

Physics-Informed Global Extraction of the Universal Small- x Dipole Amplitude

Si-Wei Dai,^{1,2,*} Fu-Peng Li,^{3,4,†} Long-Gang Pang,^{1,2,‡} Guang-You Qin,^{1,2,§} Shu-Yi Wei,^{5,¶} Han-Zhong Zhang,^{1,2,**} and Wenbin Zhao^{1,2,††}

¹Key Laboratory of Quark and Lepton Physics (MOE) & Institute of Particle Physics,
Central China Normal University, Wuhan 430079, China

²Artificial Intelligence and Computational Physics Research Center,
Central China Normal University, Wuhan 430079, China

³Key Laboratory of Nuclear Physics and Ion-beam Application (MOE) &
Institute of Modern Physics, Fudan University, Shanghai 200433, China

⁴Shanghai Research Center for Theoretical Nuclear Physics,
NSFC and Fudan University, Shanghai 200438, China

⁵Institute of Frontier and Interdisciplinary Science,
Key Laboratory of Particle Physics and Particle Irradiation (MOE),
Shandong University, Qingdao, Shandong 266237, China

We extract the universal small- x dipole scattering amplitude $N(r, x_B)$ from a global analysis based on a physics-informed neural network (PINN), without imposing a priori MV-type parametrization of the initial condition. The network provides a smooth and differentiable surrogate for $N(r, x_B)$, whose rapidity dependence is constrained by the collinearly improved Balitsky–Kovchegov evolution equation, while its functional form is simultaneously constrained by Deep Inelastic Scattering (DIS) data for the reduced total and charm cross sections, exclusive J/ψ photoproduction measurements, and a positivity requirement for the momentum-space dipole amplitude. The resulting single universal amplitude consistently describes all fitted observables within a unified framework, alleviating the long-standing tension between total and charm channels encountered in conventional small- x fits based on rigid parametric ansätze. Within the fitted kinematic domain, the best extracted PINN solution yields a smooth, non-negative momentum-space dipole over the full transverse-momentum range examined. Our results provide a robust and well-behaved input for Color Glass Condensate phenomenology across a broad class of high-energy processes.

Introduction. At high energies, Quantum Chromodynamics (QCD) enters a regime where nonlinear gluon recombination tames the growth of gluon densities, leading to gluon saturation. This regime is systematically described by the Color Glass Condensate (CGC) effective field theory [1–7]. Within this framework, the small- x evolution of Wilson-line correlators is governed by the JIMWLK renormalization group equation [8–14]. In the mean-field and large- N_c limit it reduces to the Balitsky–Kovchegov (BK) equation [15, 16], which is a closed evolution equation for the dipole amplitude $N(r, x_B)$. This amplitude is the universal input for a wide class of small- x observables, including inclusive DIS structure functions, exclusive diffractive processes and inclusive hadron productions in e - p/A and p - p/A collisions.

DIS measurements, in particular of reduced total and heavy-flavor production cross sections at small- x provide the clean lever arm to probe saturation dynamics. DIS off nuclei is especially compelling because saturation effects are parametrically enhanced in heavy nuclei [17]. The future Electron-Ion Collider (EIC) [18–20] will enable pre-

cision measurements over a broad kinematic range in the gluon longitudinal momentum fraction x_B and photon virtuality Q^2 . Establishing nonlinear QCD dynamics at collider energies and quantifying its impact on the small- x_B structure of protons and nuclei will therefore require global analyses that simultaneously confront multiple, complementary DIS channels. In particular, combining inclusive and heavy-flavor data is essential because they probe different dipole sizes and hence different regions of the dipole amplitude.

Despite substantial progress, several intertwined issues remain in small- x_B global analyses. First, achieving a simultaneous description of DIS reduced total cross section σ_r and the charm reduced cross section $\sigma_r^{c\bar{c}}$ is challenging [21–25]. Because charm production preferentially probes smaller dipoles, $\sigma_r^{c\bar{c}}$ imposes particularly stringent constraints on the short-distance behavior of $N(r, x_B)$ and on the conventionally assumed non-perturbative initial condition for the evolution. Second, widely used initial-condition parameterizations can yield a momentum-space dipole, $S(k_T, x_B)$ defined in Eq. (2), that becomes negative over parts of the k_T range, which is in tension with the Fourier-positivity constraints discussed in Refs. [21, 25, 26]. This problem has been highlighted for some 25 years. Enforcing Fourier positivity is important for a consistent momentum-space interpretation of dipole amplitude and for its connection to the gluon content of the hadron [23, 27–31].

In fact, these tensions stem from the functional form assumed for the initial condition. Most analyses adopt

* swdai@mails.ccnu.edu.cn

† fpli@fudan.edu.cn

‡ lgpang@ccnu.edu.cn

§ guangyou.qin@ccnu.edu.cn

¶ shuyi@sdu.edu.cn

** zhanghz@ccnu.edu.cn

†† WenbinZhao@ccnu.edu.cn

a McLerran–Venugopalan (MV) form [32–34], motivated by the large- A limit of a nucleus, where many color sources per unit area ($\propto A^{1/3}$) yield Gaussian fluctuations and a Glauber–Mueller shape. For the *proton* this motivation is weaker, and the MV parameters are typically fitted rather than fixed. Although improved BK evolution with higher-order impact factors has advanced the small- x_B DIS description [24, 25, 35, 36], key tensions persist even in NLO Bayesian fits: the charm $\chi^2/\text{d.o.f.} \gtrsim 2.5$ and the negative regions in $S(k_T, x_B)$ [36, 37]. It was also raised by Giraud and Peschanski [26] as a possible deficiency of the dipole formalism. This suggests the initial-condition form, not the perturbative order alone, may be limiting. It motivates a more flexible yet dynamics-constrained dipole amplitude [25, 36, 38, 39].

In this Letter we introduce a physics-informed neural network (PINN) [40–42] to determine the universal dipole amplitude $N(r, x_B)$ as a differentiable function. Unlike a purely data-driven fit, the PINN enforces the small- x evolution dynamics by including the collinearly improved BK (ciBK) [25, 43–46] evolution residual in the loss function, while experimental data and basic physical constraints anchor the solution. Rather than prescribing a parametric initial condition $N(r, x_0)$, we infer it nonparametrically from data under the evolution constraint. We perform a joint global analysis of DIS observables (including the reduced total, and reduced charm cross section) and exclusive diffractive J/ψ photoproduction, with a Fourier-positivity regularization imposed on the momentum-space dipole S -matrix. Uncertainties are quantified using deep ensembles [47]. We demonstrate that the extracted dipole amplitude simultaneously accommodates reduced total, charm quark cross sections, exclusive vector meson constraints and remains consistent with Fourier-positivity in the data-constrained small- x region.

Physics-informed global analysis. In the small- x limit, the dipole amplitude evolution is governed by the BK equation. Beyond leading-logarithmic accuracy, large collinear corrections can destabilize the evolution [48]. This instability is cured by the ciBK equation, which resums the dominant collinear enhancements to all orders [25, 43–46]. In this work we adopt the ciBK prescription of Refs. [25, 46, 49], formulated in the rapidity of the dense target and including the rapidity-shift (kinematic-constraint) structure of the evolution. Unlike standard global fits that impose a specific analytic ansatz for the non-perturbative initial condition at a starting point x_0 and then evolve to smaller x , which can introduce parametrization bias, we represent the dipole amplitude across the full (r, x_B) domain with a smooth surrogate, $N(r, Y) \equiv \mathcal{N}_\theta(r, Y)$, $Y = \ln(x_0/x_B)$. Here x_0 is the starting point, θ denotes trainable parameters, and constrain $\mathcal{N}_\theta$ directly by the ciBK dynamics. Concretely, we penalize the integro-differential residual

$$\mathcal{R}_{\text{ciBK}}(r, Y) = \frac{\partial \mathcal{N}_\theta(r, Y)}{\partial Y} - \int d^2 \mathbf{r}_1 K_{\text{ciBK}}(\mathbf{r}, \mathbf{r}_1, \mathbf{r}_2) \mathcal{F}[\mathcal{N}_\theta], \quad (1)$$

with $\mathbf{r}_2 \equiv \mathbf{r} - \mathbf{r}_1$, and the residual $\mathcal{R}_{\text{ciBK}}(r, Y)$ enters the total objective in Eq. (3) as a physics-based penalty that enforces the ciBK evolution for $\mathcal{N}_\theta(r, Y)$. Here K_{ciBK} is the collinearly improved kernel and $\mathcal{F}$ denotes the nonlinear interaction functional. Their explicit forms are given in the Supplemental Material. The amplitude $N(r, Y)$ in Eq. (1) denotes the normalized impact-parameter-integrated dipole amplitude. We assume a factorized profile, so that $\sigma_{\text{dip}}(r, Y) = 2 \int d^2 b N(r, \mathbf{b}, Y) = 2\pi R_p^2 N(r, Y) = \sigma_0 N(r, Y)$. This factorized-profile approximation is justified in the dilute regime, and becomes approximate in the saturation region, where the saturation scale can develop a nontrivial impact-parameter dependence; see the recent discussion in Ref. [50, 51]. The ciBK equation employed here is built on the LO BK equation, supplemented by running-coupling corrections and the all-order resummation of the dominant logarithms [25, 43–46]. While it does not include the full finite NLO BK kernel, this resummation serves as a practical prescription, given that the exact NLO BK equation is numerically unstable [52]. By capturing the double- and single-logarithmic terms of the exact NLO kernel, this approach provides an approximation of the NLO+NLL evolution, establishing itself as an effective approach in small- x phenomenology. A full NLO+NLL treatment, which also retains the genuine $\mathcal{O}(\alpha_s^2)$ non-logarithmic kernel terms, is prohibitively slow for a global PINN analysis and is left for future work. The running coupling uses the Brodsky–Lepage–Mackenzie prescription with a variable number of active flavors set by the dipole scale relative to the heavy-quark thresholds; see Sec.A1 of the Supplemental Material for details. In evaluating ciBK residual in Eq. (1), the rapidity shift makes the ciBK evolution non-local in Y , so one must specify a boundary condition for $N(r, Y < Y_0)$. In this work we adopt the frozen-boundary prescription, $N(r, Y < Y_0) = N(r, Y = Y_0)$. We start the evolution at $x_0 = 0.03$, which reduces sensitivity to the boundary prescription and yields a smoother extrapolation towards $x \lesssim 10^{-2}$. The smooth neural surrogate also enables a direct evaluation of $\partial_Y N$ via automatic differentiation and stable Fourier transforms to momentum space, avoiding interpolation artifacts from a discrete (r, Y) grid.

The ciBK equation alone does not uniquely fix $\mathcal{N}_\theta$; we therefore determine it through a global fit to the DIS reduced total (σ_r), charm quark ($\sigma_r^{c\bar{c}}$) cross section, and total exclusive J/Ψ photoproduction. Even though our theoretical evolution initialized at higher momentum fractions, our actual fit to experimental data is strictly confined to the $x_B < 0.004$ region. While $x_B \sim 0.01$ is widely regarded as the onset of gluon saturation, we restrict to $x_B < 0.004$ to avoid the tail of the transitional region where DGLAP-dominated collinear and non-linear ciBK dynamics overlap [53, 54]. This conservative cut also mitigates the numerical impact of missing higher-order corrections in the DIS impact factors and reduces potential contamination from valence-quark contributions. We employ an impact-parameter inde-

pendent dipole amplitude and focus on total cross sections. For J/ψ we include the standard real-part and skewness corrections [1, 55–59], use the Boosted Gaussian wave function [58], and allow an overall normalization $K_{\text{factor}}^{\text{VM}}$ [60, 61]. We take $m_{u,d,s}^{\text{eff}} = 0.14$ GeV and $m_c = 1.4$ GeV. Throughout, we employ leading-order impact factors, leaving a full NLO treatment for future work, since NLO calculations are several orders of magnitude slower than LO, particularly for heavy quarks computation, it involves up to a seven-dimensional numerical integral [24, 25, 35, 36]. The details of our calculations are presented in the Supplemental Material.

We further impose essential physical constraints through penalty terms. In the large-dipole limit ($r \rightarrow r_{\text{max}}$), we implement saturation unitarity requirement by enforcing the black-disk limit $N(r_{\text{max}}, Y) \rightarrow 1$ [72, 73]. In the small-dipole limit ($r \rightarrow r_{\text{min}}$), color transparency requires $N(r, Y) \propto r^{2\gamma}$ [74] with γ the anomalous dimension. Based on the Bochner theorem [75], Fourier positivity of the dipole amplitude implies the necessary condition $0 < \gamma \leq 1$ [26]. Rather than fixing a power-law ansatz for $N(r, Y)$, we impose this condition on the local logarithmic slope, $\gamma = \partial \ln N(r, Y) / \partial \ln r^2$ by requiring $0 < \gamma \leq 1$ in the small-dipole region $r < 10^{-3} \text{ GeV}^{-1}$ for $Y \geq \ln(0.03/0.01) \simeq 1.1$. Since this anomalous-dimension bound is necessary but not sufficient for Fourier positivity, we also directly penalize negative values of the momentum-space dipole S -matrix,

$$S(k_T, Y) = \int d^2 r e^{i\mathbf{k}_T \cdot \mathbf{r}} [1 - N(r, Y)]. \quad (2)$$

The total objective in our PINN now reads

$$\mathcal{L}(\theta, \mathbf{p}) = w_1 \mathcal{L}_{\text{ciBK}} + w_2 \mathcal{L}_{\text{data}} + w_3 \mathcal{L}_{\text{phy}}, \quad (3)$$

where w_i are numerical balancing factors controlling the relative importance of each term. Their values are determined through gradient-based loss-weight calibration, as described in Sec. A3 of the Supplemental Material. The $\mathcal{L}_{\text{ciBK}} = \|\mathcal{R}_{\text{ciBK}}\|^2$ is the ciBK residual defined in Eq. (1), $\mathcal{L}_{\text{data}}$ quantifies agreement with the experimental data, and $\mathcal{L}_{\text{phy}}$ encodes additional physical constraints. Optimization is performed with the AdamW algorithm [76] together with a cosine-annealing learning-rate schedule [77], ensuring stable and high-precision convergence. We simultaneously fit the network parameters θ and the physical parameters $\mathbf{p} = \{R_p, K_{\text{factor}}^{\text{VM}}, C^2\}$. Here R_p is the transverse proton size, entering the normalization through $\sigma_0/2 = \pi R_p^2$ with $\sigma_{\text{dip}}(r, x_B) = \sigma_0 N(r, x_B)$; $K_{\text{factor}}^{\text{VM}}$ multiplies the exclusive vector-meson cross section to absorb residual uncertainties in the non-perturbative J/ψ wave function; and C^2 fixes the running-coupling scale $\mu^2 = 4C^2/r^2$ in the ciBK kernel. The infrared-frozen coupling is fixed to $\alpha_{\text{fr}} = 0.7$. In this framework, the effective anomalous dimension is determined locally from gradients of $\mathcal{N}_\theta(r, Y)$, rather than treated as a fixed constant. Uncertainties are quantified with a hybrid strategy combining the Monte

Carlo replica method and a deep ensemble approach [47]. To propagate experimental errors, we generate statistically independent pseudodata replicas by sampling each data point from a Gaussian distribution centered on the measured central value with a width given by its uncertainty, and train an independent network for each replica. Each network outputs both the continuous amplitude $N(r, Y)$ and a data-dependent predictive variance needed to construct theory predictions. We therefore take $\mathcal{L}_{\text{data}}$ to be a negative log-likelihood, which naturally incorporates heteroscedastic (aleatoric) uncertainties. This yields a robust uncertainty band for the extracted universal amplitude $N(r, Y)$. Detailed expressions for $\mathcal{L}_{\text{data}}$ and $\mathcal{L}_{\text{phy}}$, together with a full description of the PINN framework, are provided in the Supplemental Material.

Results and Discussion. The global fit includes the HERA reduced cross section σ_r at all available energies $\sqrt{s} = 224.9, 251.5, 300.3$, and 318 GeV [62, 63], the reduced charm cross section $\sigma_r^{c\bar{c}}$ at $\sqrt{s} = 318$ GeV [64], and exclusive diffractive J/ψ photoproduction [65–71]. Unless stated otherwise, we restrict to the measurements with $x_B \leq 0.004$ and $2 < Q^2 < 45 \text{ GeV}^2$.

Figure 1 shows representative results at $\sqrt{s} = 318$ GeV for σ_r and $\sigma_r^{c\bar{c}}$. The predictions obtained from the best PINN-extracted amplitude describe the reduced total cross-section data with $\chi^2/N_{\text{dof}} = 1.482$, while maintaining a good description of the charm reduced cross section with $\chi^2/N_{\text{dof}} = 1.569$. The same joint optimization simultaneously determines the phenomenological parameters (deep-ensemble 1σ) as $R_p = 0.870 \pm 0.048 \text{ fm}$, $K_{\text{factor}}^{\text{VM}} = 1.079 \pm 0.020$ and $C^2 = 0.955 \pm 0.225$. Reconciling DIS reduced total and charm observables within small- x evolution frameworks is known to be challenging. Recent analyses [35, 36] emphasize that charm production probes short dipole distances and therefore provides particularly stringent constraints on the short-distance behavior and on the profile entering the evolution. Our simultaneous fit yields a universal dipole amplitude together with physically reasonable parameter values. Further detailed fit results for σ_r at various center-of-mass energies, as well as the evaluation of the longitudinal structure function F_L serving as a crucial cross-check, are provided in the Supplemental Material.

The extracted amplitude also describes diffractive J/ψ production within the same global framework. Figure 2 shows the total cross section for $\gamma p \rightarrow J/\psi p$ as a function of the photon-proton energy W , compared with experimental data [65–71]. The PINN prediction provides an excellent description, with $\chi^2/N_{\text{dof}} = 1.134$. The overall normalization is fixed to $K_{\text{factor}}^{\text{VM}} = 1.079$ to account for uncertainties in the vector-meson wave function, and no additional process-dependent initial conditions are introduced. This offers a nontrivial universality check of the inferred $N(r, x_B)$ across inclusive DIS (including charm) and exclusive vector-meson photoproduction.

Figure 3 shows the extracted dipole amplitude $N(r, x_B)$ at three representative x_B values. As a consistency check, we solve the ciBK equation numerically

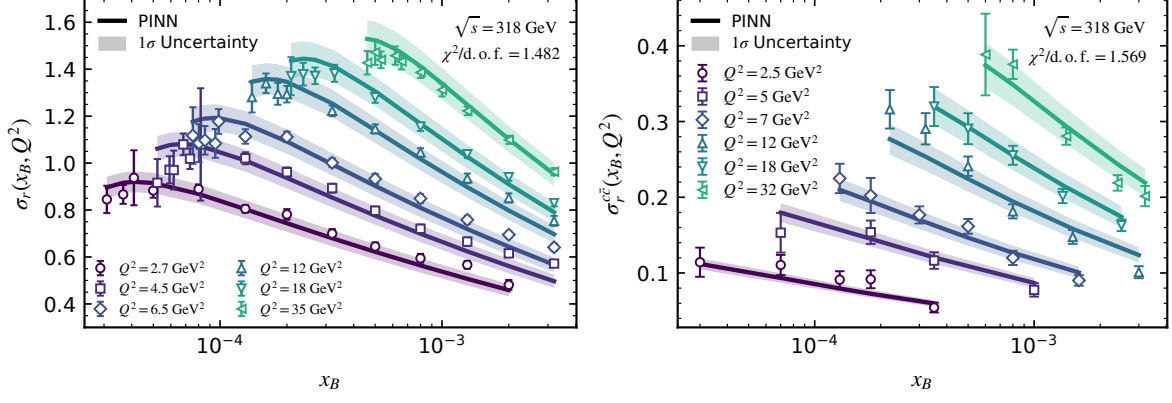


FIG. 1. The reduced total and charm cross sections, σ_r and σ_r^{cc} , are shown vs. x_B in selected Q^2 bins at $\sqrt{s} = 318$ GeV, compared with HERA data [62–64]. Solid curves are the best-fit PINN prediction (minimum global $\chi^2/\text{d.o.f.}$); bands show 1σ uncertainties. The quoted $\chi^2/\text{d.o.f.}$ uses the solid curve; for σ_r it includes all available energies.

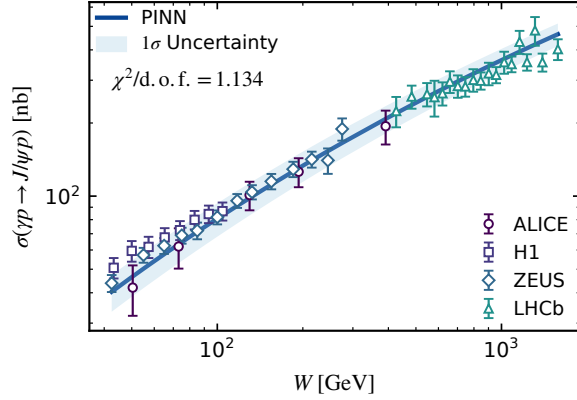


FIG. 2. Total cross section for exclusive J/ψ photoproduction as a function of center-of-mass energy W , compare to ALICE [65, 66], H1 [67, 68], ZEUS [69] and LHCb [70, 71] data. Solid curve is the PINN prediction with the smallest global $\chi^2/\text{d.o.f.}$; shaded band is 1σ uncertainties. The quoted $\chi^2/\text{d.o.f.}$ value is evaluated using the solid curve.

using $N(r, x_B = x_0) = N_{\text{PINN}}(r, x_B = x_0)$ with $x_0 = 0.03$ as the initial condition to consistently accommodate the non-local rapidity shift in the evolution kernel, and compare the evolved solutions with PINN output. We observe excellent agreement throughout the kinematic regions. Once trained, the surrogate evaluates $N(r, x_B)$ at any desired x_B in a single forward pass, avoiding repeated and computationally expensive grid-based ciBK evolutions in global analyses. The inferred pre-asymptotic shape, together with the momentum-space positivity regularization, is further illustrated by comparing $N_{\text{PINN}}(r, x_B = 0.01)$, a standard reference scale in saturation phenomenology, with an MV-type parametrization [32, 33],

$$N_{\text{MV-type}}(r) = 1 - \exp \left[-\frac{(r^2 Q_{s0}^2)^\gamma}{4} \ln \left(\frac{1}{r\Lambda} + e_c \cdot e \right) \right]. \quad (4)$$

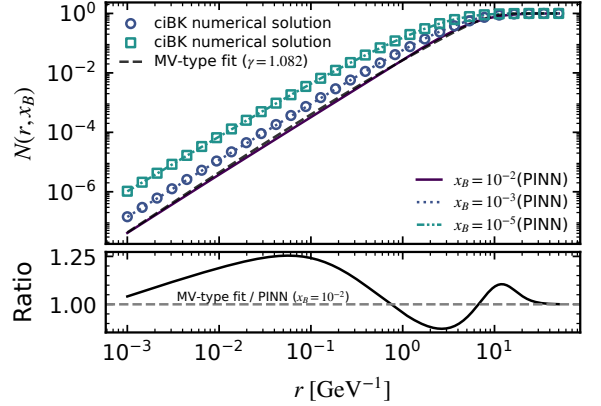


FIG. 3. Top: Solid curves show the dipole amplitude $N(r, x_B)$ predicted by the PINN for $x_B \in \{10^{-2}, 10^{-3}, 10^{-5}\}$. Markers denote numerical ciBK evolution results obtained with the same initial condition as used in the PINN. The dashed curve corresponds to the best-fit MV-type parametrization to $N_{\text{PINN}}(r, x_B = 0.01)$.

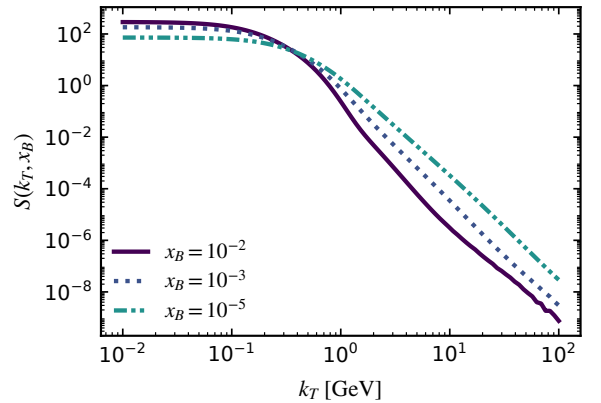


FIG. 4. Fourier transform of the dipole amplitude at $x_B \in \{10^{-2}, 10^{-3}, 10^{-5}\}$.

Fitting Eq. (4) to the PINN profile at $x_B = 0.01$ yields $(\gamma, Q_{s0}^2, \Lambda, e_c) = (1.082, 0.080 \text{ GeV}^2, 0.283 \text{ GeV}, 0.649)$. The ratio of the fitted MV-type form to the PINN result, shown in the bottom panel of Figure 3, exhibits systematic deviations of up to $\sim 25\%$ over the range $r \in [10^{-2}, 10^{-1}] \text{ GeV}^{-1}$. This indicates that the extracted initial dipole amplitude cannot be faithfully captured by the MV-type functional form. By moving beyond rigid parametrizations, the non-parametric PINN extraction provides the flexibility needed to capture the dipole-amplitude structure at both short and long transverse distances, which is essential for a simultaneous description of light- and heavy-quark observables and the momentum-space positivity requirement.

Finally, we examine the two-dimensional Fourier transform of the dipole S -matrix defined in Eq. (2). Momentum-space positivity is essential: within the CGC framework, the momentum-space dipole is directly connected to the gluon content of the target and enters a variety of particle-production cross sections in e - p/A and p - p/A collisions. The sign changes in the momentum-space dipole have been reported for several widely used initial conditions, even when the DIS reduced cross section is well described [22, 23, 25, 31, 36]. Although BK evolution generally mitigates these issues, they can persist in phenomenologically relevant kinematics. Figure 4 shows $S(k_T, x_B)$ at three representative Bjorken- x_B values. In our PINN global fit we impose a positivity regularization by penalizing regions where $S(k_T, x_B) < 0$. In the region ($x_B \leq 0.01$), the extracted PINN solution yields a smooth k_T dependence and remains positive over the range shown, $k_T \in [0, 100] \text{ GeV}$. This demonstrates that momentum-space positivity can be achieved simultaneously with a good description of inclusive DIS reduced cross section (including total and charm quark) and diffractive vector-meson production, provided sufficient functional flexibility and an explicit positivity regularization are included. We have also checked that relaxing the $S(k_T) > 0$ constraint does yield a smaller $\chi^2/\text{d.o.f.}$ but at the price of a positivity-violating S -matrix. The constraint therefore refines the extraction, yielding a dipole amplitude that is at once data-driven and physically admissible. For comparison, Ref. [36] reports oscillatory behavior in $S(k_T, x_B)$ even at $x_B = 10^{-3}$ and 10^{-6} using a Bayesian-fit approach.

Conclusion. We have performed a global and dynamics-constrained determination of the universal small- x dipole amplitude $N(r, x_B)$ using a physics-informed neural network. In our approach, the amplitude is represented by a smooth surrogate $\mathcal{N}_\theta(r, x_B)$ and is constrained directly at the functional level by the ciBK evolution equation together with precision DIS and diffractive data. The initial profile is inferred non-parametrically from data while remaining consistent with ciBK dynamics. Within the fitted kinematic region, the resulting universal amplitude simultaneously describes the reduced total cross section, the

reduced charm cross section, and exclusive J/ψ photo-production. Another central outcome is a well-behaved momentum-space representation: by incorporating the Fourier-positivity penalty on the momentum-space S -matrix, we obtain a PINN solution whose $S(k_T, x_B)$ remains non-negative for $x_B \leq 0.01$ over the transverse-momentum range shown. This addresses a long-standing issue in dipole-model initial conditions and enables consistent use of the extracted amplitude in downstream CGC calculations requiring a positive momentum-space input. It also motivates going beyond the McLerran-Venugopalan parametrization as an initial condition for the proton.

The present study employs an impact-parameter independent dipole amplitude and leading-order impact factors, and we restrict the fit to $x_B \leq 0.004$ experimental data to suppress potential DGLAP-dominated contributions and reduce sensitivity to pre-asymptotic effects associated with the rapidity-shift non-locality. Natural next steps are to add impact-parameter dependence and NLO impact factors with NLO + NLL evolution; beyond the mean-field BK level, the LO [78–80] and NLO JIMWLK [81, 82] equations can also be solved with the score-based method and linked to our PINN framework. Together, these would enable more stringent universality tests and direct confrontation with forthcoming EIC precision data.

More broadly, our framework moves beyond a purely “data-driven” fit toward “physics-driven” inference. PINN acts as a differentiable surrogate that simultaneously enforces the evolution equation, accommodates precise measurements, and incorporates first-principles constraints such as unitarity and color transparency. By combining these complementary sources of information, the network is guided not only by statistical agreement with data but also by the underlying QCD dynamics, helping ensure that the extracted amplitude remains physically consistent across the full kinematic domain. This strategy is inherently generalizable and offers a robust template for embedding fundamental laws into data-driven inference in scientific settings where dynamical equations and first-principles constraints coexist with experimental measurements.

We have made available the tabulated PINN solution for $N(r, x_B)$ over the fitted domain, together with its momentum-space Fourier transform, in Ref. [83].

ACKNOWLEDGMENTS

Acknowledgments We thank Haowu Duan, Heikki Mäntysaari, Farid Salazar, and Dionysis N. Triantafyllopoulos for helpful comments. This work was supported in part by the NSFC under grant No. 12225503, No. 12535010, No. 12435009, No. 12405156, and by the Outstanding Leading Talent Team Program of Central China Normal University (XJ2026000302). F.P. L acknowledges support from the NSFC under grant

No. 12325507, No. 12547102, and No. 12147101, and the National Key Research and Development Program of China under grant No. 2022YFA1604900. This work was supported also in part by the Cross Research Project

of Fundamental Research Funds for Central Universities of Central China Normal University in 2025: “Advanced Detection and Artificial Intelligence at the Frontiers of Physics” (No.30101250317).

-
- [1] Y. V. Kovchegov and E. Levin, *Quantum Chromodynamics at High Energy*, vol. 33 (Oxford University Press, 2013), ISBN 978-1-009-29144-6, 978-1-009-29141-5, 978-1-009-29142-2, 978-0-521-11257-4, 978-1-139-55768-9.
 - [2] L. McLerran, Nucl. Phys. A **702**, 49 (2002).
 - [3] E. Iancu and R. Venugopalan, *The Color glass condensate and high-energy scattering in QCD* (2003), pp. 249–3363, hep-ph/0303204.
 - [4] H. Weigert, Prog. Part. Nucl. Phys. **55**, 461 (2005), hep-ph/0501087.
 - [5] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, Ann. Rev. Nucl. Part. Sci. **60**, 463 (2010), 1002.0333.
 - [6] J. L. Albacete and C. Marquet, Prog. Part. Nucl. Phys. **76**, 1 (2014), 1401.4866.
 - [7] A. Morreale and F. Salazar, Universe **7**, 312 (2021), 2108.08254.
 - [8] J. Jalilian-Marian, A. Kovner, L. D. McLerran, and H. Weigert, Phys. Rev. D **55**, 5414 (1997), hep-ph/9606337.
 - [9] J. Jalilian-Marian, A. Kovner, A. Leonidov, and H. Weigert, Nucl. Phys. B **504**, 415 (1997), hep-ph/9701284.
 - [10] J. Jalilian-Marian, A. Kovner, A. Leonidov, and H. Weigert, Phys. Rev. D **59**, 014014 (1998), hep-ph/9706377.
 - [11] E. Iancu and L. D. McLerran, Phys. Lett. B **510**, 145 (2001), hep-ph/0103032.
 - [12] E. Ferreira, E. Iancu, A. Leonidov, and L. McLerran, Nucl. Phys. A **703**, 489 (2002), hep-ph/0109115.
 - [13] E. Iancu, A. Leonidov, and L. D. McLerran, Phys. Lett. B **510**, 133 (2001), hep-ph/0102009.
 - [14] E. Iancu, A. Leonidov, and L. D. McLerran, Nucl. Phys. A **692**, 583 (2001), hep-ph/0011241.
 - [15] I. Balitsky, Nucl. Phys. B **463**, 99 (1996), hep-ph/9509348.
 - [16] Y. V. Kovchegov, Phys. Rev. D **60**, 034008 (1999), hep-ph/9901281.
 - [17] H. Kowalski, T. Lappi, and R. Venugopalan, Phys. Rev. Lett. **100**, 022303 (2008), 0705.3047.
 - [18] R. Abdul Khalek et al., Nucl. Phys. A **1026**, 122447 (2022), 2103.05419.
 - [19] E. C. Aschenauer, S. Fazio, J. H. Lee, H. Mäntysaari, B. S. Page, B. Schenke, T. Ullrich, R. Venugopalan, and P. Zurita, Rept. Prog. Phys. **82**, 024301 (2019), 1708.01527.
 - [20] A. Accardi et al., Eur. Phys. J. A **52**, 268 (2016), 1212.1701.
 - [21] J. L. Albacete, N. Armesto, J. G. Milhano, and C. A. Salgado, Phys. Rev. D **80**, 034031 (2009), 0902.1112.
 - [22] J. L. Albacete, N. Armesto, J. G. Milhano, P. Quiroga-Arias, and C. A. Salgado, Eur. Phys. J. C **71**, 1705 (2011), 1012.4408.
 - [23] T. Lappi and H. Mäntysaari, Phys. Rev. D **88**, 114020 (2013), 1309.6963.
 - [24] H. Mäntysaari and B. Schenke, Phys. Rev. D **98**, 034013 (2018), 1806.06783.
 - [25] B. Ducloué, E. Iancu, G. Soyez, and D. N. Triantafyllopoulos, Phys. Lett. B **803**, 135305 (2020), 1912.09196.
 - [26] B. G. Giraud and R. Peschanski, Phys. Lett. B **760**, 26 (2016), 1604.01932.
 - [27] F. Gelis and R. Venugopalan, Phys. Rev. D **69**, 014019 (2004), hep-ph/0310090.
 - [28] J. P. Blaizot, F. Gelis, and R. Venugopalan, Nucl. Phys. A **743**, 13 (2004), hep-ph/0402256.
 - [29] J. P. Blaizot, F. Gelis, and R. Venugopalan, Nucl. Phys. A **743**, 57 (2004), hep-ph/0402257.
 - [30] A. Dumitru, A. Hayashigaki, and J. Jalilian-Marian, Nucl. Phys. A **765**, 464 (2006), hep-ph/0506308.
 - [31] P. Caucal, Z.-B. Kang, P. Korcyl, F. Salazar, B. Schenke, T. Stebel, R. Venugopalan, and W. Zhao (2025), 2512.21466.
 - [32] L. D. McLerran and R. Venugopalan, Phys. Rev. D **49**, 2233 (1994), hep-ph/9309289.
 - [33] L. D. McLerran and R. Venugopalan, Phys. Rev. D **49**, 3352 (1994), hep-ph/9311205.
 - [34] L. D. McLerran and R. Venugopalan, Phys. Rev. D **50**, 2225 (1994), hep-ph/9402335.
 - [35] H. Hänninen, H. Mäntysaari, R. Paatelainen, and J. Penttala, Phys. Rev. Lett. **130**, 192301 (2023), 2211.03504.
 - [36] C. Casuga, H. Hänninen, and H. Mäntysaari, Phys. Rev. D **112**, 034003 (2025), 2506.00487.
 - [37] C. Casuga, M. Karhunen, and H. Mäntysaari, Phys. Rev. D **109**, 054018 (2024), 2311.10491.
 - [38] M. Gao, Z.-B. Kang, J. Penttala, and D. Y. Shao (2025), 2510.26779.
 - [39] W. Kou and X. Chen (2026), 2601.16391.
 - [40] M. Raissi, P. Perdikaris, and G. E. Karniadakis, J. Comput. Phys. **378**, 686 (2019), 1711.10561.
 - [41] L. von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Gieselsbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy, et al., IEEE Transactions on Knowledge and Data Engineering **35**, 614 (2023).
 - [42] S. Cuomo, V. S. D. Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, CoRR **abs/2201.05624** (2022), 2201.05624.
 - [43] G. Beuf, Phys. Rev. D **89**, 074039 (2014), 1401.0313.
 - [44] E. Iancu, J. D. Madrigal, A. H. Mueller, G. Soyez, and D. N. Triantafyllopoulos, Phys. Lett. B **744**, 293 (2015), 1502.05642.
 - [45] E. Iancu, J. D. Madrigal, A. H. Mueller, G. Soyez, and D. N. Triantafyllopoulos, Phys. Lett. B **750**, 643 (2015), 1507.03651.
 - [46] B. Ducloué, E. Iancu, A. H. Mueller, G. Soyez, and D. N. Triantafyllopoulos, JHEP **04**, 081 (2019), 1902.06637.
 - [47] B. Lakshminarayanan, A. Pritzel, and C. Blundell, *Simple and scalable predictive uncertainty estimation using deep ensembles* (2017), 1612.01474.
 - [48] E. Iancu, J. D. Madrigal, A. H. Mueller, G. Soyez, and

- D. N. Triantafyllopoulos, Nucl. Phys. A **956**, 557 (2016), 1601.06525.
- [49] E. Iancu, A. H. Mueller, D. N. Triantafyllopoulos, and S. Y. Wei, JHEP **07**, 196 (2021), 2012.08562.
- [50] W. Li and A. M. Stasto, Eur. Phys. J. C **82**, 562 (2022), 2201.10579.
- [51] K. Kutak, W. Li, A. Stasto, and R. Straka, Phys. Rev. D **111**, 116021 (2025), 2502.15465.
- [52] T. Lappi and H. Mäntysaari, Phys. Rev. D **93**, 094004 (2016), 1601.06598.
- [53] J. L. Albacete, J. G. Milhano, P. Quiroga-Arias, and J. Rojo, Eur. Phys. J. C **72**, 2131 (2012), 1203.1043.
- [54] P. Agostini et al. (LHeC, FCC-he Study Group), J. Phys. G **48**, 110501 (2021), 2007.14491.
- [55] K. J. Golec-Biernat and M. Wusthoff, Phys. Rev. D **59**, 014017 (1998), hep-ph/9807513.
- [56] H. Kowalski and D. Teaney, Phys. Rev. D **68**, 114005 (2003), hep-ph/0304189.
- [57] J. R. Forshaw, R. Sandapen, and G. Shaw, Phys. Rev. D **69**, 094013 (2004), hep-ph/0312172.
- [58] H. Kowalski, L. Motyka, and G. Watt, Phys. Rev. D **74**, 074016 (2006), hep-ph/0606272.
- [59] A. G. Shuvaev, K. J. Golec-Biernat, A. D. Martin, and M. G. Ryskin, Phys. Rev. D **60**, 014015 (1999), hep-ph/9902410.
- [60] T. Lappi and H. Mäntysaari, Phys. Rev. C **87**, 032201 (2013), 1301.4095.
- [61] H. Mäntysaari, H. Roch, F. Salazar, B. Schenke, C. Shen, and W. Zhao, Phys. Rev. D **113**, 014038 (2026), 2507.14087.
- [62] H. Abramowicz et al. (H1, ZEUS), Eur. Phys. J. C **75**, 580 (2015), 1506.06042.
- [63] F. D. Aaron et al. (H1, ZEUS), JHEP **01**, 109 (2010), 0911.0884.
- [64] H. Abramowicz et al. (H1, ZEUS), Eur. Phys. J. C **78**, 473 (2018), 1804.01019.
- [65] B. B. Abelev et al. (ALICE), Phys. Rev. Lett. **113**, 232504 (2014), 1406.7819.
- [66] S. Acharya et al. (ALICE), Eur. Phys. J. C **79**, 402 (2019), 1809.03235.
- [67] A. Aktas et al. (H1), Eur. Phys. J. C **46**, 585 (2006), hep-ex/0510016.
- [68] C. Alexa et al. (H1), Eur. Phys. J. C **73**, 2466 (2013), 1304.5162.
- [69] S. Chekanov et al. (ZEUS), Eur. Phys. J. C **24**, 345 (2002), hep-ex/0201043.
- [70] R. Aaij et al. (LHCb), JHEP **10**, 167 (2018), 1806.04079.
- [71] R. Aaij et al. (LHCb), SciPost Phys. **18**, 071 (2025), 2409.03496.
- [72] J. L. Albacete, N. Armesto, J. G. Milhano, C. A. Salgado, and U. A. Wiedemann, Phys. Rev. D **71**, 014003 (2005), hep-ph/0408216.
- [73] C. Marquet, R. B. Peschanski, and G. Soyez, Phys. Lett. B **628**, 239 (2005), hep-ph/0509074.
- [74] S. Munier and R. B. Peschanski, Phys. Rev. Lett. **91**, 232001 (2003), hep-ph/0309177.
- [75] S. Bochner, *Lectures on Fourier Integrals*, vol. 42 of *Annals of Mathematics Studies* (Princeton University Press, 1959), originally published as *Vorlesung über Fouriersche Integrale*, Leipzig Verlag, 1932.
- [76] I. Loshchilov and F. Hutter, in *International Conference on Learning Representations* (2019).
- [77] I. Loshchilov and F. Hutter, CoRR **abs/1608.03983** (2016), 1608.03983.
- [78] A. H. Mueller, Phys. Lett. B **523**, 243 (2001), hep-ph/0110169.
- [79] T. Lappi and H. Mäntysaari, Eur. Phys. J. C **73**, 2307 (2013), 1212.4825.
- [80] S. Schlichting and B. Schenke, Phys. Lett. B **739**, 313 (2014), 1407.8458.
- [81] A. Kovner, M. Lublinsky, and Y. Mulian, Phys. Rev. D **89**, 061704 (2014), 1310.0378.
- [82] A. Kovner, M. Lublinsky, and Y. Mulian, JHEP **08**, 114 (2014), 1405.0418.
- [83] S. Dai, F. Li, L.-G. Pang, G. Qin, S.-y. Wei, H. Zhang, and W. Zhao, *Physics-informed global extraction of the universal small- x dipole amplitude* (2026), URL <https://doi.org/10.5281/zenodo.20794812>.
- [84] S. J. Brodsky, G. P. Lepage, and P. B. Mackenzie, Phys. Rev. D **28**, 228 (1983).
- [85] F. Takahashi et al. (Particle Data Group), Int. J. Mod. Phys. A **41**, 2630011 (2026).
- [86] K. He, X. Zhang, S. Ren, and J. Sun, CoRR **abs/1512.03385** (2015), 1512.03385.
- [87] S. Wang, Y. Teng, and P. Perdikaris, SIAM Journal on Scientific Computing **43**, A3055 (2021).
- [88] Y. Wang, Y. Yao, J. Guo, and Z. Gao, Journal of Computational Physics **510**, 113112 (2024), ISSN 0021-9991.
- [89] Z. Chen, V. Badrinarayanan, C. Lee, and A. Rabinovich, CoRR **abs/1711.02257** (2017), 1711.02257.
- [90] V. Andreev et al. (H1), Eur. Phys. J. C **74**, 2814 (2014), 1312.4821.

I. SUPPLEMENTAL MATERIAL

A. Physics-Informed Framework for Dipole Evolution

This Supplementary Material section provides the details about the PINN framework used to extract the universal dipole scattering amplitude. We first outline the collinearly-improved Balitsky-Kovchegov (ciBK) equation that governs the small- x evolution and the running coupling, and then detail the neural network architecture along with the formulation of the loss function.

A1. The Collinearly-Improved Balitsky-Kovchegov Equation and the Running Coupling

The small- x evolution of the dipole scattering amplitude $N(r, Y)$ is governed by the Balitsky-Kovchegov (BK) equation. To achieve high-precision phenomenology compatible with experimental data, we employ the collinearly-improved BK (ciBK) equation, which resums large collinear logarithms to all orders and enforces exact kinematic constraints. The evolution equation in rapidity Y reads [25, 46]

$$\begin{aligned} \frac{\partial N(r, Y)}{\partial Y} = & \frac{\bar{\alpha}_{\text{BLM}}(r)}{2\pi} \int d^2\mathbf{r}_1 \frac{r^2}{r_1^2 r_2^2} \left[\frac{r^2}{\min(r_1^2, r_2^2)} \right]^{\pm A_1 \bar{\alpha}_{\text{BLM}}} \\ & \times \mathcal{K}_{\text{DLA}}(\rho) \left[N(r_1, Y - \delta_1) + N(r_2, Y - \delta_2) \right. \\ & \left. - N(r, Y) - N(r_1, Y - \delta_1)N(r_2, Y - \delta_2) \right], \end{aligned} \quad (\text{S1})$$

where $\mathbf{r}_2 = \mathbf{r} - \mathbf{r}_1$ and the evolution kernel factorizes into three components: the leading-order dipole kernel supplemented by the running-coupling prescription, the DLA factor resumming the double collinear logarithms to all orders, and the single transverse logarithmic (STL) correction factor. Here $\bar{\alpha}_{\text{BLM}}$ denotes the Brodsky-Lepage-Mackenzie (BLM) running coupling [84], whose explicit three-dipole form is given in Eq. (S5). These four elements are defined as follows.

(i) The kinematic constraints (rapidity shifts):

The lifetime ordering of successive soft-gluon emissions, required for the consistency of the evolution at small x , is enforced through the rapidity shifts entering the arguments of the amplitudes in the nonlinear functional,

$$\delta_i = \max \left\{ 0, \ln \frac{r^2}{r_i^2} \right\}, \quad i = 1, 2, \quad (\text{S2})$$

where r_1 and r_2 denote the sizes of the two daughter dipoles produced in the emission and r is the size of the parent dipole. These constraints relate the rapidity available to the daughter dipoles to their transverse sizes: an emission into a much smaller daughter dipole incurs a finite rapidity shift, suppressing configurations that would

otherwise violate time ordering. In this way the shifts resum the dominant double collinear logarithms generated by strictly time-ordered emissions to all orders, while leaving the conventional BK kernel structure unchanged.

(ii) The double logarithmic approximation (DLA) factor.

In a formulation without the kinematic constraints, the same double collinear logarithms are resummed in closed form by the DLA factor, given by the Bessel function

$$\mathcal{K}_{\text{DLA}}(\rho) = \frac{J_1(2\sqrt{\bar{\alpha}_s \rho^2})}{\sqrt{\bar{\alpha}_s \rho^2}} = 1 - \frac{\bar{\alpha}_s \rho^2}{2} + \frac{(\bar{\alpha}_s \rho^2)^2}{12} + \dots, \quad (\text{S3})$$

where $\rho = \sqrt{L_{r_1 r} L_{r_2 r}}$ with $L_{r_i r} = \ln(r_i^2/r^2)$. Expanding the shifted amplitudes $N(r_i, Y - \delta_i)$ in powers of $\bar{\alpha}_s$ reproduces, order by order, the same double-logarithmic tower $(\bar{\alpha}_s \rho^2)^n$ that is resummed by $\mathcal{K}_{\text{DLA}}(\rho)$. The rapidity shifts δ_i and the DLA factor therefore constitute two equivalent prescriptions for resumming the same double collinear logarithms, and retaining both simultaneously would double-count these contributions. Since the double-logarithmic resummation is already fully carried by the kinematic constraints δ_i , we keep only the leading term, $\mathcal{K}_{\text{DLA}}(\rho) = 1$. This choice stabilizes the evolution speed and avoids the spurious oscillations associated with the higher-order terms of the Bessel expansion.

(iii) The single transverse logarithmic (STL) correction factor.

Beyond the double collinear logarithms, the next-to-leading-order corrections to the BK kernel also generate single transverse logarithms, $\ln(r^2/r_i^2)$, which are not captured by the time-ordering constraints. These are resummed to all orders by the STL correction factor,

$$\mathcal{K}_{\text{STL}} = \left[\frac{r^2}{\min(r_1^2, r_2^2)} \right]^{\pm A_1 \bar{\alpha}_{\text{BLM}}}, \quad A_1 = \frac{11}{12}, \quad (\text{S4})$$

where the coefficient A_1 originates from the one-loop DGLAP anomalous dimension. The sign in the exponent is fixed by the relative size of the dipoles: the plus sign applies when $r^2 < \min(r_1^2, r_2^2)$ and the minus sign when $r^2 > \min(r_1^2, r_2^2)$, so that the factor consistently suppresses the kernel in the region of strongly ordered dipole sizes. Together with the kinematic constraints and the leading DLA term, this factor completes the collinear improvement of the BK kernel adopted in the present analysis.

(iv) The running coupling prescription.

The running-coupling corrections form an important and well-understood class of the NLO contributions to the BK equation, being the subset enhanced by the one-loop β -function. Among the prescriptions proposed in the literature, we adopt the BLM prescription (also dubbed as “fast apparent convergence” [45]). In this prescription, the effective coupling entering the kernel is written as a smooth interpolating coupling $\bar{\alpha}_{\text{BLM}}(r, r_1, r_2)$, which depends on the sizes of all three dipoles involved in the splitting. The detailed BLM for-

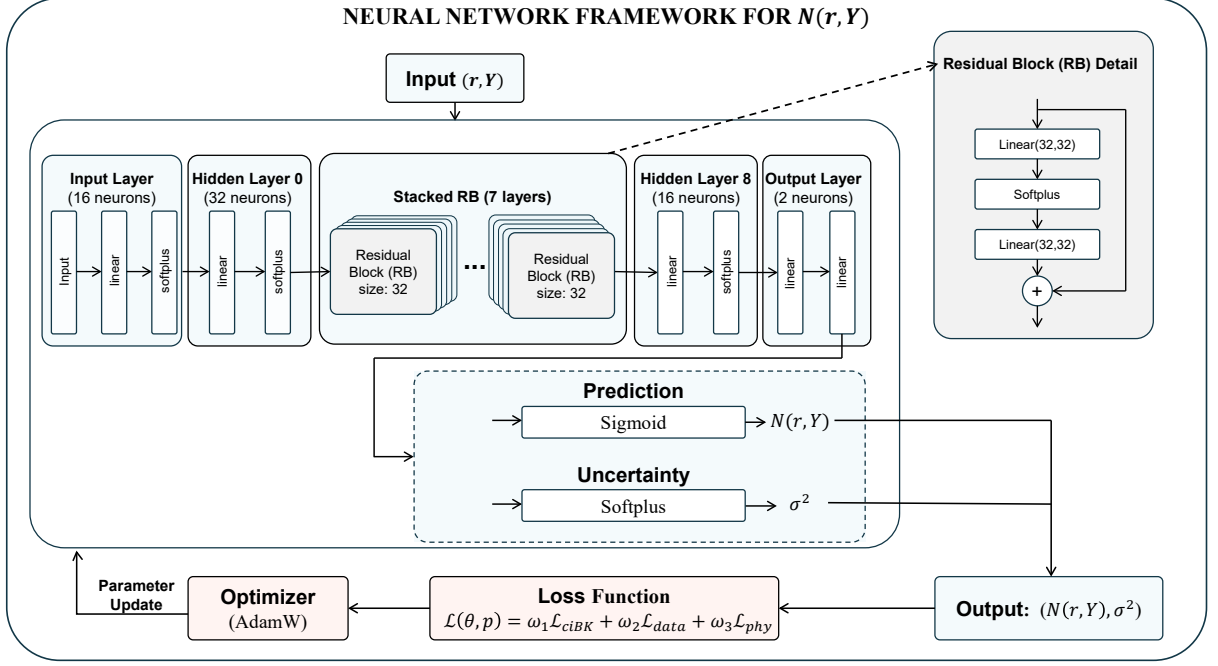


FIG. S1. Schematic of the PINN framework. The network maps the kinematic inputs (r, Y) to the dipole amplitude $N(r, Y)$. Training is guided by the total loss $\mathcal{L}(\theta, \mathbf{p}) = w_1 \mathcal{L}_{\text{ciBK}} + w_2 \mathcal{L}_{\text{data}} + w_3 \mathcal{L}_{\text{phy}}$. See the text for details of the three loss components.

mula is the Eq. (S5). This prescription absorbs the dominant β_0 -dependent part of the NLO corrections into the kernel and effectively selects the smallest dipole as the relevant scale, thereby minimizing the residual higher-order corrections. It is defined as

$$\bar{\alpha}_{\text{BLM}}(r, r_1, r_2) = \left[\frac{1}{\bar{\alpha}_s(r)} + \frac{r_1^2 - r_2^2}{r^2} \frac{\bar{\alpha}_s(r_1) - \bar{\alpha}_s(r_2)}{\bar{\alpha}_s(r_1)\bar{\alpha}_s(r_2)} \right]^{-1}, \quad (\text{S5})$$

with $\bar{\alpha}_s(r) \equiv N_c \alpha_s(r)/\pi$. In any limit where one of the three dipoles is much smaller than the other two, Eq.(S5) reduces to the minimal-dipole choice $\bar{\alpha}_s(r_{\min})$.

The one-loop coupling $\alpha_s(r)$ is evaluated at the coordinate-space scale $\mu^2 = 4C^2/r^2$, where the dimensionless constant C originates from the Fourier transform between coordinate and momentum space and is treated as a free parameter in the global fit. The coupling incorporates heavy-quark flavor thresholds: the number of active flavors increases from $n_f = 3$ to 4 and to 5 as μ^2 crosses m_c^2 and m_b^2 , respectively. To regularize the Landau pole in the infrared, we freeze the coupling through a smooth shift λ_0 [72],

$$\alpha_s(r) = \frac{1}{\beta_0^{(n_f)} \ln \left(\frac{4C^2}{r^2 \Lambda_{(n_f)}^2} + \lambda_0 \right)}, \quad \beta_0^{(n_f)} = \frac{11N_c - 2n_f}{12\pi}. \quad (\text{S6})$$

with $\lambda_0 = \exp[1/(\beta_0^{(3)} \times \alpha_{\text{fr}})]$, which ensures that the coupling saturates smoothly to a maximal value α_{fr} in the deep infrared. Continuity of $\alpha_s(r)$ across the thresholds $\mu^2 = m_f^2$ ($f = c, b$) is enforced by recursively matching

the scale parameters,

$$\Lambda_{(n_f-1)} = m_f \left[\left(\frac{m_f^2}{\Lambda_{(n_f)}^2} + \lambda_0 \right)^{\frac{\beta_0^{(n_f)}}{\beta_0^{(n_f-1)}}} - \lambda_0 \right]^{-1/2}. \quad (\text{S7})$$

The matching is anchored at the Z pole, where $\Lambda_{(5)}$ is fixed by the world-average $\alpha_s(M_Z)$. Throughout the global extraction we fix $N_c = 3$, $m_c = 1.4$ GeV, $m_b = 4.18$ GeV, $M_Z = 91.2$ GeV, $\alpha_s(M_Z^2) = 0.118$ [85], and the frozen coupling $\alpha_{\text{fr}} = 0.7$ [22].

A2. Physics-Informed Neural Network Framework and Loss Function

Figure S1 illustrates the PINN architecture, which takes the dipole size r and the rapidity $Y = \ln(x_0/x_B)$ as inputs. The network is based on a Residual Neural Network (ResNet) [86], which mitigates vanishing-gradient issues in deep multilayer perceptrons. Our baseline configuration comprises seven residual blocks with 32 neurons each, using the softplus activation $\ln(1 + e^x)$ for the hidden nonlinearities. A final sigmoid activation, $\sigma(x) = 1/(1 + e^{-x})$, constrains the output to $(0, 1)$, ensuring that the dipole amplitude $N(r, Y)$ remains physical.

The optimization objective, i.e., the loss function, is formulated as:

$$\mathcal{L}(\theta, \mathbf{p}) = w_1 \mathcal{L}_{\text{ciBK}} + w_2 \mathcal{L}_{\text{data}} + w_3 \mathcal{L}_{\text{phy}}, \quad (\text{S8})$$

where θ denotes the trainable neural network weights and biases, and $\mathbf{p}$ represents optimizable physical parameters. The weights w_i are numerical balancing factors controlling the relative importance of the different loss terms. They are determined through gradient-level calibration before production training and then held unchanged during the subsequent optimization.

The term $\mathcal{L}_{\text{ciBK}} = \|\mathcal{R}_{\text{ciBK}}\|^2$ is the ciBK residual, which is defined as

$$\mathcal{R}_{\text{ciBK}}(r, Y) = \frac{\partial \mathcal{N}_\theta(r, Y)}{\partial Y} - \int d^2 \mathbf{r}_1 K_{\text{ciBK}}(\mathbf{r}, \mathbf{r}_1, \mathbf{r}_2) \mathcal{F}[\mathcal{N}_\theta]. \quad (\text{S9})$$

Here, the collinear-improved kernel incorporates the BLM prescription along with the resummation factors discussed above:

$$K_{\text{ciBK}}(\mathbf{r}, \mathbf{r}_1, \mathbf{r}_2) = \frac{\bar{\alpha}_{\text{BLM}}(r)}{2\pi} \frac{r^2}{r_1^2 r_2^2} \mathcal{K}_{\text{DLA}}(\rho) \mathcal{K}_{\text{STL}}. \quad (\text{S10})$$

The associated nonlinear functional $\mathcal{F}[\mathcal{N}_\theta]$ enforces the exact kinematic constraints:

$$\mathcal{F}[\mathcal{N}_\theta] = \mathcal{N}_\theta(r_1, Y - \delta_1) + \mathcal{N}_\theta(r_2, Y - \delta_2) - \mathcal{N}_\theta(r, Y) - \mathcal{N}_\theta(r_1, Y - \delta_1) \mathcal{N}_\theta(r_2, Y - \delta_2). \quad (\text{S11})$$

During the network optimization, this continuous residual is evaluated on a set of collocation points generated via two-dimensional Latin hypercube sampling (LHS) over the (r, Y) domain. These points are resampled at the start of each training epoch. Compared with a dense, fixed uniform grid, this dynamic resampling reduces the per-epoch computational cost while ensuring thorough coverage of the continuous kinematic space over the course of training.

The data loss $\mathcal{L}_{\text{data}}$ measures agreement with experimental data. Consistent with the deep-ensemble setup, we use a negative log-likelihood,

$$\mathcal{L}_{\text{data}} = \sum_{\mathcal{O}} \sum_{i=1}^{N_{\mathcal{O}}} \lambda_{\mathcal{O}} \left[\frac{(\mathcal{O}_i^{\text{th}} - \mathcal{O}_i^{\text{exp}})^2}{2((\sigma_i^{\text{NN}})^2 + (\sigma_i^{\text{exp}})^2)} + \frac{1}{2} \ln((\sigma_i^{\text{NN}})^2 + (\sigma_i^{\text{exp}})^2) \right], \quad (\text{S12})$$

where the outer sum runs over observable classes $\mathcal{O} \in \{\sigma_r, \sigma_r^{c\bar{c}}, \sigma^{\text{VM}}\}$, and $\lambda_{\mathcal{O}}$ are dataset-specific weights. These weights are introduced to balance the relative contributions of different observables, preventing the overall data loss from being strictly dominated by datasets with a significantly larger number of data points (e.g., the inclusive σ_r). For the i -th point, $\mathcal{O}_i^{\text{exp}}$ and σ_i^{exp} are the measured central value and total experimental uncertainty. The corresponding prediction $\mathcal{O}_i^{\text{th}}$ is computed from the network output, namely the universal dipole amplitude $N(r, Y)$. The term σ_i^{NN} is the data-dependent predictive uncertainty produced by the network's variance branch during training.

The physical loss $\mathcal{L}_{\text{phy}}$ imposes key theoretical constraints, including the large- r limit, momentum-space

positivity, and the small- r color-transparency behavior. Its explicit form is

$$\begin{aligned} \mathcal{L}_{\text{phy}} = & \lambda_{\text{IR}} \|N(r_{\text{max}}, Y) - 1\|_{r_{\text{max}}=50 \text{ GeV}^{-1}}^2 \\ & + \lambda_{\text{pos}} \|\min(0, S(k_T, Y))\|^2 \\ & + \lambda_{\text{UV}} \left(\left\| \max \left(0, \frac{\partial \ln N(r, Y)}{\partial \ln r^2} - 1 \right) \right\|^2 \right. \\ & \left. + \left\| \min \left(0, \frac{\partial \ln N(r, Y)}{\partial \ln r^2} \right) \right\|^2 \right)_{r < 10^{-3} \text{ GeV}^{-1}}, \end{aligned} \quad (\text{S13})$$

where the penalty terms enforce the following physical behaviors:

- **Asymptotic large- r constraint (Black-disk limit):** Because $r_{\text{max}} = 50 \text{ GeV}^{-1}$ (about 10 fm) is far larger than the proton radius, the dipole amplitude should saturate to the black-disk limit. We therefore penalize deviations from $N(r_{\text{max}}, Y) = 1$.
- **Asymptotic small- r constraint (Color transparency):** In the deep perturbative regime ($r < 10^{-3} \text{ GeV}^{-1}$), color transparency implies $N(r, Y) \sim r^{2\gamma}$, where γ is an effective anomalous dimension. The third term constrains the local logarithmic slope $\gamma \equiv \partial \ln N(r, Y) / \partial \ln r^2$ to the physically allowed range $[0, 1]$, preserving the small- r asymptotics while retaining the functional flexibility required by QCD evolution.
- **Fourier-positivity constraint:** The second term penalizes negative values of the momentum-space S -matrix $S(k_T, Y)$, enforcing positivity of the Fourier-transformed distribution during training.

Here, λ_{IR} , λ_{pos} , and λ_{UV} are penalty coefficients that control the relative strength of the corresponding physical constraints. These coefficients, together with w_i , are numerical balancing factors rather than additional physical fit parameters. In the default production training, the loss weights are kept fixed and set to $(w_1, w_2, w_3) = (10000, 1, 1)$, $(\lambda_{\sigma_r}, \lambda_{\sigma_r^{c\bar{c}}}, \lambda_{\sigma^{\text{VM}}}) = (20, 25, 10)$ and $(\lambda_{\text{IR}}, \lambda_{\text{pos}}, \lambda_{\text{UV}}) = (1, 10000, 1000)$. Their values are determined through gradient-based loss-weight calibration, as described in the next subsection.

The loss is minimized using the AdamW optimizer [76] with a cosine-annealing learning-rate schedule [77], ensuring stable and high-precision convergence. Training was run for 3000 epochs on a GPU, with the learning rate initialized at 10^{-3} .

A3. Gradient-based loss-weight calibration

Our PINN objective combines multiple loss terms with different physical dimensions, numerical scales, and gradient magnitudes, which often leads to ill-conditioned optimization [87, 88]. Inspired by gradient-normalization

ideas from multi-task learning and PINNs [88, 89], we calibrate the loss weights using a gradient-based procedure before production training. The total loss is

$$\mathcal{L}_{\text{total}} = \sum_i w_i \mathcal{L}_i, \quad (\text{S14})$$

where i indexes the ciBK residual, experimental data, Fourier-positivity, and physical-constraint losses. The weights w_i are determined from the unweighted gradient magnitudes. We define the L_2 norm

$$G_i = \|\nabla_{\theta} \mathcal{L}_i\|, \quad (\text{S15})$$

computed during a warm-up stage, take the median $G^{\text{med}} = \text{median}(G_i)$ as a reference, and set

$$w_i = \frac{G^{\text{med}}}{G_i}. \quad (\text{S16})$$

This brings all weighted gradient norms $w_i G_i$ to the same order of magnitude, preventing any single term from dominating the optimization solely because of its numerical scale. The weights are then kept fixed for the remainder of training.

B. Theoretical Formalism for the Observables and Supplementary Results

This section first summarizes the theoretical formalism used to compute the observables, and then presents comprehensive results for the reduced cross section σ_r over all available HERA kinematic range. Additionally, we evaluate the longitudinal structure function F_L as a cross-check, complementing the representative comparisons shown in the main text.

B1. Reduced cross sections and longitudinal structure function: σ_r , F_L and σ_r^{cc}

The DIS cross section is commonly written in terms of the reduced cross section,

$$\sigma_r(y, x_B, Q^2) = F_2(x_B, Q^2) - \frac{y^2}{1 + (1 - y)^2} F_L(x_B, Q^2), \quad (\text{S17})$$

where $y = Q^2/(sx_B)$ is the inelasticity, x_B the Bjorken variable, Q^2 the photon virtuality, $\sqrt{s}$ the lepton-proton center-of-mass energy.

The structure functions F_2 and F_L are related to the transverse (T) and longitudinal (L) virtual-photon proton cross sections via

$$F_2 = \frac{Q^2}{4\pi^2 \alpha_{em}} \left(\sigma_T^{\gamma^* p} + \sigma_L^{\gamma^* p} \right), \quad (\text{S18})$$

$$F_L = \frac{Q^2}{4\pi^2 \alpha_{em}} \sigma_L^{\gamma^* p}. \quad (\text{S19})$$

Here α_{em} is the electromagnetic coupling.

In the leading-order eikonal approximation within the color dipole picture [1], the $\gamma^* p$ interaction proceeds in two steps. First, the virtual photon fluctuates into a quark-antiquark pair, $\gamma^* \rightarrow q\bar{q}$, described by the light-cone wave function $\Psi^{\gamma^* \rightarrow q\bar{q}}$. The resulting color dipole then scatters off the proton. The corresponding $\gamma^* p$ cross sections for transverse and longitudinal photon polarizations are

$$\sigma_{T,L}^{\gamma^* p} = 2 \sum_f \int d^2 \mathbf{b} d^2 \mathbf{r} \frac{dz}{4\pi} |\Psi_{T,L}^{\gamma^* \rightarrow q\bar{q}}(\mathbf{r}, Q^2, z)|^2 \times N(\mathbf{r}, \mathbf{b}, x_B), \quad (\text{S20})$$

Here $\mathbf{r}$ is the transverse size of the $q\bar{q}$ dipole, z is the fraction of the photon longitudinal momentum carried by the quark, $\mathbf{b}$ is the impact parameter, and $N(\mathbf{r}, \mathbf{b}, x_B)$ denotes the dipole-proton scattering amplitude. The sum runs over the light quark flavors and the charm quark.

In this work, we neglect the impact-parameter dependence of the dipole-proton scattering amplitude. Accordingly, we replace the impact-parameter integral by a constant effective proton area, $\int d^2 \mathbf{b} \rightarrow \sigma_0/2$, where $\sigma_0/2$ is the proton transverse area.

The squared photon light-cone wave functions, summed over quark helicities and, for the transverse case, averaged over the two transverse photon polarizations, are [1]

$$|\Psi_L(r, Q^2, z)|^2 = \frac{8N_c}{\pi} \alpha_{em} e_q^2 Q^2 z^2 (1 - z)^2 K_0^2(\varepsilon r), \quad (\text{S21})$$

$$|\Psi_T(r, Q^2, z)|^2 = \frac{2N_c}{\pi} \alpha_{em} e_q^2 \left\{ [z^2 + (1 - z)^2] \varepsilon^2 K_1^2(\varepsilon r) + m_f^2 K_0^2(\varepsilon r) \right\}, \quad (\text{S22})$$

where $r = |\mathbf{r}|$, $\varepsilon^2 = z(1 - z)Q^2 + m_f^2$, e_f is the fractional charge of quark flavor f , and K_0 and K_1 are modified Bessel functions of the second kind of order zero and one, respectively. For the light and charm quarks we use the effective masses $m_{\text{light}} = 0.14$ GeV and $m_c = 1.4$ GeV, respectively.

In practice, when computing $\sigma_{T,L}^{\gamma^* p}$, we apply the standard kinematical shift and introduce the modified Bjorken variable

$$\tilde{x} = x_B \left(1 + \frac{4m_f^2}{Q^2} \right). \quad (\text{S23})$$

for each quark flavor.

B2. Diffractive Vector-Meson Production

In exclusive vector-meson production at small Bjorken- x , the relevant longitudinal momentum fraction is set by

the hard scale $Q^2 + M_V^2$ and the photon-proton center-of-mass energy W ,

$$x \equiv \frac{Q^2 + M_V^2}{W^2 + Q^2}. \quad (\text{S24})$$

The diffractive scattering amplitude for $\gamma^* p \rightarrow V p$ reads [58]

$$\begin{aligned} \mathcal{A}_{T,L}^{\gamma^* p \rightarrow V p}(x, Q^2, \Delta) = i \int d^2 \mathbf{r} \int_0^1 \frac{dz}{4\pi} \int d^2 \mathbf{b} (\Psi_V^* \Psi)_{T,L} \\ \times e^{-i[\mathbf{b} - (1-z)\mathbf{r}] \cdot \Delta} \frac{d\sigma_{q\bar{q}}}{d^2 \mathbf{b}}, \end{aligned} \quad (\text{S25})$$

where Δ is the momentum transfer, related to the Mandelstam variable t by $t = -\Delta^2$, and $d\sigma_{q\bar{q}}/d^2 \mathbf{b}$ is the dipole-proton cross section differential in the impact parameter $\mathbf{b}$. Assuming impact-parameter independence, the b -integrated dipole cross section becomes

$$\sigma_{q\bar{q}}(x, r) \equiv \int d^2 \mathbf{b} \frac{d\sigma_{q\bar{q}}}{d^2 \mathbf{b}} = \sigma_0 N(x, r). \quad (\text{S26})$$

The transverse and longitudinal photon and vector-meson overlap functions are given by [58]

$$\begin{aligned} (\Psi_V^* \Psi)_T = \hat{e}_f e \frac{N_c}{\pi z(1-z)} \left\{ m_f^2 K_0(\varepsilon r) \phi_T(r, z) \right. \\ \left. - [z^2 + (1-z)^2] \varepsilon K_1(\varepsilon r) \partial_r \phi_T(r, z) \right\}, \end{aligned} \quad (\text{S27})$$

$$\begin{aligned} (\Psi_V^* \Psi)_L = \hat{e}_f e \frac{N_c}{\pi} 2Qz(1-z) K_0(\varepsilon r) \left[M_V \phi_L(r, z) \right. \\ \left. + \delta \frac{m_f^2 - \nabla_r^2}{M_V z(1-z)} \phi_L(r, z) \right], \end{aligned} \quad (\text{S28})$$

where $\hat{e}_f$ is the fractional electric charge of the charm quark, $e = \sqrt{4\pi\alpha_{\text{em}}}$, and $M_V = 3.097$ GeV is the J/ψ mass. The scalar functions $\phi_{T,L}(r, z)$ encode the meson wave function, in this work we use the boosted-Gaussian parametrization for J/ψ [58].

The t -differential cross section for exclusive vector-meson production is given by [58]

$$\frac{d\sigma_{T,L}^{\gamma^* p \rightarrow V p}}{dt} = \frac{R_g^2}{16\pi} \left| \mathcal{A}_{T,L}^{\gamma^* p \rightarrow V p} \right|^2 (1 + \beta^2), \quad (\text{S29})$$

where $\beta \equiv \text{Re } \mathcal{A} / \text{Im } \mathcal{A}$ accounts for the real part of the scattering amplitude contribution. It can be calculated from the effective small- x exponent of the amplitude,

$$\beta = \tan\left(\frac{\pi\lambda}{2}\right), \quad \lambda = \frac{\partial \ln(\mathcal{A}_{T,L}^{\gamma^* p \rightarrow V p})}{\partial \ln(1/x)}. \quad (\text{S30})$$

The skewness (off-forward) correction is encoded in R_g , which in the Shuvaev prescription reads [59]

$$R_g = \frac{2^{2\lambda+3}}{\sqrt{\pi}} \frac{\Gamma(\lambda + \frac{5}{2})}{\Gamma(\lambda + 4)}. \quad (\text{S31})$$

Integrating Eq. (S29) over t yields the total cross section,

$$\sigma_{T,L}^{\gamma^* p \rightarrow V p} = \int dt \frac{d\sigma_{T,L}^{\gamma^* p \rightarrow V p}}{dt}, \quad (\text{S32})$$

$$\sigma_{\text{total}}^{\gamma^* p \rightarrow V p} = \sigma_T^{\gamma^* p \rightarrow V p} + \sigma_L^{\gamma^* p \rightarrow V p}. \quad (\text{S33})$$

In the present study we focus on J/ψ photoproduction with $Q^2 = 0$, for which only the transverse contribution is relevant.

B3. Reduced cross sections and longitudinal structure function at all energies

This section presents supplementary comparisons for the inclusive reduced cross section σ_r at all available HERA center-of-mass energies, and for the longitudinal structure function F_L as a cross-check. Figure S2 compares the PINN predictions with the combined HERA measurements of the inclusive reduced total cross section σ_r at the center-of-mass energies $\sqrt{s} = 224.9, 251.5, 300.3$, and 318 GeV. The good agreement over the measured kinematic range provides a global validation of the extracted dipole amplitude in inclusive DIS.

Figure S3 compares the PINN predictions with the F_L measurements. The significance of this cross-check lies in isolating the longitudinal photon contribution through Eq. (S19). Since σ_r depends on both F_2 and F_L (see Eq. (S17)), these constraints are not fully independent, and the F_L comparison provides a nontrivial consistency check of the extracted universal amplitude. The resulting uncertainty band for F_L is comparatively narrow in our fit, consistent with the indirect constraint provided by σ_r . The agreement with the data supports the robustness of the extracted $N(r, Y)$ across observables that emphasize different photon polarizations.

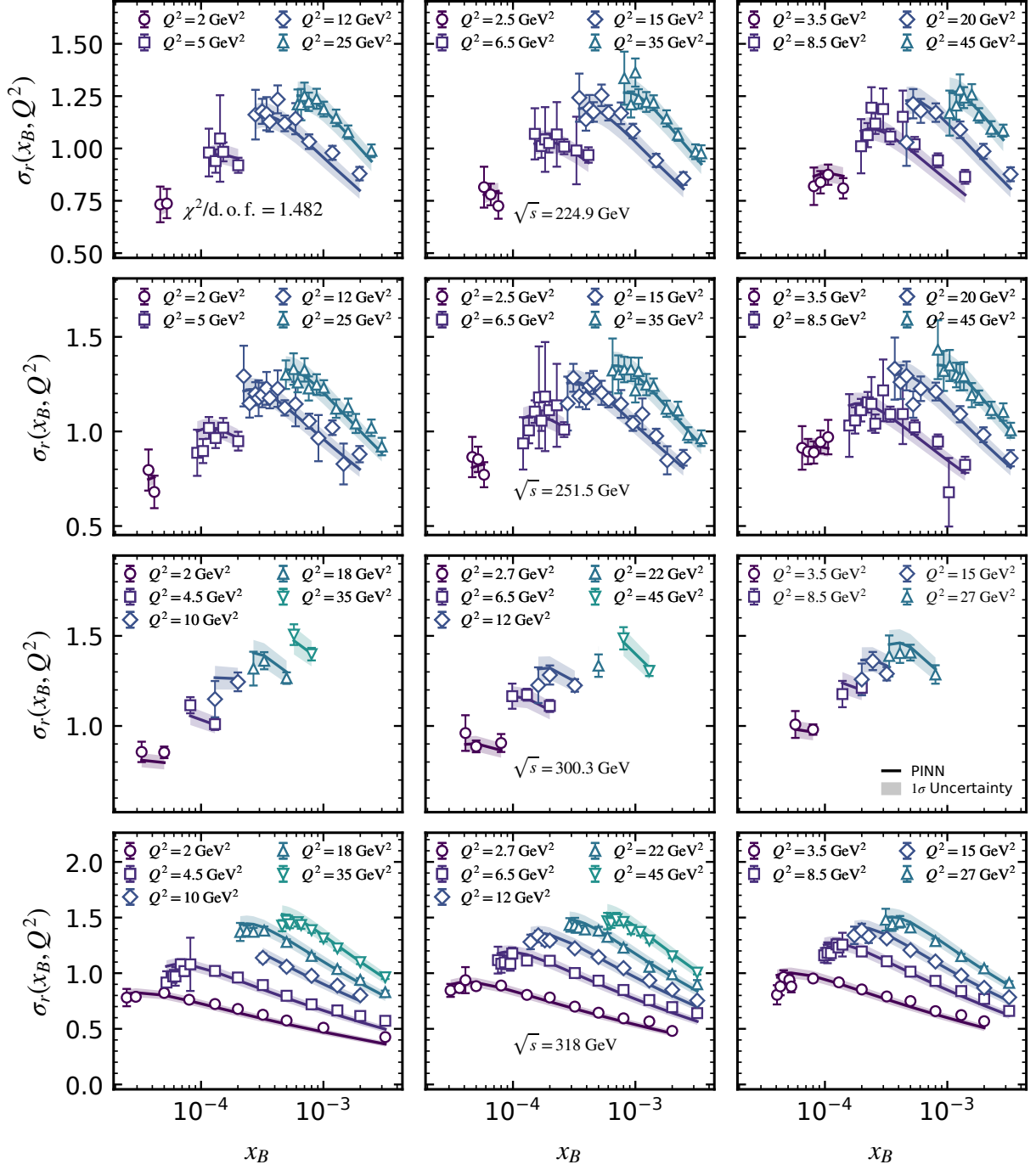


FIG. S2. The reduced total cross section $\sigma_r(x_B, Q^2)$ compared with combined HERA e^+p data [62, 63] at all available center-of-mass energies, $\sqrt{s} = 224.9, 251.5, 300.3, 318$ GeV. The quoted $\chi^2/\text{d.o.f.}$ uses the solid curve, and it includes all available energies.

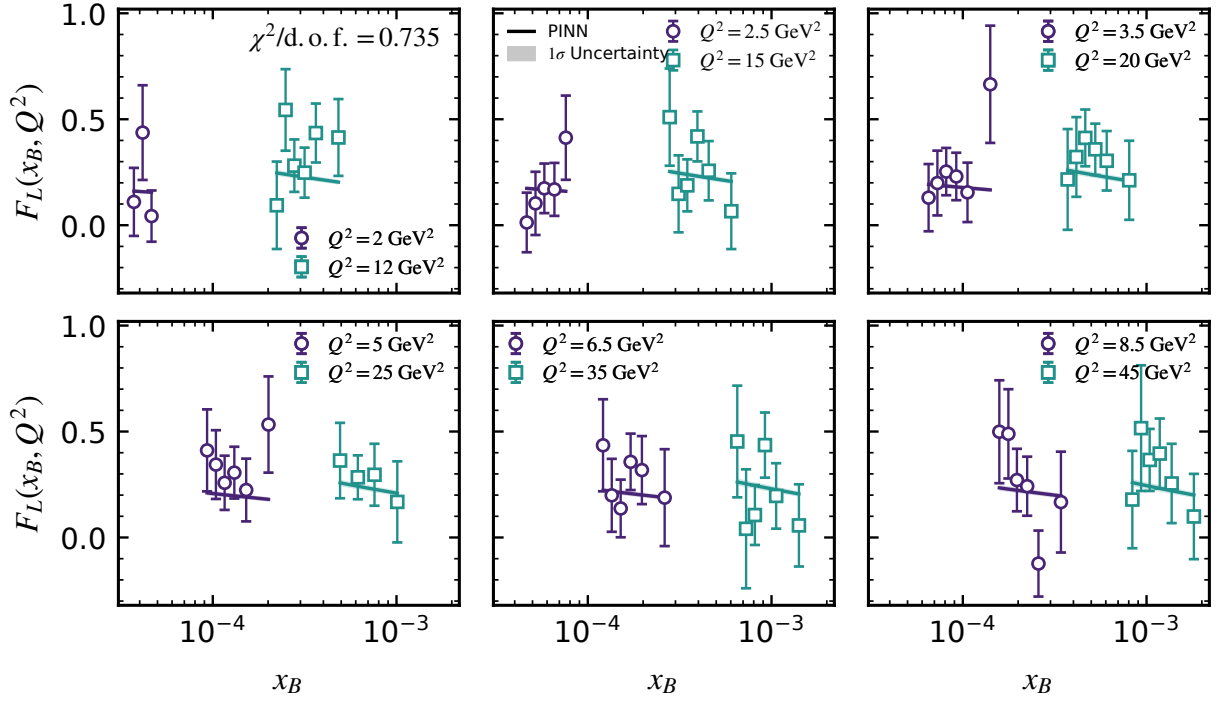


FIG. S3. The longitudinal structure function $F_L(x_B, Q^2)$ compared with HERA e^+p data [90] at different Q^2 at all available center-of-mass energies.