

Determination of the Initial Condition for the Balitsky–Kovchegov Equation with Transformers

Meisen Gao (高梅森)

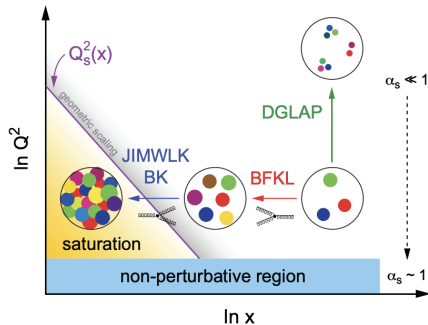
East China University of Science and Technology

with Zhong-Bo Kang (UCLA), Jani Penttala (UCLA), Ding Yu Shao (Fudan University)

[arXiv:2510.26779](https://arxiv.org/abs/2510.26779)

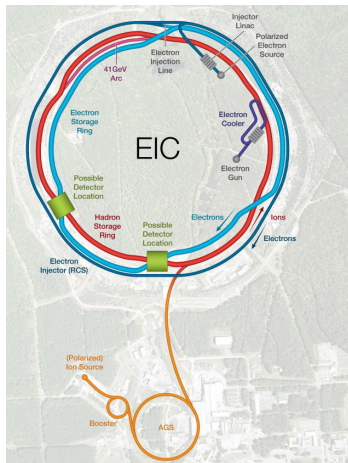
2025.11.02 • Beijing, China

Small- x QCD and Gluon Saturation



- For $x \ll 1$, gluon multiplication drives high occupancies; nonlinear recombination generates a **saturation scale** $Q_s(x)$ and *geometric scaling*.
- The **Color Glass Condensate (CGC)** describes this regime via Wilson lines; in DIS the key quantity is the (impact-parameter integrated) **dipole amplitude** $N(r, x)$.
- Its energy evolution is governed by the **BK** equation: linear BFKL growth minus a nonlinear damping term $\Rightarrow$ saturation (Balitsky '96; Kovchegov '99–'00).

From HERA to EIC: Gluon Saturation



EIC UG (2023), *EIC White Paper*.

- Mapping this regime is a central objective for the **Electron–Ion Collider** program (EIC White Paper 2012; EIC Yellow Report 2021).
- Kinematic reach down to $x \sim 10^{-4} - 10^{-5}$ at moderate Q^2 ; heavy nuclei amplify $Q_s(x)$, giving cleaner saturation signals.
- Key channels: inclusive/diffractive DIS, F_L , charm, di-hadrons, exclusive vector mesons— directly sensitive to $N(r, x)$ and $Q_s(x)$.
- *This talk*: determine the nonperturbative initial condition $N(r, x_0)$ from **HERA** DIS.

Initial Condition and Fitting Challenge

- BK evolution requires a nonperturbative **initial condition** $N(r, x_0)$ at moderately small x_0 , fixed by data.
- We use a generalized MV ansatz with (Q_{s0}, γ, e_c) :

$$N(r, x_0) = 1 - \exp \left[- \left(\frac{r^2 Q_{s0}^2}{4} \right)^\gamma \ln \left(\frac{1}{\Lambda_{\text{QCD}} r} + e e_c \right) \right].$$

- Prior small- x dipole fits (e.g., Albacete *et al.* 2011; Lappi & Mäntysaari 2013).
- NLO developments: Beuf *et al.* (2020); Hänninen *et al.* (2023).
- Data: H1 & ZEUS combined inclusive e^+p (EPJC **75** (2015) 580).
- **This work:** LO dipole DIS with **rcBK** (Balitsky prescription), b -integrated; fast scans via **transformer** emulators (cf. Casuga *et al.* 2024).

Inclusive DIS in the Dipole Picture

- Neutral-current reduced cross section (at small x):

$$\sigma_r(x, y, Q^2) = F_2(x, Q^2) - \frac{y^2}{Y_+} F_L(x, Q^2), \quad Y_+ \equiv 1 + (1 - y)^2.$$

- Dipole picture: $\gamma^* \rightarrow q\bar{q}$ followed by dipole-proton scattering,

$$\sigma_{T,L}(x, Q^2) = \sum_f \int_0^1 dz \int d^2\mathbf{r} |\Psi_{T,L}(z, \mathbf{r}; Q^2)|^2 \sigma_{\text{dip}}(x, r),$$

with impact-parameter integration

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

- Photon wavefunctions (massless limit):

$$|\Psi_L|^2 \propto Q^2 z^2 (1 - z)^2 K_0^2(\bar{Q}r), \quad |\Psi_T|^2 \propto [z^2 + (1 - z)^2] K_1^2(\bar{Q}r),$$

with $\bar{Q}^2 = z(1 - z)Q^2$ and $K_{0,1}$ modified Bessel K functions.

Balitsky–Kovchegov (BK) Evolution

- Goal: describe the $x \downarrow$ (energy $\uparrow$) evolution of the **dipole amplitude** $N(r, x)$ in the CGC/dipole picture.
- In rapidity $Y = \ln(1/x)$ the (impact-parameter integrated) LO BK equation reads

$$\frac{\partial N(\mathbf{x}_{01}, Y)}{\partial Y} = \frac{\alpha_s N_c}{2\pi^2} \int d^2 \mathbf{x}_2 K_{\text{LO}}(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) \left[N(\mathbf{x}_{02}, Y) + N(\mathbf{x}_{12}, Y) - N(\mathbf{x}_{01}, Y) - N(\mathbf{x}_{02}, Y)N(\mathbf{x}_{12}, Y) \right],$$
$$K_{\text{LO}}(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) = \frac{x_{01}^2}{x_{02}^2 x_{12}^2}, \quad x_{ij} \equiv |\mathbf{x}_i - \mathbf{x}_j|, \quad r \equiv x_{01}.$$

- The first three (linear) terms reproduce BFKL gluon radiation; the last (nonlinear) term subtracts double counting of simultaneous scatterings $\Rightarrow$ gluon recombination $\Rightarrow$ **saturation**.
- **Limits:** color transparency $r \rightarrow 0$: $N \propto r^2$ (weak scattering); black-disk $r \rightarrow \infty$: $N \rightarrow 1$ (unitarity bound).
- **This work:** we solve and emulate the **running–coupling** BK (rcBK) with the Balitsky prescription; the $\alpha_s(r^2)$ choice follows the standard rcBK practice.

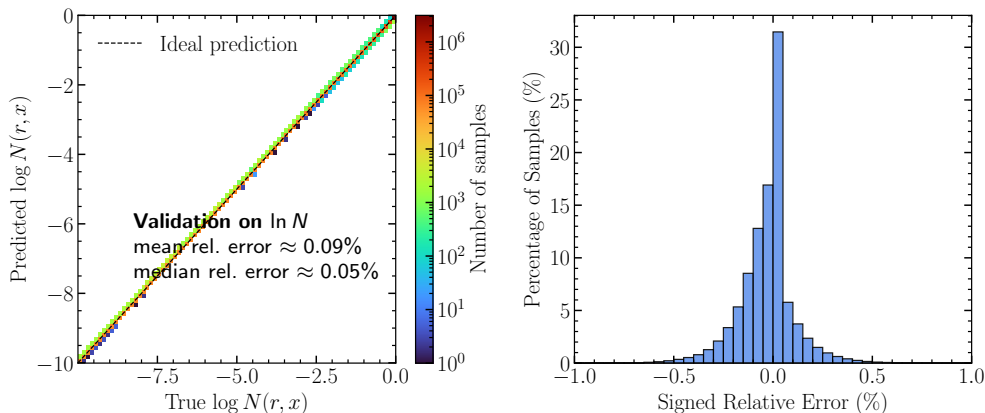
Generating BK Solutions for Training

- To enable a fit, we first generate a large set of BK solutions spanning the parameter space:
- We consider two possible starting points for evolution: $x_0 = 0.01$ (standard) and a higher $x_0 = 0.05$ (to test sensitivity to the starting scale).
- For each x_0 , we sample $\sim 10^4$ random initial parameter sets $(Q_{s0}^2, \gamma, e_c, C^2)$ using Latin Hypercube sampling to cover the space efficiently.
- For each sampled set, we solve the running-coupling BK equation (**rcBK, Balitsky prescription**) numerically (using a custom Julia solver):
 - $N(r, x_0)$ given by the initial model (MV-type) with those parameters.
 - Evolve $N(r, x)$ to lower x up to $Y \equiv \ln(x_0/x) \approx 16$ (i.e. down to $x \simeq x_0 e^{-16}$; for $x_0 = 0.01$ this is $\sim 10^{-9}$).
 - Compute $N(r, x)$ on a grid of r from 10^{-6} to 10^2 GeV^{-1} (covers all relevant dipole sizes).
- This yields a comprehensive dataset of $\mathcal{O}(10^7)$ points $\{r, x, N(r, x)\}$ (about 4.7×10^7 values of N in total) covering wide ranges of parameters and kinematics.

Transformer Network for $N(r, x)$ Emulation

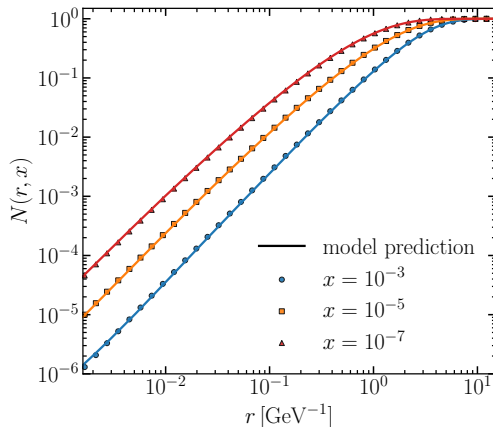
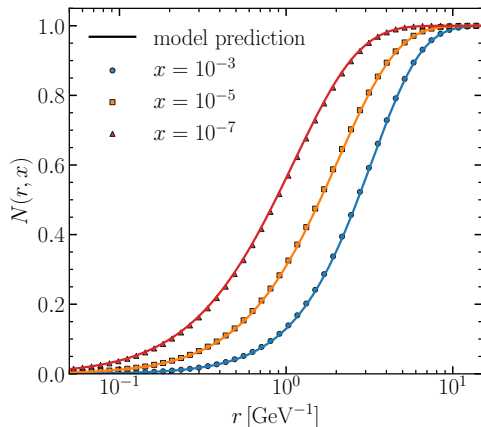
- **Inputs (6 scalars):** $(\log_{10} Q_{s0}^2, \gamma, e_c, \log_{10} C, \log_{10} r, x)$; standardized per-dimension.
- **Tokenization & PE:** each scalar $\rightarrow$ shared linear projection (token); add sinusoidal positional encodings.
- **Encoder:** Transformer encoder (4 layers, 8 heads, $d_{\text{model}}=128$, FF=512, dropout 0.1; pre-norm + residual).
- **Head & constraint:** mean-pool tokens $\rightarrow$ 2-layer MLP $\rightarrow$ logit z ; apply *sigmoid* so $0 < N_{\text{pred}} = \sigma(z) < 1$ (unitarity).
- **Loss (log-MSE):**
$$\mathcal{L} = \text{MSE}(\log N_{\text{pred}}, \log N_{\text{true}}),$$
robust in both dilute ($N \ll 1$) and saturation ($N \rightarrow 1$) regimes.
- **Training setup:** 90/10 train/val; AdamW (lr $3 \times 10^{-4} \rightarrow 3 \times 10^{-6}$, cosine anneal), weight decay 10^{-4} , early stopping; AMP mixed precision; gradient clipping (max-norm 1.0).
- **Rationale:** self-attention captures interactions among (Q_{s0}, γ, e_c, C) and kinematics (r, x) without hand-crafted features; highly parallelizable.

BK Dipole Amplitude Emulator Performance



- **Left:** parity plot ($\ln N_{\text{pred}}$ vs. $\ln N_{\text{true}}$) tightly follows the diagonal.
- **Right:** signed relative-error histogram for N (zoomed to $\pm 1\%$); most entries are per-mille.

Example: Emulator vs. Exact BK Solution

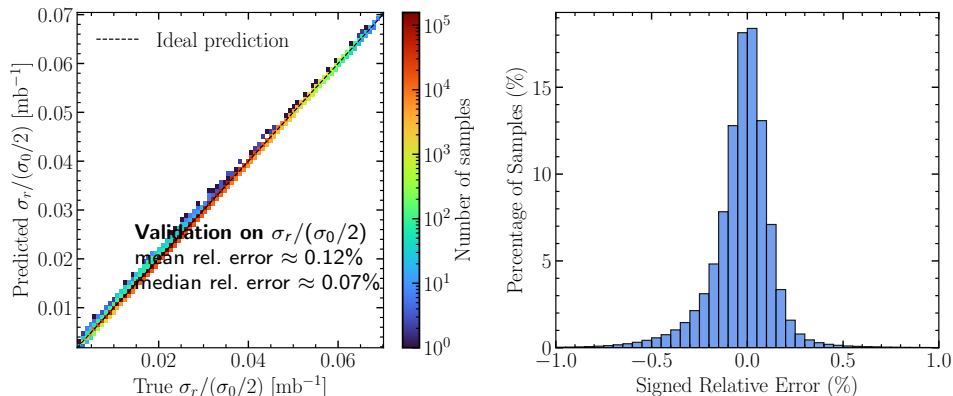


- **Points:** exact BK; **Lines:** transformer emulator.
- Agreement is essentially pointwise across many decades in r and all three x . The emulator reproduces both the *dilute* limit ($N \ll 1$) and the *saturation* limit ($N \rightarrow 1$).

Emulating the DIS Cross Section

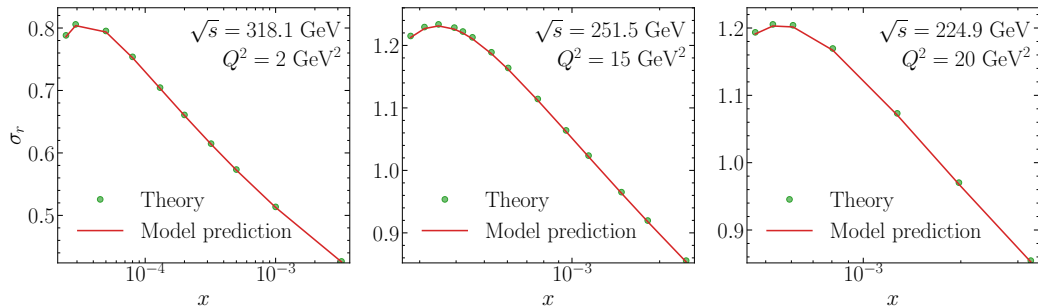
- Goal: compare to data via the reduced cross section $\sigma_r(x, Q^2)$.
- Direct evaluation requires integrating $N(r, x)$ with photon wavefunctions; scanning many parameter sets makes this costly.
- **Solution:** a second Transformer directly emulates $\sigma_r/(\sigma_0/2)$ from the 6 inputs $(\log_{10} Q_{s0}^2, \gamma, e_c, \log_{10} C, \log_{10} x, \log_{10} Q^2)$.
- **Training data:** from each BK solution, compute σ_r (both T and L) on kinematics covering HERA; total $\sim 4 \times 10^6$ examples.
- **Model:** same encoder as for $N(r, x)$; regression target is $\sigma_r/(\sigma_0/2)$.
- **Loss:** mainly MSE with a small Smooth- L_1 term for robustness to occasional large residuals.
- **Accuracy:** $\approx 0.1\%$ mean relative error on held-out data, sufficient for precision fits.

DIS Cross Section Emulator Performance



- **Left:** parity plot closely follows the diagonal (near-perfect prediction).
- **Right:** signed relative-error histogram; most entries well below 10⁻³.
- Precision comfortably exceeds experimental needs (data uncertainties are at the few-percent level).

Validating Cross Section Predictions



- Representative parameter set *not used* during training.
- **Green markers:** direct dipole-model integration; **Red curves:** emulator prediction.
- Near pointwise agreement across the full x range and all Q^2 bins $\Rightarrow$ emulator reproduces the x - Q^2 dependence of σ_r .
- Supports using the surrogate for fast scans and a comprehensive HERA fit.

Fitting to HERA Data: Setup

- **Dataset:** HERA combined inclusive e^+p DIS (H1+ZEUS, 2015), $x \lesssim 10^{-2}$ with $1 \lesssim Q^2 \lesssim 50 \text{ GeV}^2$.

- **Objective:** minimize

$$\chi^2 = \sum_i \frac{[\sigma_{r,i}^{\text{data}} - \sigma_{r,i}^{\text{th}}(\boldsymbol{\theta})]^2}{(\delta\sigma_{r,i}^{\text{data}})^2},$$

where $\boldsymbol{\theta} = (\sigma_0/2, Q_{s0}^2, e_c, \gamma, C^2)$.

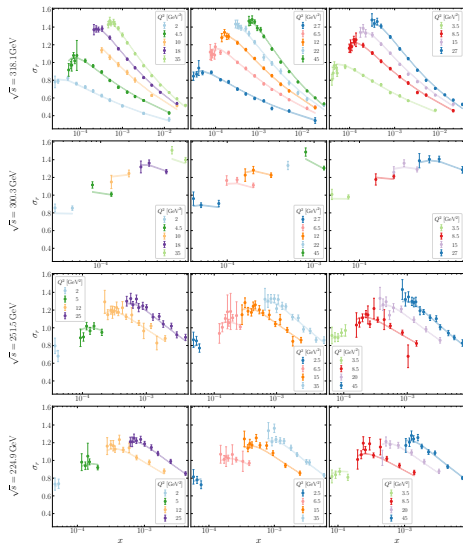
- **Fast evaluation:** emulators make σ_r^{th} essentially instantaneous $\Rightarrow$ robust scans/uncertainty estimates.
- **Minimizer:** Minuit (`iminuit`).
- **Uncertainties:** replica method (200 pseudo-datasets with Gaussian fluctuations); fit each replica, use the spread for parameter/prediction bands.
- **Fit variants (4):** $x_0 = 0.01$ with γ fixed/free; $x_0 = 0.05$ with γ fixed/free.
- **Goal:** compare x_0 choices and assess sensitivity/validity of rcBK when starting at $x_0 = 0.05$.

Fit Results: Extracted Parameters

Parameter	Prior Range	$x_0 = 0.01$		$x_0 = 0.05$	
		4-param ($\gamma = 1$)	5-param (γ free)	4-param ($\gamma = 1$)	5-param (γ free)
Q_{s0}^2 [GeV ²]	[0.01, 0.11]	$0.063^{+0.001}_{-0.004}$	$0.068^{+0.024}_{-0.015}$	$0.025^{+0.0035}_{-0.0025}$	$0.045^{+0.0021}_{-0.0017}$
e_c	[0.5, 70.0]	$29.0^{+8.7}_{-3.1}$	$18.7^{+20.1}_{-9.6}$	$34.0^{+35.3}_{-3.2}$	$41.0^{+9.2}_{-2.6}$
C^2	[2.0, 20.0]	$4.45^{+0.99}_{-0.59}$	$4.82^{+1.56}_{-2.63}$	$17.4^{+2.5}_{-6.1}$	$10.2^{+1.6}_{-0.8}$
$\sigma_0/2$ [mb]	[12.0, 20.0]	$14.5^{+0.6}_{-0.4}$	$14.8^{+0.9}_{-2.4}$	$19.4^{+0.5}_{-2.0}$	$16.6^{+0.2}_{-0.4}$
γ	[0.9, 1.3]	1.00 (fixed)	$1.006^{+0.037}_{-0.019}$	1.00 (fixed)	$1.138^{+0.033}_{-0.012}$
χ^2/dof	—	0.854	0.857	1.471	1.195

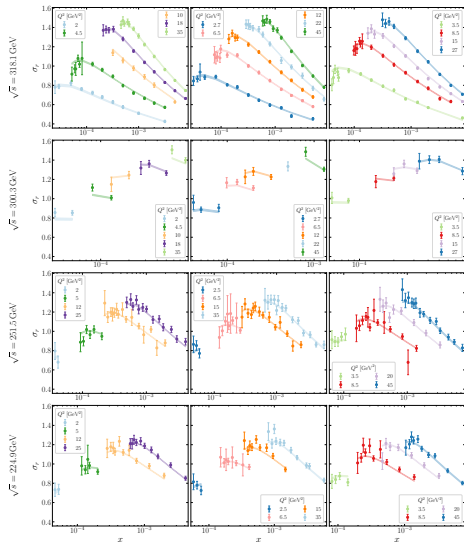
- Uncertainties are 95% credible intervals from 200 replicas. “4-param” fixes $\gamma = 1$; “5-param” lets γ float.
- For $x_0 = 0.01$: $Q_{s0}^2 \approx 0.06$ GeV², $C^2 \approx 4$ –5, $\sigma_0/2 \approx 14$ –15 mb; $\gamma \simeq 1$.
- For $x_0 = 0.05$: smaller Q_{s0}^2 (later evolution start), preference for larger C^2 ; in 5-param, $\gamma \approx 1.14$.
- Fit quality: $\chi^2/\text{dof} \approx 0.86$ at $x_0 = 0.01$; ~ 1.20 at $x_0 = 0.05$ (5-param).

Comparison to HERA Data ($x_0 = 0.05$ fit)



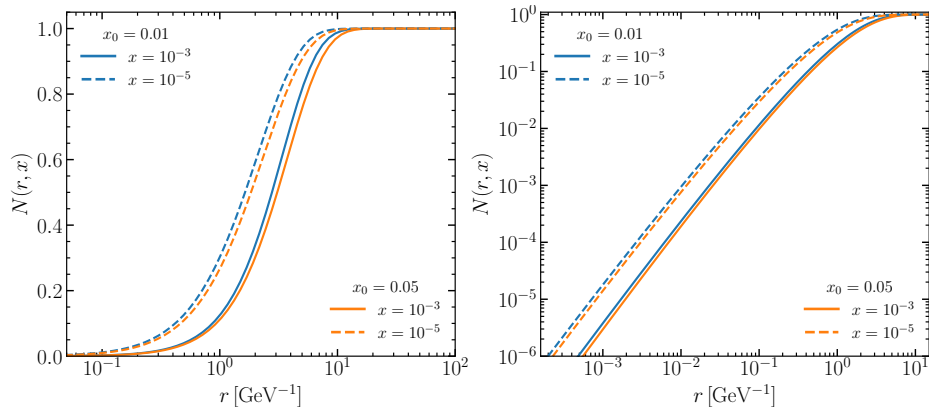
- Reduced cross section σ_r vs. x at fixed Q^2 , compared to HERA combined e^+p data (points with errors).
- Curves/bands show the $x_0 = 0.05$ fit (5-parameter, γ free); bands indicate 2σ from replicas.

Comparison to HERA Data ($x_0 = 0.01$ fit)



- Reduced cross section σ_r vs. x at fixed Q^2 , compared to HERA combined e^+p data (points with errors).
- Excellent agreement with HERA data; $\chi^2/\text{dof} \approx 0.86$.

Dipole Amplitude from Different x_0 Fits



- Evolving both fits to a common x yields nearly identical $N(r, x)$; the small residuals around $r \sim 1$ GeV⁻¹ are minor and largely absorbed by the fitted normalization $\sigma_0/2$ in σ_r .

Summary

- Transformer emulators deliver fast, high-fidelity predictions for BK $N(r, x)$ and DIS σ_r .
- Enable an efficient fit to HERA small- x data; best performance at $x_0 = 0.01$.
- Extracted parameters are consistent with prior studies; uncertainties quantified via replicas.

Outlook

- Include NLO ingredients in the BK kernel and DIS wavefunctions.
- Introduce impact-parameter dependence $N(\mathbf{r}, \mathbf{b}; x)$.
- Toward global analyses: inclusive & diffractive DIS, heavy flavor, and pA.

Thank you!