_{\bar{s}}$	...	...	...	$\pm 14$	$\pm 0.02$
$\chi^2/N_{\text{data}}$	1.29	1.59	1.69	1.66	...

Fit 3, we use the corresponding values for those parameters from Fit 2 and zeros for the rest. Fit 4 uses Eq. (27) for both quarks and antiquarks so that the treatment of all three light-quark flavors is the same. In addition to the parameters from Fit 3, Fit 4 contains six more parameters for the antiquarks. The result of Fit 4 leads to a larger  $N_{\bar{q}}$  value to compensate for the fact that  $\alpha_{\bar{s}}$  and  $\beta_{\bar{s}}$  are now present in the fit. However, the motivation behind performing Fit 4 in this way is to generalize the  $\mathcal{N}_q(x)$  in a flavor-independent fashion for both quarks and antiquarks. Fit 4 is the final fit that we will use to generate pseudodata for testing the DNN fits and for calculating the model's accuracy. The last column, Fit 5, uses

the analytical generating function to fit the proton-DNN model in fine bins. The error of the model is propagated to each bin, and a MINUIT fit is performed providing a highly accurate parametrization of  $N_q$  represented in this particular function form. We provide this as a basis for comparison.

### C. DNN method testing

We develop a systematic method of constructing, optimizing, and testing the DNN fits by using pseudodata to ensure a quality extraction from the experimental data. Our approach uses a combination of Monte Carlo sampling and synthetic data generation. The pseudodata points are randomly generated by sampling within multi-Gaussian distributions centered around each experimental data point, with variance given by the experimental uncertainty. Many pseudodata DNN fits (instances) are performed together to obtain the uncertainty of the resulting DNN model (mean and distribution). The general approach is to use existing experimental data to parametrize a fit function and then use it to generate new synthetic data (replicas) with similar characteristics. The pseudodata are generated with a known Siverson function so that the extraction technique can be explicitly tested. An error bar is assigned to each new data point which is taken directly from the experimental uncertainties reported for the complete set of kinematic bins. This approach aims to produce pseudodata that simulate the experimental data as closely as possible with particular sensitivity to phase space. It does this so that the test metrics are also relevant for the real experimental extraction. To do this the pseudodata generator must be very well-tuned to the kinematic range of the experimental data. Hence, the *generating function* contains as much feature space information as possible. It is important to emphasize here that the metrics that we use to quantify the improvement in the *Second Iteration* compared with the *First Iteration* are sensitive to phase space. The *accuracy* (proximity of the mean of the DNN fits to the *true* Siverson) is defined as

$$\epsilon_q(x, k_{\perp}) = \left( 1 - \frac{|\Delta^N f_{q/p^{\uparrow}}^{(\text{true})} - \Delta^N f_{q/p^{\uparrow}}^{(\text{mean})}|}{\Delta^N f_{q/p^{\uparrow}}^{(\text{true})}} \right) \times 100\%, \quad (29)$$

and *precision* (the standard deviation of replicas) as

$$\sigma_q(x, k_{\perp}) = \sqrt{\frac{\sum_i (\Delta^N f_{q/p^{\uparrow}}^{(i)} - \Delta^N f_{q/p^{\uparrow}}^{(\text{mean})})^2}{N}}. \quad (30)$$

The *generating function* used to produce the *true* value of the Siverson is improved in the process of optimizing the DNN hyperparameters. This approach improves the *generating function* and the DNN fit with each iteration. As a result, more realistic data can be generated in each iteration, which enables better hyperparameter optimization and

testing for the experimental data in the subsequent iteration. Note that experimental data still refers to pseudodata replicas that are generated using the real experimental data rather than the *generating function*.

In the pseudodata test, the same number of replicas are used in the *First Iteration* and in the *Second Iteration*. The number of replicas should be kept the same across consecutive iterations to control statistical error variation from the replicas. To most accurately propagate the experimental uncertainty using the replica approach requires a sufficient number of replicas so that only negligible statistical error from the replicas is added. For the present study, our pseudodata uncertainty is simplified and is represented by a single error bar which contains the experimental statistical error and systematic error provided by the data publication.<sup>4</sup>

These are the steps to perform the DNN *method testing* with pseudodata using the *generating function*:

- (1) Generate pseudodata for the SIDIS Sivers asymmetry using the *generating function* (The initial function used is from Fit 4 in Table II).
- (2) *First Iteration*: construct a DNN fit and tune its hyperparameters by training with the pseudodata from Step 1. Use 10% of the pseudodata for the validation in each epoch.
- (3) Determine the optimum number of epochs by analyzing the training and validation losses for each replica in Step 2.
- (4) Improve the *generating function*:
  - (a) Perform an intermediate DNN fit to the experimental data. This fit is performed using the optimized hyperparameters from the above Steps 2 and 3.
  - (b) Use the trained DNN fit from (a) to infer the asymmetry over the 3D kinematics  $(x, z, p_{hT})$  in fine bins.<sup>5</sup>
  - (c) Perform a MINUIT fit on the fine-binned data to obtain the new *generating function* parametrization.
- (5) Perform Step 1 again using the improved *generating function*.
- (6) *Second Iteration*: Perform Step 2 again with the pseudodata generated from Step 5.
- (7) Perform a comparison of the Sivers functions extracted at the *First Iteration* vs *Second Iteration* in terms of the *accuracy*, *precision* and the magnitude of the *loss function* at the final epoch.

<sup>4</sup>A more detailed analysis could be performed if a full covariance matrix from the experimental analysis was provided, but this type of information is often difficult to obtain after the experiment results are published.

<sup>5</sup>When generating fine-binned kinematics for the Sivers asymmetry data for a given kinematic variable, the averaged values are used for other variables including  $Q^2$  as the  $Q^2$  evolution is near negligible over the considered kinematics.

The accuracy metric is a critical part of the testing and is used to understand the resulting extraction methods' strengths and weaknesses. A truly robust method would be consistently able to make accurate predictions for any *generating function* parametrization over the relevant kinematic range. To verify the robustness of our extraction method, we test it with many artificial Sivers functions and their corresponding pseudodata at the final stage.

The hyperparameters' optimization process is a systematic process of trial and error with architecture changes like adding or subtracting layers and the number of nodes per layer. The initial learning rate varies as does the batch size. Care is taken to achieve the lowest training and validation loss. We monitor the stability of the loss by examining the magnitude and frequency of fluctuations. A more moderate and even trend downward is optimal. After the *generating function* is improved, in principle, it holds more information. The number of layers and nodes is increased to make a better fit to the new, more realistic, pseudodata in the *Second Iteration*. If there are too few hidden layers and nodes, it becomes apparent when the training loss does not decrease compared with previous fits. Too many additional layers and nodes result in instability in the training loss. Tuning continues until little improvement in accuracy and precision can be determined.

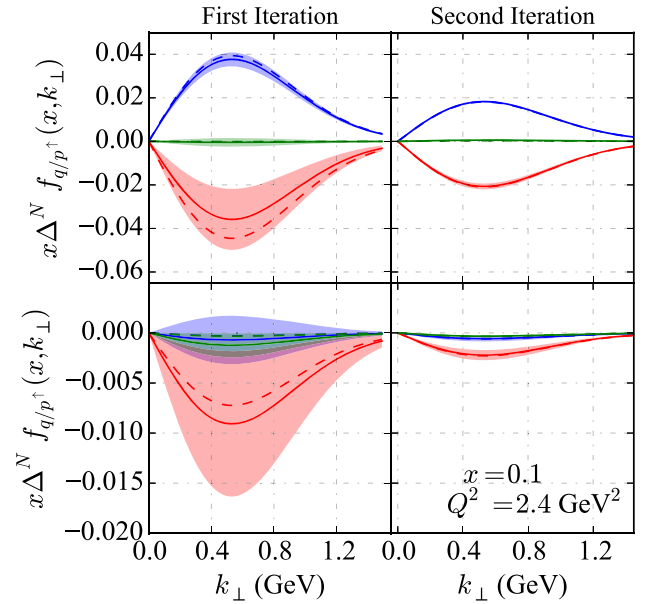


FIG. 6. A comparison of the DNN model before (left column) and after (right column) hypertuning with the improved *generating function*. In the *First Iteration*, the DNN model is based on Fit 4, whereas in the *Second Iteration*, the model is based on the hypertuned DNN fit. Additionally, hypertuning optimization leads to higher accuracy and precision with the same number of epochs. The dotted line in each case represents the generating function or the true value while the solid line represents the mean of DNN fits. In the top two plots,  $u$  is blue,  $s$  is green and  $d$  is red. Similarly for the antiquark in the lower two plots.

TABLE III. The summary of the optimized sets of hyperparameters: The indications in the table are  $C_0^i$  and  $C_0^f$  for results from the pseudodata from the generating function,  $C_p^i$ , and  $C_p^f$  for results from SIDIS data from experiments associated with the polarized-proton target, and  $C_d^i$  and  $C_d^f$  for results from SIDIS data from experiments associated with the polarized-deuterium target, where  $i$  and  $f$  indicate the *First Iteration* and *Second Iteration* respectively. The initial learning rate is also listed ( $\times 10^{-4}$ ) as is the final training loss ( $\times 10^{-3}$ ). The accuracy and precision in each case are the maxima over the phase space.

Hyperparameter	$C_0^i$	$C_0^f$	$C_p^i$	$C_p^f$	$C_d^i$	$C_d^f$
Hidden layers	5	7	5	7	5	8
Nodes/layer	256	256	550	550	256	256
Learning rate	1	0.125	5	1	10	1
Batch size	200	256	300	300	100	20
Number of epochs	1000	1000	300	300	200	200
Training loss	0.6	0.05	1.5	1	2	1
$\varepsilon_u^{\max}$	95.67	99.27	55.21	94.04	56.80	93.02
$\varepsilon_{\bar{u}}^{\max}$	42.62	98.09	52.57	96.70	34.83	91.40
$\varepsilon_d^{\max}$	80.46	98.89	55.69	93.13	52.44	89.27
$\varepsilon_{\bar{d}}^{\max}$	74.59	97.08	55.37	95.04	46.60	92.58
$\varepsilon_s^{\max}$	45.53	79.27	49.54	90.64	36.34	93.41
$\varepsilon_{\bar{s}}^{\max}$	59.27	91.13	33.89	82.51	65.57	91.45
$\sigma_u^{\max}$	3	0.1	5	2	2	0.4
$\sigma_{\bar{u}}^{\max}$	2	0.2	6	2	8	2
$\sigma_d^{\max}$	10	1	20	6	2	1
$\sigma_{\bar{d}}^{\max}$	7	4	20	8	7	1
$\sigma_s^{\max}$	2	0.2	4	1	6	2
$\sigma_{\bar{s}}^{\max}$	1	0.1	4	2	6	3

Figure 6 illustrates the results from the *First Iteration* (two plots on the left-hand side) and the *Second Iteration* (two plots on the right-hand side). The sets of hyperparameters in the *First Iteration* and the *Second Iteration* are given in the columns for each DNN model in Table III. The indications in the table are  $C_0^i$  and  $C_0^f$  for results from the pseudodata from the generating function,  $C_p^i$ , and  $C_p^f$  for results from SIDIS data from experiments associated with the polarized-proton target, and  $C_d^i$  and  $C_d^f$  for results from SIDIS data from experiments associated with the polarized-deuterium target. Here  $i$  and  $f$  indicate the *First Iteration* and *Second Iteration* respectively. The listed learning rate (multiplied by  $10^{-4}$ ) is the initial learning rate as a dynamically decreasing learning rate is used. (This is explained in Sec. IV). The accuracy  $\varepsilon_q(x, k_\perp)$  is defined in Eq. (29), and the results in this table correspond to the maximum deviation of the mean of the replicas from the true values  $\varepsilon_q^{\max}$ , whereas the precision  $\sigma_q(x, k_\perp)$  is defined in Eq. (30), and the results are the maximum standard deviations of the replicas  $\sigma_q^{\max}$  and are in the units of  $\times 10^{-3}$ .

It is worth noting that the improvement in both *accuracy*  $\varepsilon_q^{\max}$  and *precision*  $\sigma_q^{\max}$  can be observed from the closeness of the solid line (mean of the 1000 DNN replicas) to the dashed line (*generating function*) for quarks (upper plots) and antiquarks (lower plots). The improvement of the DNN model in each case is significant to the point where it is difficult to distinguish the solid line and the corresponding

dashed line, indicating a high degree of accuracy and precision. Also, we observed that the training loss in the *Second Iteration* is about 1 order of magnitude less than the one from the *First Iteration*.

#### D. DNN model from real data

In contrast to the testing of the DNN model with pseudodata from the *generating function*, we now describe the steps to apply the DNN fit method to real experimental data. The pseudodata test from Sec. IV C is using the combined proton and deuteron data as in all previous work on global fits of the Siverson function. In the following extraction with real experimental data, the proton and deuteron data are fitted separately. To take full advantage of the information provided by the model testing in the previous section, the steps from Sec. IV C are performed again separately for proton and deuteron data. The starting hyperparameters for the first DNN fit in this section including the architecture, initial learning rate, batch size, as well as optimal number of epochs, are all determined based on the best accuracy and precision using the well-tuned pseudodata from the generating function first in each case. They provide more information on the initial hyperparameter so that even the *First Iteration* is a quality fit. In the following steps, two distinct DNN models are developed, one for the *proton* quarks Siverson asymmetry and one for *neutron* quark Siverson asymmetry. The following procedure is common to both.

- (1) *First Iteration*: Construct a DNN fit and tune its hyperparameters by training with the experimental data. Use 10% of the data for the validation in each epoch.<sup>6</sup>
- (2) Identify the optimum number of epochs when the validation loss exceeds the training loss [see Fig. 7 as an example].
- (3) Perform a DNN fit, with the optimized hyperparameters from Step 1 and the number of epochs determined from Step 2, using all the data without leaving any for validation.
- (4) Improve the *generating function*:
  - (a) Use the tuned DNN model in Step 3 to infer the asymmetry over the 3D kinematics  $(x, z, p_{hT})$  in fine bins.
  - (b) Perform a MINUIT fit to obtain the new *generating function*.
  - (c) Produce pseudodata for the SIDIS asymmetry using the *generating function* in the previous step.
- (5) Perform Step 1 and Step 2 with the pseudodata from Step 4 (from the improved *generating function*).
- (6) *Second Iteration*: perform a DNN fit with the optimized hyperparameters from Step 5 using experimental data without leaving any data for validation.
- (7) Perform a comparison of the Sivers functions extracted at the *First Iteration* vs *Second Iteration* in terms of the *accuracy*, *precision* and the magnitude of the *loss function* at the final epoch. See Fig. 8 for the qualitative improvement of the *Second Iteration* compared with the *First Iteration*.

Although the architectural specifics such as the number of hidden layers, nodes per layer, and learning rate may be modified in the hyperparameter optimization step, the number of epochs and the number of replicas remain the same. The overall feedforward architecture structure remains consistent as well for simplicity. Table III shows the optimized set of hyperparameter configurations are represented with  $C_{\{0,p,d\}}^{i,f}$  where  $i, f$  represent the *First Iteration* and the *Second Iteration*, and  $\{0, p, d\}$  represent with *pseudodata*, with *proton* data and with *deuteron* data respectively.

Clearly, the pseudodata from the *generating function* still play a critical role in the tuning and testing of the fit of the real experimental data. Accuracy is necessarily determined using pseudodata so that the mean of the final DNN model can be compared directly to the *true* Sivers asymmetry. This procedure for determining accuracy assumes that the *true* Sivers asymmetry from the experiment can be approximated by the well-tuned *generating function* after the final

<sup>6</sup>For this section we list this as *First Iteration* though the real experimental data have already been fit to improve the *generating function* for the pseudodata test. With every subsequent iteration, the pseudodata test and the DNN fit to the experimental data improve.

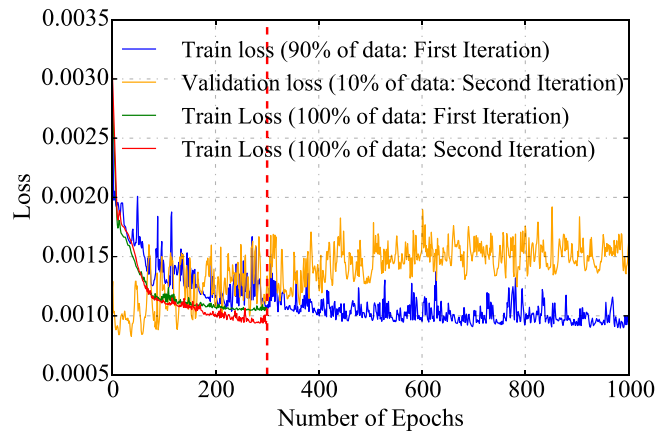


FIG. 7. An example for the determination of the optimum number of epochs (about 300 in this case, marked by the vertical red dashed line) at the *First Iteration* based on crossover coordinates between the train (blue curve) and the validation (orange curve) losses. Then, the train-loss behaviors until the optimum number of epochs with all data points at the *First Iteration* and at the *Second Iteration* are represented by the green curve and the red curve respectively.

iteration. There is a systematic error associated with analyzing accuracy this way, but this type of error can be estimated.

After completing the extraction method described above, we perform a systematic uncertainty assessment on the overall method of extraction. To test the reliability of the extraction, we adjust the parameters of the *generating function* and repeat the extraction process by again following the full set of steps for several variations of pseudodata. By using the Sivers asymmetry generated from the optimized *generating function*, the absolute differences between the

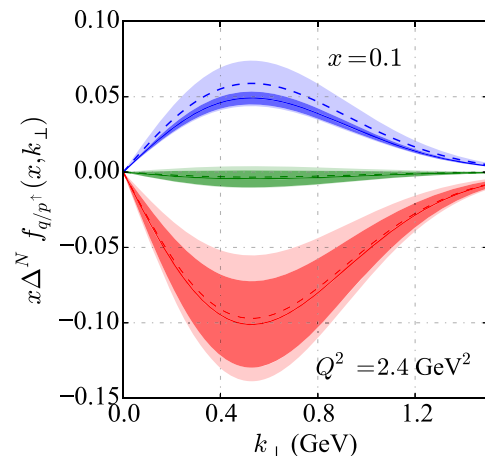


FIG. 8. The qualitative improvement of the extracted Sivers functions for  $u$  (blue),  $d$  (red), and  $s$  (green) quarks at  $x = 0.1$  and  $Q^2 = 2.4 \text{ GeV}^2$  using the optimized *proton*-DNN model at the *Second Iteration* (solid lines with dark-colored error bands with 68% C.L.), compared with the *First Iteration* (dashed lines with light-colored error bands with 68% C.L.).

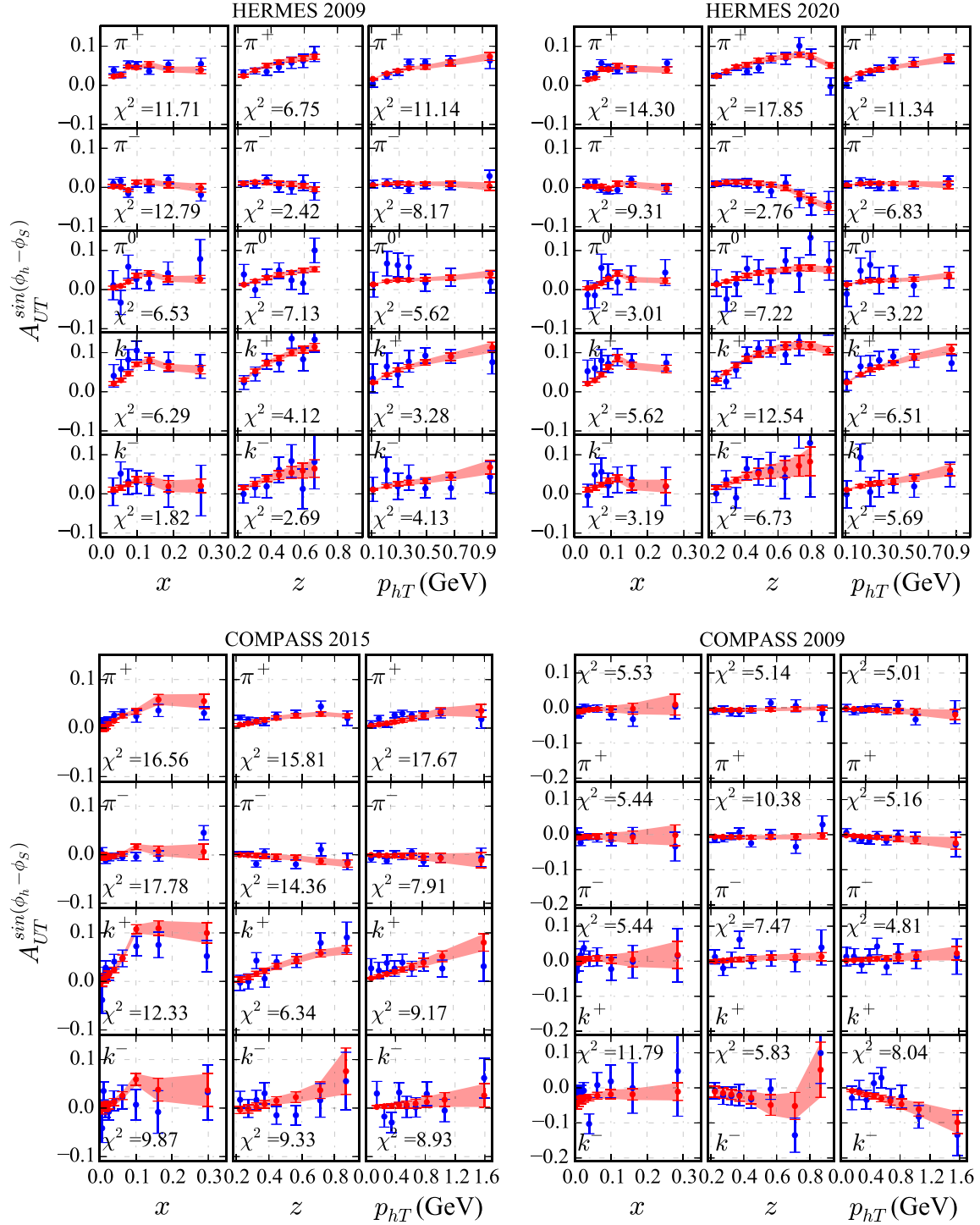


FIG. 9. The DNN fit results of the SIDIS Sivers asymmetries (red) accompanied by 68% C.L. error bands in comparison with the actual data (blue). The *proton*-DNN model is trained with HERMES2009, HERMES2020, and COMPASS2015, whereas the *deuteron*-DNN model is trained with the COMPASS2009 data. The calculated partial  $\chi^2$  values are provided as quantitative assessments for all kinematic bins.



mean of the DNN model and the true value over  $k_{\perp}$  were used to estimate the systematic uncertainties of the final DNN model from this extraction technique.

## V. RESULTS

In this section, we present the results from two separate DNN models: the *proton*-DNN and the *deuteron*-DNN, along with their optimized hyperparameters. The number of parameters in the optimized proton-DNN model is 11 million, whereas the deuteron-DNN model is 2.8 million. Only SIDIS data were used to train the DNN models in this exploratory AI-based extraction technique. The optimized hyperparameter configurations are provided for both models in columns  $\mathcal{C}_p^{i(f)}$  and  $\mathcal{C}_d^{i(f)}$ , with the subscripts  $p$  and  $d$ , respectively, in Table III. To quantitatively represent the improvement made by performing the steps mentioned in the previous section, we present our accuracy and precision results in the lower part of Table III.

### A. DNN fit to SIDIS data

We now explore the results and compare some of our final fits and projections with those of other global fits. Note that the  $\chi^2$  values are calculated values for each kinematic bin based on Pearson's reduced  $\chi^2$  statistic, and indicated in our plots are calculated after the analysis is complete, rather than as a part of the minimization process.

The plots of SIDIS Sivers asymmetry data and our resulting DNN models (for *proton* and *deuteron*) are shown in Fig. 9. Each plot includes the partial  $\chi^2$  values for the particular  $x$ ,  $z$ , and  $p_{hT}$  bins for each hadron type. HERMES2009 [59] (*top left*), HERMES2020 [56] (*top right*), and COMPASS 2015 [60] (*bottom left*) are described with the *proton*-DNN model with root mean square error (RMSE) of 0.0225, whereas the COMPASS2009 [52] (*bottom right*) dataset is described with the *deuteron* DNN with RMSE of 0.0220. In comparison with [4], there are some improvements in describing the proton SIDIS data on  $\pi^{\pm}$  and  $K^+$  in HERMES 2009, which can be noticed quantitatively based on the partial  $\chi^2$  values from the DNN model. This indicates that the possible effects attributed to the TMD evolution [6,8,48] and assumed to be the cause of the larger  $\chi^2$  values in [4] for proton data on  $\pi^+$  may have been somewhat integrated into the DNN model. Although HERMES 2020 [56] reported SIDIS data in 1D kinematic bins as well as with 3D kinematic bins, in our fits we use the data in the form of 1D kinematic bins to be consistent with the rest of the datasets in our fits.

The *deuteron*-DNN model's description of COMPASS2009 [52] data is shown in the bottom left subfigure of Fig. 9. Without applying any cuts on the data, the DNN model yields a RMSE of 0.0220, covering the full range in  $x$ ,  $z$ , and  $p_{hT}$  kinematic projections from the

COMPASS2009 dataset. This is in contrast to the limited kinematic coverage considered in [5,7,9,51]; notably, the data points at  $p_{hT} > 1$  GeV are described somewhat better by the *deuteron*-DNN model compared with the fits in [4,26]. This suggests that performing dedicated fits to data specific to polarized nucleon targets enables better information extraction, which is true for both DNN and other fitting approaches. The advantage of DNNs in this case is to perform well even with limited data.

We did not include JLab [57,58] data in our *deuteron*-DNN model fits to use it as a projection test for the neutron Sivers asymmetry. Our projection indicates good agreement with the  $^3\text{He}$  data, but both the data and the projection are largely consistent with zero. It is also important to note that in this work we are *not* imposing any isospin symmetry condition ( $f_{1T}^{\perp u} = f_{1T}^{\perp d}$  and/or  $f_{1T}^{\perp \bar{u}} = f_{1T}^{\perp \bar{d}}$ ) for the SIDIS data with the deuteron target as was done in [7]. The successful construction of the two different proton and neutron Sivers functions may indicate that our DNN approach can be particularly useful for analyzing data from polarized nucleons in different nuclei, potentially opening up a new way of exploring the nuclear effects associated with TMDs.

### B. Sivers in momentum space

The extracted Sivers functions, including the systematic uncertainties from the DNN models at  $x=0.1$  and  $Q^2 = 2.4$  GeV, are shown in Fig. 10 represented by the *mean* with 68% C.L. *error bands*. The corresponding optimized hyperparameter configurations for the *proton*-DNN model and *deuteron*-DNN model are  $\mathcal{C}_p^f$  and  $\mathcal{C}_d^f$ , respectively, as given in Table III. The Sivers functions extracted using the *deuteron*-DNN model show consistency with zero, considering the accompanying systematic uncertainties. However, this is still a significant result, given the limitation in statistics from the SIDIS data with a deuteron target. The Sivers functions extracted using the *proton*-DNN model have small systematic uncertainties. Note that we use  $\Delta^N f_{q/p\uparrow}(x, k_{\perp})$  notation, as in [4], to represent the Sivers functions in our plots, and one can use Eq. (3) to convert to  $f_{1T}^{\perp q}(x, k_{\perp})$  notation.

Comparing the extracted Sivers functions in  $k_{\perp}$  space to other extractions in the literature can also be useful, although we have not included those curves in our plots. In summary, the *proton*-DNN model extractions are relatively precise, with narrower error bands compared with those in [3,4,7–9,11,26].

### C. Sivers first transverse moment

The first transverse moment of the Sivers functions can be obtained through  $d^2k_{\perp}$  integration of the Sivers functions [4],



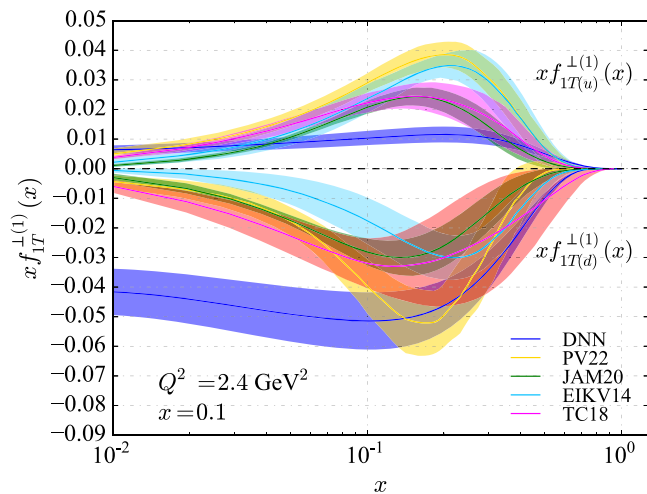


FIG. 12. The extracted Sivers functions for valence  $u(d)$  quarks from the *proton*-DNN model represented in the upper (lower) half of the figure, with the results from PV22 [51], JAM20 [5], EIKV [8], TC18 [61].

which was also a similar observation from a theoretical calculation based on  $SU(2)$  chiral Lagrangian [63] and the predictions at large- $N_c$  limit of QCD [62]. The central values extracted in [27] are qualitatively similar to the features seen in Fig. 11 which are small but nonzero within the uncertainties. Additionally, the corresponding central values extracted in [4] are both negative but consistent with zero.

The first transverse moments  $x f_{1T}^{\perp(1)q}(x)$ , in the case of  $SU(3)_{\text{flavor}}$ , from our DNN result, are more precise (narrower error bands) than those in [4,23,26]. However, the error bands are slightly larger than those in JAM20 [5], which includes more data from SIDIS, DY and SIA,  $pp$  collisions, and parametrizations for Sivers, Collins, and transversity TMDs together.

## D. Projections

### 1. SIDIS Projections

In Fig. 13, we compare the SIDIS Sivers asymmetries (in red) projected onto the HERMES2020 3D kinematic bins with the experiment measurements (in blue). These results are obtained using our *proton*-DNN model and are accompanied by 68% C.L. error bands. We also provide calculated partial  $\chi^2$  values for each kinematic bin as a quantitative assessment. Unlike in [7], we have made projections for all the data points since we have not applied any data cuts. There are relatively larger partial  $\chi^2/N_{pt}$  (as was also observed in [7]), but only in a couple of  $K^+$  and  $K^-$  bins. In Fig. 14, we present the projected SIDIS Sivers asymmetries for the JLab kinematics [57,58], obtained using our *deuteron*-DNN model. The figure includes 68% C.L. error bands and a comparison with the JLab

neutron Sivers asymmetry data. These results are consistent with those reported in [4,8,9,61].

### 2. DY Projections

The resulting DNN model based on the SIDIS Sivers asymmetries is capable of projecting the Sivers asymmetries in DY experiments which could be sensitive to either valence quarks or sea quarks depending on the relevant kinematic coverage. For example, the COMPASS2017-polarized DY Sivers asymmetry measurements [53] are dominated by the valence quarks, and the upcoming SpinQuest (E1039) experiment's polarized DY asymmetry measurements [64] will be dominated by the sea quarks. For these DY projections, we follow the block diagram represented in Fig. 5, which includes the assumption of the *sign-change* of the Sivers function in DY relative to the SIDIS process mentioned in Eq. (1). Therefore, using the trained *proton*-DNN model, we make projections for the DY Sivers asymmetries for the COMPASS2017 experiment [53] with a proton target and a pion beam using CTEQ6L [36] and JAM21PionPDFnlo [65] for proton PDFs and pion PDFs respectively. Meanwhile, using both the *proton*-DNN model and *deuteron*-DNN model, we make predictions for the SpinQuest experiment.<sup>7</sup> The kinematic inputs are  $x_1$  (beam),  $x_2$  (target),  $x_F (= x_1 - x_2)$ ,  $q_T$  (transverse component of the virtual photon), and  $Q_M$  (dilepton invariant mass).

The projected DY Sivers asymmetries for the COMPASS2017 kinematics using the trained *proton*-DNN model in comparison with the data [53] points are represented in the Fig. 15. Although the projections are based on the assumption of conditional universality, it is worth noting that without this assumption, negative asymmetry projections were observed. However, for clarity, the projections without assuming conditional universality are not shown in Fig. 15. When comparing the projections of the *proton*-DNN model with the predictions from [4,8,49], it is evident that the mean of the *proton*-DNN model projection, in terms of  $x_F$ , is more consistent with the measured mean Sivers asymmetry in the experiment (refer to Fig. 6 in [53]). At the same time, the *proton*-DNN model projection has relatively smaller uncertainty. It is important to note that the predictions mentioned in the cited works were based on different  $Q^2$ -evolution schemes, while the *proton*-DNN model incorporates DGLAP evolution through LHAPDF [37]. In [7], only two data points from the COMPASS2017 [53] data were included in their fits, resulting in larger uncertainties in the projected asymmetry values for the remaining data points when compared with the projections generated by the *proton*-DNN model. The increasing trend of the projected DY Sivers asymmetries with respect to the  $q_T$  kinematic variable in [7] is consistent

<sup>7</sup>The kinematic bins from the SeaQuest experiment are used [66].

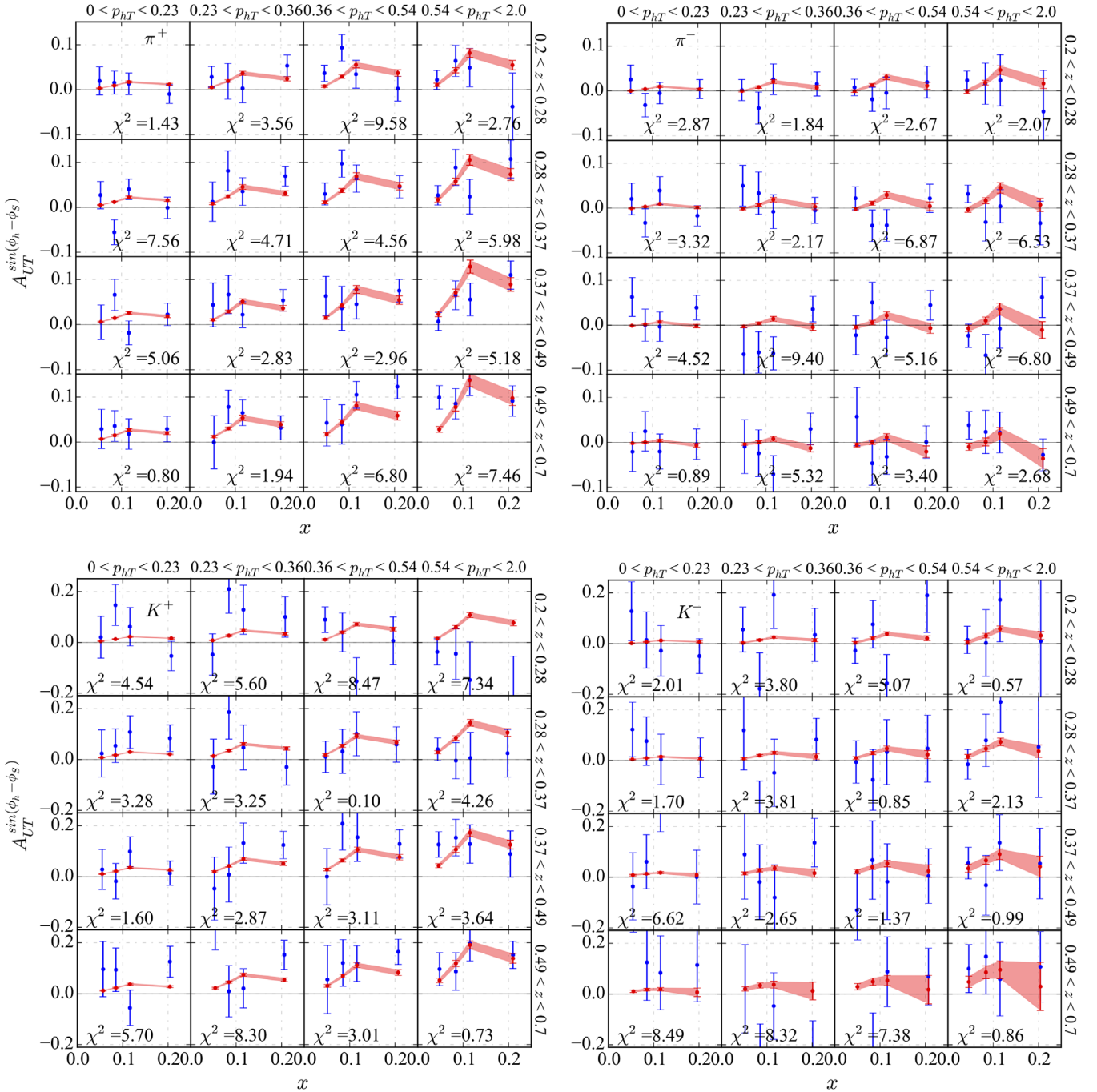


FIG. 13. Projections of the of HERMES 2020 data for 3D kinematic bins, using the *proton*-DNN model including 68% C.L. error bands (in red) in comparison with the actual data points (in blue).

with the projections presented in the middle-right plot of Fig. 15 generated by the *proton*-DNN model, while in [9], the corresponding trend exhibits a very small negative slope in relation to  $q_T$ .

A nonzero sea-quark Siverts asymmetry is inferring that the sea quarks have nonzero orbital angular momentum. The *proton*-DNN model predictions exhibit consistency with the nonzero Siverts asymmetry from the *sea* quarks, with higher precision compared with existing predictions

[4,8,49], for the SpinQuest kinematics [64,67]. Additionally, in this work, we report our projections for the polarized Drell-Yan Siverts asymmetries for a deuteron target at the SpinQuest experiment, as shown in Fig. 16 by the orange-colored bands. The mean of the deuteron DNN model in all kinematic projections  $x_1, x_2, x_F$ , and  $q_T$  are consistent with zero. The *proton*-DNN model predicts a positive slope with respect to  $q_T$  for a *proton* target as shown in the lower-right plot of Fig. 16. To date, with the

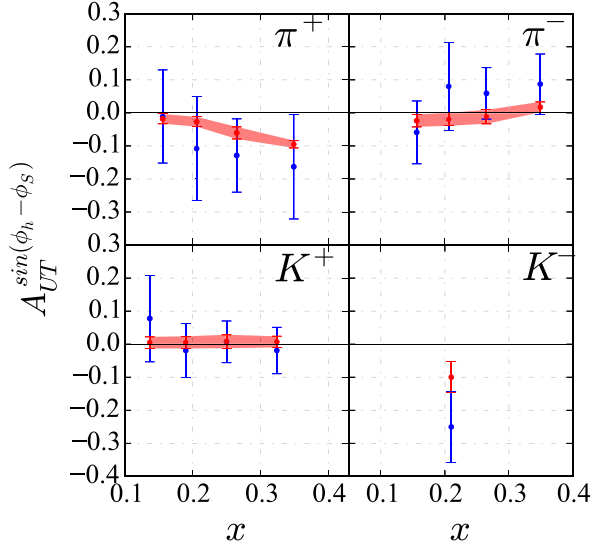


FIG. 14. The *deuteron*-DNN model projections (red) with 68% C.L. error bands, for JLab kinematics [57,58] in comparison with measured data (blue) without the systematic uncertainty.

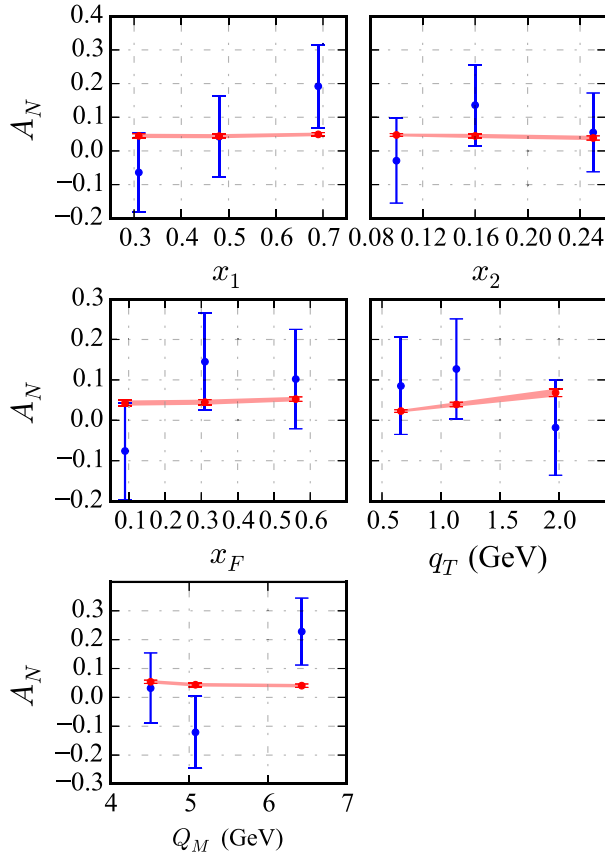


FIG. 15. The *proton*-DNN model's predictions (red) including 68% C.L. error bands, for Sivers asymmetries in  $x_1, x_2, x_F, q_T$ , and  $Q_M$  kinematic projections for COMPASS DY kinematics [53] in contrast with the measured data (blue).

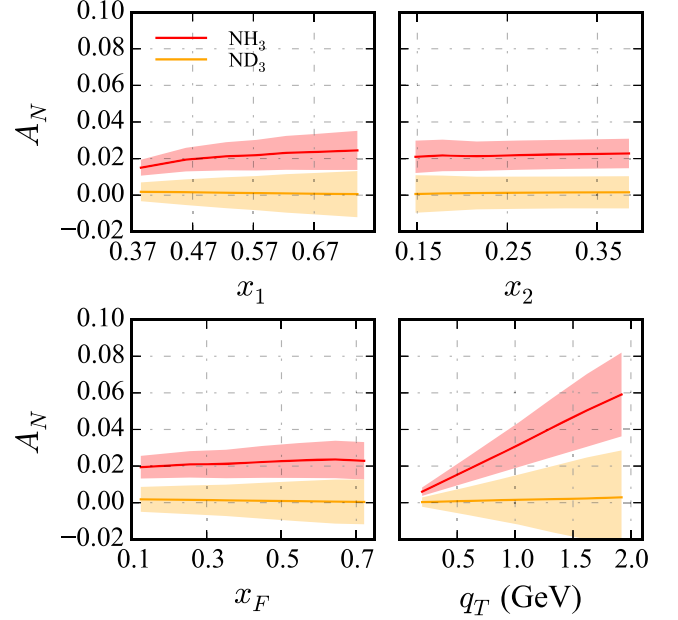


FIG. 16. The *proton*-DNN model (red) and the *deuteron*-DNN model (orange) predictions including 68% C.L. error bands, for Sivers asymmetries in  $x_1, x_2, x_F$ , and  $q_T$  kinematic projections for the SpinQuest DY kinematics [64,66].

exception of this work, no predictions have been made for the polarized DY Sivers asymmetry using a *deuteron* target, which will be measured during the SpinQuest experiment. A noteworthy aspect of the forthcoming SpinQuest experiment is that, in addition to measuring the Sivers asymmetry from *proton* and *deuteron* targets, it will also ascertain the transversity distributions of both quarks and gluons, utilizing a tensor-polarized deuteron Spin 1 target, as proposed in [68].

## VI. SYSTEMATIC STUDIES

Using DNNs in global fits following a similar schema as laid out in this work may afford the loosening of some of the strict kinematic cuts to the experimental data while still preserving the validity of TMD factorization. The methodology is still very new, but some distinct advantages are clear with our DNN approach compared with standard analytical fitting.

### A. Systematic study of data cuts

Neural networks, especially deep neural networks, have a profound capacity for learning complex patterns and relationships in data to model nonlinear and high-dimensional functions without any prior information on the function dimensionality or form. When performing global fits the kinematic dependencies can be intricate and difficult to capture using conventional analytical approaches leading to the need to limit the data to achieve a decent fit. DNNs can adapt to these complexities and, in some cases, benefit from

the lack of kinematic cuts, making the resulting model more robust and still sensitive to a range of kinematic variables.

The TMD factorization formula for the SIDIS hadronic tensor  $W^{\mu\nu}$  was defined in [6,16],

$$\begin{aligned}
 W^{\mu\nu} = & \sum_f |\mathcal{H}_f(Q^2, \mu)|^{\mu\nu} \\
 & \times \int d^2k_\perp d^2p_\perp \delta^{(2)}(z_h k_\perp + p_\perp - p_{hT}) \\
 & \times F_{f/N^\dagger}(x, z_h k_\perp, S; \mu, \zeta_F) D_{h/f}(z_h, p_\perp; \mu, \zeta_D) \\
 & + Y(p_{hT}, Q^2), \tag{34}
 \end{aligned}$$

where all nonperturbative information (the soft part) is encoded in  $F_{f/N^\dagger}$  and the  $D_{h/f}$  whereas the perturbatively calculable hard part is denoted by  $|\mathcal{H}_f(Q^2, \mu)|^{\mu\nu}$ . For preserving the validity of TMD factorization normally strict kinematic cuts are applied. However, applying stringent cuts to already limited experimental data reduces the statistical significance of the model and may lead to the loss of valuable information. We explore a range of possible cuts with the intention of preserving as much data as possible and taking care to not introduce bias by only selecting data that lead to better fits. The unique feature of DNNs to perform well even when trained on a broader range of data with complex correlations serves as a major advantage over fitting analytically. It then serves, whenever there is not a direct conflict with the necessary factorization theorem, to include data points from regions that might otherwise be excluded in traditional kinematic cuts allowing the DNN model to build implicit inclusion of the necessary corrections.

DNNs have the ability to implicitly capture higher-order effects that would be near impossible to obtain using a direct analytical fit unless the initial Ansatz is very lucky or the function form has been proven to contain the necessary physics. TMD factorization relies on specific assumptions about the dominance of certain terms in the cross section calculation. One critical limitation is  $q_{hT} \ll Q$  which is required for the derivation of the factorization property suitable for the case of relatively low transverse momentum. In this factorization scheme approximations are made that have errors of order  $(q_{hT}/Q)$ . For  $q_{hT}$  greater than  $Q$ , the conventional formalism, with integrated fragmentation functions, would no longer be valid. This directly leads to the restriction of SIDIS data to a region where  $p_{hT} \ll zQ$ , which can severely reduce the available data. TMD factorization loses accuracy at large  $q_{hT}$ , with fractional errors characterized as  $(q_{hT}/Q)^\alpha$ . The Collins and Soper approach [16] gives  $(m/Q)$  errors for the full range of  $q_{hT}$  which treats the TMD term as a first approximation to the cross section and allows for the application of a correction by applying an additive approximation from the ordinary collinear factorization. Such corrections can be implicitly

captured when training a DNN model over the full range of  $p_{hT}$ . This is the case even if the assumptions such corrections are based on are not relevant in all kinematic regions of the applicable data. The only requirement is that TMD factorization does not break down at a rigid boundary but instead, remains valid but at a cost to the models' accuracy. The scale of such errors can be numerically estimated so there is an advantage to using as much data as possible and then studying the systematic effects of certain limitations. This approach allows the implicit inclusion of higher-order effects in the DNN model, providing the means for a more comprehensive analysis that can handle a wider kinematic range without sacrificing factorization validity. However, careful consideration and validation are necessary to ensure the reliability of the model as well as accurate quantification of the systematic error as a function of  $p_{hT}$ .

The applicability of TMD factorization was ensured by applying cuts to SIDIS data based on various criteria in the literature. Although the limit  $q_T \ll Q$  [12,13,15,16,69–72] covers the conventional TMD factorization formalism, the large- $Q$  requirement is needed for suppressing the power corrections  $\sim m^2/Q^2$  and  $\sim \Lambda^2/Q^2$ , where  $\Lambda$  is a general nonperturbative scale of QCD. Since  $m$  and  $\Lambda$  are  $\sim 1$  GeV, the  $Q > 2$  GeV condition was applied on top of the  $\delta = p_{hT}/(zQ) \leq 0.3$  condition in [7,11] although the phenomenological region for  $\delta$  is 0.2–0.3 in order for the TMD factorization to be valid. It is practically well known by the effort on global fits, that accommodating more data points is a bigger challenge when using a smaller lower bound for  $\delta$ . Therefore, a more conservative limit  $q_T/Q < 0.75$  was used in [9] to retain a large enough dataset to perform a meaningful fit. On the other hand, a combination of cuts,  $Q^2 > 1.63$  GeV<sup>2</sup>,  $0.2 < z < 0.6$ ,  $0.2 < p_{hT} < 0.9$  GeV, was applied in [5], and also discussed in detail in [73] with the standpoint that data which satisfy  $p_{hT} \ll Q$  may not satisfy  $q_T \ll Q$  depending on the value of  $z_h$  because  $q_T \simeq p_{hT}/z_h$ , and therefore be difficult to describe in a TMD approach. In addition to Siverson function extractions this is also the case for other TMDs extractions [74].

In this exploratory effort with DNNs, such later-mentioned power corrections are not directly imposed. In addition to the data basic cut  $Q^2 > 1$  GeV<sup>2</sup> we performed  $Q^2 > 2$  GeV<sup>2</sup> and  $p_{hT} < zQ$  cuts separately with the proton-DNN model to understand the impact on the extracted Siverson functions. In addition, a dedicated DNN fit has been performed with the JAM20 [5] cuts: i.e.  $Q^2 > 1.63$  GeV<sup>2</sup>,  $0.2 < z < 0.6$ ,  $0.2 < p_{hT} < 0.9$  GeV as a demonstration of impact from such a selective combination of cuts on the extracted Siverson functions.

The results are plotted in Fig. 17 only representing the Siverson functions for  $u, d$  flavors. In summary, all cuts analyzed which respect  $Q^2 > 2$  GeV<sup>2</sup> and  $p_{hT} < zQ$ , except the JAM20 [5] cuts, are consistent with the DNN

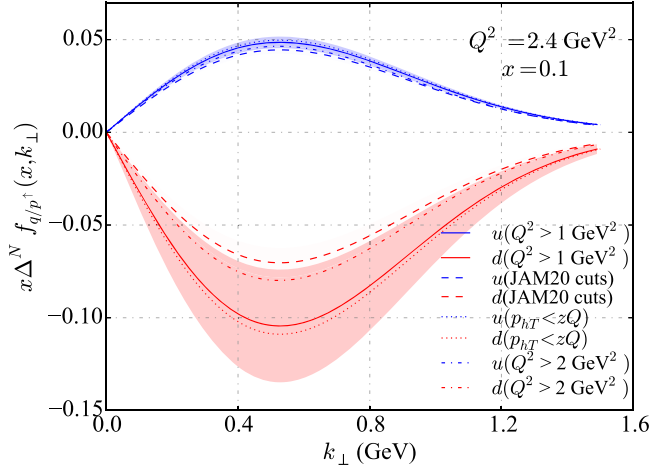


FIG. 17. Solid lines with light band represent the  $u$  (in blue),  $d$  (in red) Sivers functions using the cut  $Q^2 > 1 \text{ GeV}^2$ . These resulting DNN models made from the cuts from all tests are also shown.

model that only contains the  $Q^2 > 1 \text{ GeV}^2$  cut which is the same as without any cuts at all. Note that  $Q^2 > 1 \text{ GeV}^2$  is the recommended generic cut in [16] assuming the application of corrections and error estimates for increasing  $q_T$ . The deviation measured in this study of cuts that is strictly dependent on  $q_T$  is only of  $\sim 2\%$  so it is clearly advantageous to incorporate a wide range of  $q_T$  to ensure that there is implicit inclusion of these corrective contributions to the hadronic tensor from the  $Y$  term built into the DNN model. Figure 18 shows the resultant Sivers functions from proton DNN for all six light-quark flavors with relatively small uncertainty bands caused by the selection of cuts in JAM20 [5]. Note that the uncertainty band represents the statistical component with 68% C.L. from 1000 replica models.

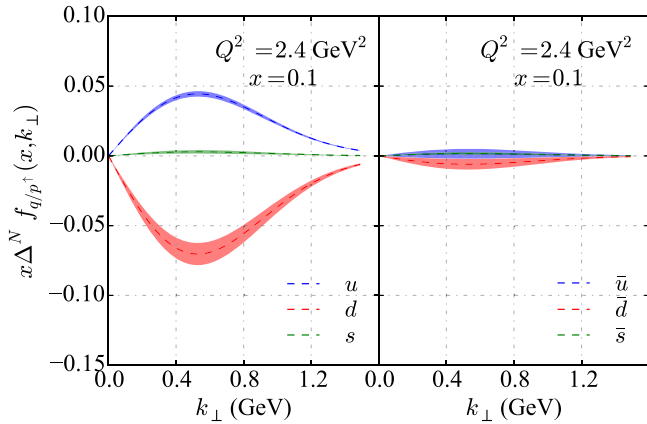


FIG. 18. Sivers functions from a retrained DNN model using the cuts [65] to the data demonstrating that being selective with the data can reduce the error bands of the fit but may also add an unintentional bias.

## B. Systematic study on choice of $h(k_\perp)$

In the original framework [4], the Sivers function is written as the factorized form, Eq. (4), where  $h(k_\perp)$  is understood to be of an unknown form that is simply postulated by the authors (in [4]) based on the assumed kinematic response. Indeed, the analytical expression of  $h(k_\perp)$  has been treated with various types of ansatz and mostly with the Gaussian type of parametrization [4,5,7,35,51]. It is also entirely possible that this term has nothing to do with the proper theoretical treatment as suggested in [6]. When a term in the factorized TMD expression is required to manage some kinematic behavior but has not been formally derived, the DNN analysis presented can be particularly useful as it allows for a high-quality fit despite the dependence of  $h(k_\perp)$ . After the determination of the other terms using the DNN, like  $\mathcal{N}_q(x)$  in this case, the terms can then be separated and studied independently to determine the interpretation. If  $h(k_\perp)$  or any of the multiplicative terms are biases then the DNN can be used to compensate for the bias. This is done by building an architecture that has orders of magnitude more parameters (usually thousands) than the expression in question,  $h(k_\perp)$  in this case. This effectively reduces the weight of  $h(k_\perp)$  in the fit. This can be done progressively and systematically with intentional control of how much each term contributes to the fit. In the fit performed for this analysis, we build the DNN to optimize the loss directly so the contribution of  $h(k_\perp)$  to the final fit is small though the DNN becomes specialized to the particular  $h(k_\perp)$ . In this way, the DNN directly mitigates any type of incorrect Ansatz if the final number of DNN parameters is very large with respect to  $h(k_\perp)$ . This can be done while still rendering a high-quality numerical model encoding all information available in the data. The DNN term becomes uniquely parametrized to account for the biases while the resulting combination of terms in the model remains largely unbiased, even though each individual term may not be independently meaningful. It should be noted that for our analysis the choice of having the DNN model  $\mathcal{N}_q(x)$  rather than  $\mathcal{N}_q(x)h(k_\perp)$  or some other choice is arbitrary and done purely as an exercise to demonstrate the flexibility of the technique. To study the systematic variation of the choice of  $h(k_\perp)$ , we used the same candidate functions for  $h(k_\perp)$  that were also used in [35]. Those are

$$h(k_\perp) = \sqrt{2}e \frac{k_\perp}{m_1} e^{-k_\perp^2/m_1^2} \quad (35)$$

and

$$h(k_\perp) = \frac{2k_\perp m_1}{m_1^2 + k_\perp^2}. \quad (36)$$

The model must be trained separately with each  $h(k_\perp)$  creating entirely different models for  $\mathcal{N}_q(x)$  that result in

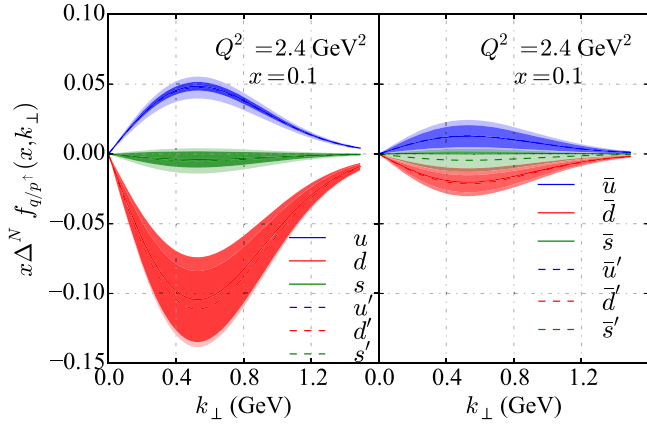


FIG. 19. Using two different  $h(k_\perp)$ . Solid line with dark band represents the Siverts functions with  $h(k_\perp) = \sqrt{2}e^{\frac{k_\perp}{m_1}}e^{-k_\perp^2/m_1^2}$ , whereas the dashed line with light band represents the Siverts functions with  $h(k_\perp) = \frac{2k_\perp m_1}{m_1^2 + k_\perp^2}$ .

the same Siverts function as shown in Fig. 19. The solid line with a dark band represents the Siverts functions with  $h(k_\perp) = \sqrt{2}e^{\frac{k_\perp}{m_1}}e^{-k_\perp^2/m_1^2}$ , whereas the dashed line with a light band represents the Siverts functions with  $h(k_\perp) = \frac{2k_\perp m_1}{m_1^2 + k_\perp^2}$ . It is clear that the DNN  $\mathcal{N}_q(x)$  is capable of incorporating both types of  $h(k_\perp)$  without affecting the Siverts functions in the final model as well as the asymmetries (with deviation less than 1%).

## VII. THE 3D TOMOGRAPHY OF THE PROTON

The TMD density of unpolarized quarks inside a proton polarized in the  $\hat{y}$  direction can be graphically represented using the relation [7,51],

$$\rho_{p\uparrow}^a(x, k_x, k_y; Q^2) = f_1^a(x, k_\perp^2; Q^2) - \frac{k_x}{m_p} f_{1T}^{\perp a}(x, k_\perp^2; Q^2), \quad (37)$$

where  $k_\perp$  is a two-dimensional vector  $(k_x, k_y)$ , and the unpolarized TMD and the Siverts function for quark flavor  $a$  are respectively represented as  $f_1^a(x, k_\perp^2; Q^2)$ , and  $f_{1T}^{\perp a}(x, k_\perp^2; Q^2)$ . The corresponding quark density distributions from our *proton*-DNN model for all light-quark flavors in  $SU(3)_{\text{flavor}}$  at  $x = 0.1$  and  $Q^2 = 2.4 \text{ GeV}^2$  are shown in Fig. 20. The observed shifts in each quark flavor are linked to the correlation between the OAM of quarks and the spin of the proton. The results shown in Fig. 20 provide evidence of nonzero OAM in the wave function of the proton's valence and sea quarks. The *proton*-DNN model calculations for the  $u$  and  $d$  quarks are similar to those reported in [7,51], where the distortion has a positive shift for the  $u$  quark and a negative shift for the  $d$  quark with respect to the  $+x$  direction. From the results in Fig. 20, the

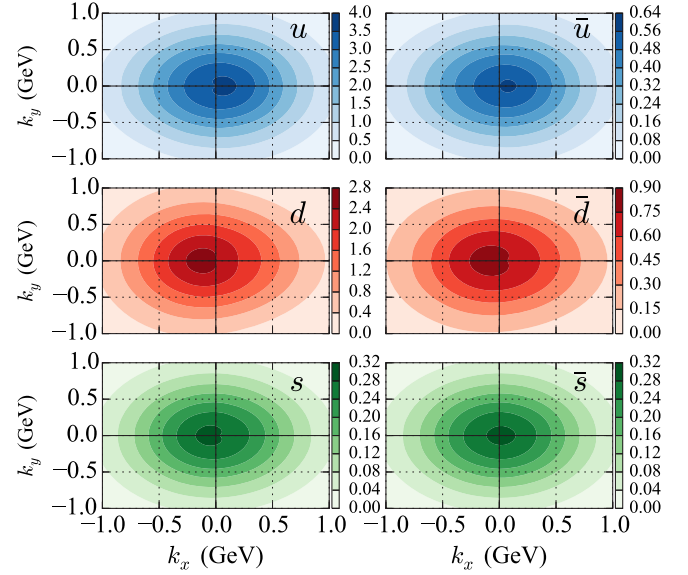


FIG. 20. Quark density distributions  $\rho_{p\uparrow}^a$  from the *proton*-DNN model (average of 1000 replicas) for the light-quark flavor  $a = \{u, \bar{u}, d, \bar{d}, s, \bar{s}\}$  inside a proton polarized along the  $+y$  direction and moving toward the reader, as a function of  $(k_x, k_y)$  at  $x = 0.1$  and  $Q^2 = 2.4 \text{ GeV}^2$ .

*proton*-DNN model demonstrates that a virtual photon traveling toward a polarized proton “sees” an enhancement of the quark distribution, in particular more  $u, \bar{u}$  quarks to its right-hand side and more  $d, \bar{d}$  quarks to its left-hand side in the momentum space. Moreover, the resultant shifts for  $\bar{u}, s$  quarks from the *proton*-DNN model are also in agreement with [7]. In the low- $x$  region, the momentum space quark density becomes almost symmetric [51], and it indicates that the Siverts effect becomes smaller and the corresponding experimentally observed asymmetry is small.

The forthcoming data from Jefferson Lab at 12 GeV, Fermilab SpinQuest experiment, and the anticipated future data from the Electron-Ion Collider [75–77], along with their extensive kinematic coverage, are expected to provide invaluable insights into the 3D structure of the nucleon. Obtaining a model-independent estimate of quark angular momentum requires parton distributions that simultaneously depend on both momentum and position [78–81]. In addition to experimental observations, lattice QCD (LQCD) computations provide a valuable tool for QCD phenomenology from first principles. For instance, LQCD has been utilized to investigate the Siverts effect and other TMD observables at different pion masses [82] as well as the generalized parton distribution at the physical pion mass [83]. Additionally, LQCD results on the Collins-Soper kernel over a range of  $b_T$  (the Fourier transform of the transverse momentum) are useful for global fits of TMD observables from different processes [84]. In this way, LQCD could complement the experimental data and open



up an avenue to enhance the DNN method to explore the 3D structure of nucleons more directly.

### VIII. EXPLORING EVOLUTION

The solution of the TMD evolution equations [12,16],

$$\mu^2 \frac{dF(x, b; \mu, \zeta)}{d\mu^2} = \frac{\gamma_F(\mu, \zeta)}{2} F(x, b; \mu, \zeta) \quad (38)$$

$$\zeta \frac{dF(x, b; \mu, \zeta)}{d\zeta} = -\mathcal{D}(b, \mu) F(x, b; \mu, \zeta), \quad (39)$$

can be written as the following simplified form in terms of the Fourier transform of  $k_\perp$  (i.e.,  $b$ ) [7] where  $F$  can be any TMD distribution

$$F(x, b; \mu, \zeta) = \left( \frac{\zeta}{\zeta_\mu(b)} \right)^{-\mathcal{D}(b, \mu)} F(x, b), \quad (40)$$

and  $\mathcal{D}(b)$  is the nonperturbative Collins-Soper kernel. Also in the literature, these scales were generally selected as

$$\mu \sim Q, \quad \zeta_F \zeta_D \sim Q^4, \quad \mu^2 = \zeta^2 = Q^2 \quad (41)$$

[6–8,16,85], and the global fits have been performed using some form of evolution factor as a function of the Collins-Soper kernel. Although the full analysis of incorporating TMD evolution from the DNN fit is beyond the scope of this work, a preliminary DNN fit has been performed by modifying the  $\mathcal{N}_q(x)$  as  $\mathcal{N}_q(x, Q^2)$  by adding a separate input node for  $Q^2$  in addition to  $x$ . Figure 21 shows the percentage of the Sivers asymmetry ( $A_{UT}^{\sin(\phi_h - \phi_s)}$ ) vs  $Q^2$  (GeV<sup>2</sup>) in comparison with [6]. The preliminary version of

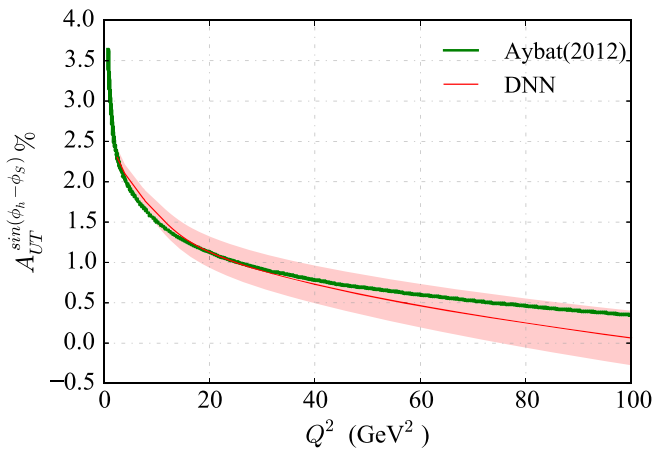


FIG. 21. The Sivers asymmetry evolution in  $Q^2$  compared to the result from [6]. The red-colored solid line and the band represent the mean and standard deviation of the  $A_{UT}^{\sin(\phi_h - \phi_s)}$  from 1000 replica models of the proton DNN at  $x = 0.12$ ,  $z = 0.32$ ,  $p_{hT} = 0.14$  GeV.

the TMD evolution from DNN is in agreement with the observation in [6] within 68% C.L. (with 1000 replica models) regarding the suppression of the full asymmetry faster than  $\sim 1/\sqrt{Q}$ , but slower than  $\sim 1/Q^2$ . Moreover, the dynamic trend and consistency with the evolution behavior seen in [6] indicates that a complete evolution treatment at large  $Q$  may be a collective effect and worthy of a deeper investigation with our DNN approach.

As presented the Sivers function satisfies DGLAP evolution but not a TMD evolution. We have however introduced the first steps of how this analysis could be performed. However, in order to perform the best evolution analysis it would be best to start with no prior analytical Ansatz.

### IX. CONCLUSION AND DISCUSSION

In conclusion, this paper demonstrates the effectiveness of using specialized DNNs as part of a fit function used to obtain a global extraction of the Sivers function from transverse single-spin asymmetry experimental data. It has been clearly shown that these tools are incredibly flexible and can be used to build accurate models even under the condition that the factorized terms contain biases. By training the model with these terms present, the DNN can account for these biases, resulting in a largely unbiased final model. This approach enables the exploration of existing formalism and the testing of new phenomenology while managing and studying biases in individual terms. The proposed method can provide a means to minimize errors and ambiguity associated with the ill-defined expressions normally constructed to meet the factorization requirements.

Our proposed method leverages AI to perform global fits and extract the Sivers distributions of unpolarized quarks in both polarized protons and neutrons. We use a generating function to ensure robustness and accuracy in the extraction process. Progressive improvement in the extracted information can be achieved by optimizing architecture and the corresponding hyperparameters using pseudodata generated in the same kinematic bins as the real data and then translating that improvement to real experimental data extraction. Our schema handles complex and sparse data effectively, with the generating function providing additional quality control and the means to quantify accuracy and precision.

As the first attempt to extract the Sivers function using AI techniques, we chose the  $\mathcal{N}_q(x)$  parametrization of the deep neural net to incorporate all  $x$ -dependent features. This initial DNN extraction of  $\mathcal{N}_q(x)$  uses SIDIS data from HERMES and COMPASS to build a Sivers function model which can then be used to make predictions for the Sivers asymmetries for both SIDIS and DY processes. The fitting method successfully extracts the Sivers function for valence quarks and light-quark flavors in a flavor-independent manner. This investigation is exploratory, with the intention

of developing tools and techniques that minimize error and maximize utility, which we hope to expand upon in further work. The trained model predicts results consistent with experimental data, demonstrating its predictive capability. Projections for the sea-quark Sivers function are made for the SpinQuest experiment with polarized proton and deuteron targets.

There are two key challenges when using data to extract TMDs. One is the applicability of TMD factorization and the other is achieving meaningful fit results. An exploratory effort has been performed to address this with our schema along with a variety of data cuts. As a result, the DNN shows the variation of  $q_T$  dependent cuts, resulting in small deviations in the extracted Sivers functions compared with the result with a wide range of  $q_T$ , and simultaneously following the TMD factorization theorem. We also analyze the TMD evolution of the Sivers asymmetry, which aligns with previous work and indicates the potential of the DNN approach. Further analysis in this direction is warranted.

Our method and techniques are intended to be simple and reproducible, highlighting the potential of AI for data analysis and information extraction. Cooperation between

experimentalists, theorists, and the growing computational efforts is crucial for accelerating advancements in the field. A global effort to standardize data collection, organization, and storage in an unbinned format with detailed covariance information must be developed to take full advantage of the AI tools now available. AI and its emerging technologies will continue to accelerate the progress of data-driven physics, but the speed of that progress largely depends on the level of foresight and cooperation within the community.

## ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy (DOE) Contract No. DE-FG02-96ER40950. The authors would like to thank Micheal Wagman, Andrea Signori, Alexei Prokudin, and Filippo Delcarro for their insightful discussions. Additionally, the authors would like to acknowledge Bakur Parsamyan for providing the COMPASS data, Luciano Pappalardo for providing the HERMES data, and Sassot Rodolfo for providing the Fragmentation Functions.

- 
- [1] D. Sivers, *Phys. Rev. D* **41**, 83 (1990).
  - [2] D. Sivers, *Phys. Rev. D* **43**, 261 (1991).
  - [3] M. Anselmino, M. Boglione, U. D'Alesio, S. Melis, F. Murgia, and A. Prokudin, *Phys. Rev. D* **79**, 054010 (2009).
  - [4] M. Anselmino, M. Boglione, U. D'Alesio, F. Murgia, and A. Prokudin, *J. High Energy Phys.* **04** (2017) 046.
  - [5] J. Cammarota, L. Gamberg, Z.-B. Kang, J. A. Miller, D. Pitonyak, A. Prokudin, T. C. Rogers, and N. Sato (Jefferson Lab Angular Momentum (JAM) Collaboration), *Phys. Rev. D* **102**, 054002 (2020).
  - [6] S. M. Aybat, A. Prokudin, and T. C. Rogers, *Phys. Rev. Lett.* **108**, 242003 (2012).
  - [7] M. Bury, A. Prokudin, and A. Vladimirov, *J. High Energy Phys.* **05** (2021) 151.
  - [8] M. G. Echevarria, A. Idilbi, Z.-B. Kang, and I. Vitev, *Phys. Rev. D* **89**, 074013 (2014).
  - [9] M. G. Echevarria, Z.-B. Kang, and J. Terry, *J. High Energy Phys.* **01** (2021) 126.
  - [10] Z.-B. Kang and J.-W. Qiu, *Phys. Rev. Lett.* **103**, 172001 (2009).
  - [11] M. Bury, A. Prokudin, and A. Vladimirov, *Phys. Rev. Lett.* **126**, 112002 (2021).
  - [12] I. Scimemi and A. Vladimirov, *J. High Energy Phys.* **06** (2020) 137.
  - [13] J. C. Collins and D. E. Soper, *Nucl. Phys.* **B193**, 381 (1981).
  - [14] J. C. Collins and D. E. Soper, *Nucl. Phys.* **B213**, 545 (1983).
  - [15] J. C. Collins, D. E. Soper, and G. Sterman, *Nucl. Phys.* **B250**, 199 (1985).
  - [16] J. Collins, *Foundations of Perturbative QCD* (Cambridge University Press, Cambridge, England, 2011).
  - [17] R. Meng, F. I. Olness, and D. E. Soper, *Nucl. Phys.* **B371**, 79 (1992).
  - [18] X. Ji, J.-P. Ma, and F. Yuan, *Phys. Lett. B* **597**, 299 (2004).
  - [19] X. Ji, J.-P. Ma, and F. Yuan, *Phys. Rev. D* **71**, 034005 (2005).
  - [20] X. Ji, J.-P. Ma, and F. Yuan, *J. High Energy Phys.* **07** (2005) 020.
  - [21] M. G. Echevarria, A. Idilbi, and I. Scimemi, *J. High Energy Phys.* **07** (2012) 002.
  - [22] S. Arnold, A. Metz, and M. Schlegel, *Phys. Rev. D* **79**, 034005 (2009).
  - [23] M. Anselmino, M. Boglione, U. D'Alesio, A. Kotzinian, F. Murgia, and A. Prokudin, *Phys. Rev. D* **72**, 094007 (2005).
  - [24] J. C. Collins, A. V. Efremov, K. Goeke, S. Menzel, A. Metz, and P. Schweitzer, *Phys. Rev. D* **73**, 014021 (2006).
  - [25] W. Vogelsang and F. Yuan, *Phys. Rev. D* **72**, 054028 (2005).
  - [26] M. Anselmino, M. Boglione, U. D'Alesio, A. Kotzinian, S. Melis, F. Murgia, A. Prokudin, and C. Türk, *Eur. Phys. J. A* **39**, 89 (2009).
  - [27] A. Bacchetta and M. Radici, *Phys. Rev. Lett.* **107**, 212001 (2011).
  - [28] K. Hornik, M. Stinchcombe, and H. White, *Neural Netw.* **2**, 359 (1989).
  - [29] B. C. Csáji, Approximation with artificial neural networks, MS Thesis, Dept. Science, Eotvos Lorand Univ., Budapest, Hungary, 2001.
  - [30] J. C. Collins, *Phys. Lett. B* **536**, 43 (2002).
  - [31] Z.-B. Kang and J.-W. Qiu, *Phys. Rev. D* **81**, 054020 (2010).
  - [32] A. Bacchetta, U. D'Alesio, M. Diehl, and C. A. Miller, *Phys. Rev. D* **70**, 117504 (2004).

- [33] M. Anselmino, U. D'Alesio, and F. Murgia, *Phys. Rev. D* **67**, 074010 (2003).
- [34] M. Anselmino, M. Boglione, U. D'Alesio, J. O. G. Hernandez, S. Melis, F. Murgia, and A. Prokudin, *Phys. Rev. D* **92**, 114023 (2015).
- [35] M. Anselmino, M. Boglione, U. D'Alesio, A. Kotzinian, F. Murgia, and A. Prokudin, *Phys. Rev. D* **71**, 074006 (2005).
- [36] D. Stump, J. Huston, J. Pumplin, W.-K. Tung, H. L. Lai, S. Kuhlmann, and J. F. Owens, *J. High Energy Phys.* **10** (2003) 046.
- [37] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, *Eur. Phys. J. C* **75**, 132 (2015).
- [38] R. D. Ball *et al.* (The NNPDF Collaboration), *Eur. Phys. J. C* **82**, 428 (2022).
- [39] D. de Florian, R. Sassot, M. Epele, R. J. Hernández-Pinto, and M. Stratmann, *Phys. Rev. D* **91**, 014035 (2015).
- [40] D. de Florian, M. Epele, R. J. Hernández-Pinto, R. Sassot, and M. Stratmann, *Phys. Rev. D* **95**, 094019 (2017).
- [41] A. Signori, A. Bacchetta, M. Radici, and G. Schnell, *J. High Energy Phys.* **11** (2013) 194.
- [42] M. Anselmino, M. Boglione, J. O. Gonzales-Hernandez, S. Melis, and A. Prokudin, *J. High Energy Phys.* **04** (2014) 005.
- [43] A. Airapetian *et al.* (HERMES Collaboration), *Phys. Rev. D* **87**, 074029 (2013).
- [44] J. C. Collins, A. V. Efremov, K. Goeke, M. G. Perdekamp, S. Menzel, B. Meredith, A. Metz, and P. Schweitzer, *Phys. Rev. D* **73**, 094023 (2006).
- [45] A. V. Efremov, K. Goeke, S. Menzel, A. Metz, and P. Schweitzer, *Phys. Lett. B* **612**, 233 (2005).
- [46] H. Dembinski, P. Ongmongkolkul *et al.*, scikit-hep/iminuit, 10.5281/zenodo.3949207 (2020).
- [47] J. C. Collins, A. V. Efremov, K. Goeke, S. Menzel, A. Metz, and P. Schweitzer, *Phys. Rev. D* **73**, 014021 (2006).
- [48] M. Anselmino, M. Boglione, and S. Melis, *Phys. Rev. D* **86**, 014028 (2012).
- [49] P. Sun and F. Yuan, *Phys. Rev. D* **88**, 114012 (2013).
- [50] X. Luo and H. Sun, *Phys. Rev. D* **101**, 074016 (2020).
- [51] A. Bacchetta, F. Delcarro, C. Pisano, and M. Radici, *Phys. Lett. B* **827**, 136961 (2022).
- [52] M. Alekseev *et al.* (COMPASS Collaboration), *Phys. Lett. B* **673**, 127 (2009).
- [53] M. Aghasyan *et al.* (COMPASS Collaboration), *Phys. Rev. Lett.* **119**, 112002 (2017).
- [54] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, Software available from <https://www.tensorflow.org> (2015).
- [55] A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. (O'Reilly Media, Sebastopol, Canada, 2019).
- [56] A. Airapetian *et al.* (The HERMES Collaboration), *J. High Energy Phys.* **12** (2020) 10.
- [57] X. Qian *et al.* (Jefferson Lab Hall A Collaboration), *Phys. Rev. Lett.* **107**, 072003 (2011).
- [58] Y. X. Zhao *et al.* (Jefferson Lab Hall A Collaboration), *Phys. Rev. C* **90**, 055201 (2014).
- [59] A. Airapetian *et al.* (HERMES Collaboration), *Phys. Rev. Lett.* **103**, 152002 (2009).
- [60] C. Adolph *et al.*, *Phys. Lett. B* **744**, 250 (2015).
- [61] M. Boglione, U. D'Alesio, C. Flore, and J. O. Gonzalez-Hernandez, *J. High Energy Phys.* **07** (2018) 148.
- [62] P. V. Pobylitsa, arXiv:hep-ph/0301236.
- [63] F. He and P. Wang, *Phys. Rev. D* **100**, 074032 (2019).
- [64] A. Klein, D. Keller, K. Liu *et al.*, E1039 FNAL proposal SEAQUEST Document 1720-v3 (2016).
- [65] P. C. Barry, C.-R. Ji, N. Sato, and W. Melnitchouk (JAM Collaboration), *Phys. Rev. Lett.* **127**, 232001 (2021).
- [66] J. Dove *et al.* (FNAL E906/SeaQuest Collaboration), arXiv:2212.12160.
- [67] A. Chen *et al.*, *Proc. Sci.*, SPIN2018 (2019) 164 [arXiv:1901.09994].
- [68] D. Keller, arXiv:2205.01249.
- [69] R. Meng, F. I. Olness, and D. E. Soper, *Phys. Rev. D* **54**, 1919 (1996).
- [70] I. Scimemi and A. Vladimirov, *Eur. Phys. J. C* **78**, 89 (2018).
- [71] A. Bacchetta, V. Bertone, C. Bissolotti, G. Bozzi, F. Delcarro, F. Piacenza, and M. Radici, *J. High Energy Phys.* **07** (2020) 117.
- [72] A. Bacchetta, V. Bertone, C. Bissolotti, G. Bozzi, M. Cerutti, F. Piacenza, M. Radici, A. Signori (The MAP Collaboration), *J. High Energy Phys.* **10** (2022) 127.
- [73] R. Boussarie *et al.*, arXiv:2304.03302.
- [74] S. Bhattacharya, Z.-B. Kang, A. Metz, G. Penn, and D. Pitonyak, *Phys. Rev. D* **105**, 034007 (2022).
- [75] D. Boer *et al.*, arXiv:1108.1713.
- [76] A. Accardi *et al.*, *Eur. Phys. J. A* **52**, 268 (2016).
- [77] R. A. Khalek *et al.*, *Nucl. Phys.* **B1026**, 122447 (2022).
- [78] X. Ji, *Phys. Rev. Lett.* **78**, 610 (1997).
- [79] C. Lorcé, B. Pasquini, X. Xiong, and F. Yuan, *Phys. Rev. D* **85**, 114006 (2012).
- [80] M. Radici, *J. Phys. Conf. Ser.* **527**, 012025 (2014).
- [81] N. G. Stefanis, C. Alexandrou, T. Horn, H. Moutarde, and I. Scimemi, *EPJ Web Conf.* **137**, 01003 (2017).
- [82] B. Yoon *et al.*, *Phys. Rev. D* **96**, 094508 (2017).
- [83] H.-W. Lin, *Phys. Rev. Lett.* **127**, 182001 (2021).
- [84] P. Shanahan, M. Wagman, and Y. Zhao, *Phys. Rev. D* **102**, 014511 (2020).
- [85] S. M. Aybat, J. C. Collins, J. W. Qiu, and T. C. Rogers, *Phys. Rev. D* **85**, 034043 (2012).