

BAYESIAN INFERENCE OF THE FLUCTUATING PROTON SHAPE IN DIS AND HADRONIC COLLISIONS*

HEIKKI MÄNTYSAARI^{a,b}, BJÖRN SCHENKE^c
CHUN SHEN^{d,e}, WENBIN ZHAO^e

^aDepartment of Physics, University of Jyväskylä
P.O. Box 35, 40014 University of Jyväskylä, Finland

^bHelsinki Institute of Physics, P.O. Box 64, 00014 University of Helsinki, Finland

^cPhysics Department, Brookhaven National Laboratory, Upton, NY 11973, USA

^dDepartment of Physics and Astronomy, Wayne State University
Detroit, Michigan 48201, USA

^eRIKEN BNL Research Center, Brookhaven National Laboratory
Upton, NY 11973, USA

*Received 31 July 2022, accepted 14 October 2022,
published online 14 December 2022*

We determine the likelihood distribution for the model parameters describing the event-by-event fluctuating proton geometry at small x by performing a Bayesian analysis within the Color Glass Condensate framework. The exclusive J/ψ production data from HERA is found to constrain the model parameters well, and we demonstrate that complementary constraints can be obtained from simulations of Pb+Pb collisions at the LHC.

DOI:10.5506/APhysPolBSupp.16.1-A33

1. Introduction

Determining the partonic structure of protons and nuclei is one of the main goals of future nuclear DIS facilities such as the Electron–Ion Collider [1]. Determining the spatial distribution of gluonic matter at small x is fundamentally interesting, and is also extremely important to provide input to simulations of heavy-ion collisions at RHIC and at the LHC where the final-state hydrodynamical evolution transforms the initial coordinate space anisotropies into the momentum space correlations.

In Deep Inelastic Scattering, the simple structure of the photon probe allows for a precise determination of the gluonic structure of protons and nuclei. Exclusive processes such as the J/ψ production are especially important, as only in exclusive scattering it is possible to determine the total

* Presented at the 29th International Conference on Ultrarelativistic Nucleus–Nucleus Collisions: Quark Matter 2022, Kraków, Poland, 4–10 April, 2022.

momentum transfer to the target hadron, which by definition is the Fourier conjugate to the impact parameter, and as such, provides access to the target geometry. In addition to the average geometry, it is important to understand the shape fluctuations that can be expected to play a major role when looking at fluctuation-dominated flow observables, *e.g.* in proton–lead collisions at the LHC.

In our recent Letter [2], we performed a Bayesian analysis to extract the likelihood distribution for the model parameters describing the event-by-event fluctuating proton shape from HERA J/ψ production data [3], assuming that the nucleon substructure can be described in terms of gluonic hot spots as suggested in Ref. [4], see also Ref. [5] for a review.

2. Constraining model parameters using HERA data

The scattering amplitude for exclusive vector meson production in the dipole picture can be written as [6]

$$\begin{aligned} \mathcal{A}^{\gamma^*+p \rightarrow V+p} &= 2i \int d^2\mathbf{r}_\perp d^2\mathbf{b}_\perp \frac{dz}{4\pi} e^{-i[\mathbf{b}_\perp - (\frac{1}{2}-z)\mathbf{r}_\perp] \cdot \mathbf{\Delta}_\perp} \\ &\times [\Psi_V^* \Psi_\gamma] (Q^2, \mathbf{r}_\perp, z) N_\Omega(\mathbf{r}_\perp, \mathbf{b}_\perp, x_\mathbb{P}) . \end{aligned} \quad (1)$$

Here, Ψ_γ is the photon light front wave function describing the $\gamma^* \rightarrow q\bar{q}$ splitting, Ψ_V is the vector meson wave function for which we use the Boosted Gaussian parametrization [6], and N_Ω is the dipole–proton scattering amplitude, where Ω refers to a particular proton configuration. The transverse size of the dipole is $\mathbf{r}_\perp$ and the proton-to-dipole distance is $\mathbf{b}_\perp$. The fraction of the photon plus momentum carried by the quark is denoted by z , and $x_\mathbb{P}$ is the fraction of the target longitudinal momentum (in the infinite momentum frame) transferred in the process.

The coherent cross section corresponding to the events where the proton remains intact reads

$$\frac{d\sigma^{\gamma^*+p \rightarrow V+p}}{d|t|} = \frac{1}{16\pi} \left| \left\langle \mathcal{A}^{\gamma^*+p \rightarrow V+p} \right\rangle_\Omega \right|^2 , \quad (2)$$

and is sensitive to the average dipole–proton interaction, and as such, to the average geometry. Here, $\langle \rangle_\Omega$ corresponds to an average over the target configurations. On the other hand, calculating the total diffractive cross section and subtracting the coherent contribution, one obtains the incoherent cross section corresponding to events in which the proton breaks up

$$\frac{d\sigma^{\gamma^*+p \rightarrow V+p^*}}{d|t|} = \frac{1}{16\pi} \left[\left\langle \left| \mathcal{A}^{\gamma^*+p \rightarrow V+p} \right|^2 \right\rangle_\Omega - \left| \left\langle \mathcal{A}^{\gamma^*+p \rightarrow V+p} \right\rangle_\Omega \right|^2 \right] . \quad (3)$$

As a variance, the incoherent cross section is sensitive to the amount of fluctuations at distance scale $\sim 1/\sqrt{|t|}$ in the scattering amplitude.

The dipole–proton scattering amplitude N is obtained using the same approach as in the IP-Glasma calculation of the initial conditions for heavy-ion collisions [7], following Ref. [4]. The local color charge density is assumed to be proportional to the local saturation scale $Q_s^2(\mathbf{b}_\perp)$ extracted from the IPsat parametrization, and as such, to the local density $T_p(\mathbf{b}_\perp)$. We introduce an event-by-event fluctuating density by writing the density profile following Ref. [4] as:

$$T_p(\mathbf{b}_\perp) = \frac{1}{N_q} \sum_{i=1}^{N_q} p_i T_q(\mathbf{b}_\perp - \mathbf{b}_{\perp,i}), \quad T_q(\mathbf{b}_\perp) = \frac{1}{2\pi B_q} e^{-\mathbf{b}_\perp^2/(2B_q)}, \quad (4)$$

and the coefficient p_i allows for different normalizations for individual hot spots. This coefficient is sampled from a log-normal distribution whose width σ is taken to be a model parameter, as well as the hot-spot size B_q and the proton size B_p , which is the width of a Gaussian probability distribution for the hot-spot positions $\mathbf{b}_{\perp,i}$. We additionally include repulsive short-range correlations between the hot spots by introducing a parameter $d_{q,\min}$ which is the smallest allowed distance between the hot spots. The remaining model parameters are an infrared regulator m , the ratio between the local color charge density and the saturation scale ($Q_s/(g^2\mu)$) that controls the overall normalization, and the number of hot spots N_q .

To determine the likelihood distribution for the model parameters, we employ the Bayesian Inference. It is a general and systematic method to constrain the probability distribution of model parameters $\boldsymbol{\theta}$ by comparing model calculations $\mathbf{y}(\boldsymbol{\theta})$ with experimental measurements $\mathbf{y}_{\text{exp}}$ [8] (J/ψ production data at $W = 75$ GeV measured by H1 [3]). According to Bayes' theorem, the posterior distribution of model parameters satisfies

$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{y}_{\text{exp}}) \propto \mathcal{P}(\mathbf{y}_{\text{exp}}|\boldsymbol{\theta})\mathcal{P}(\boldsymbol{\theta}). \quad (5)$$

Here, $\mathcal{P}(\mathbf{y}_{\text{exp}}|\boldsymbol{\theta})$ is the likelihood for model results with parameter $\boldsymbol{\theta}$ to agree with the experimental data that we calculate using the Gaussian process emulators. The final posterior distribution is determined by using the Markov Chain Monte Carlo sampling. For more details, see Ref. [2] and references therein.

3. Results

The determined posterior distribution of model parameters is shown in Fig. 1. The HERA data used to constrain the parameters corresponds to $x_{\mathbb{P}} \approx 10^{-3}$. We show separately results from two analyses, one with a fixed number of hot spots $N_q = 3$ with results shown in red/gray and in the upper right corner of the figure, the second with N_q a free parameter, and results shown in blue/black and in the lower left corner of the figure.

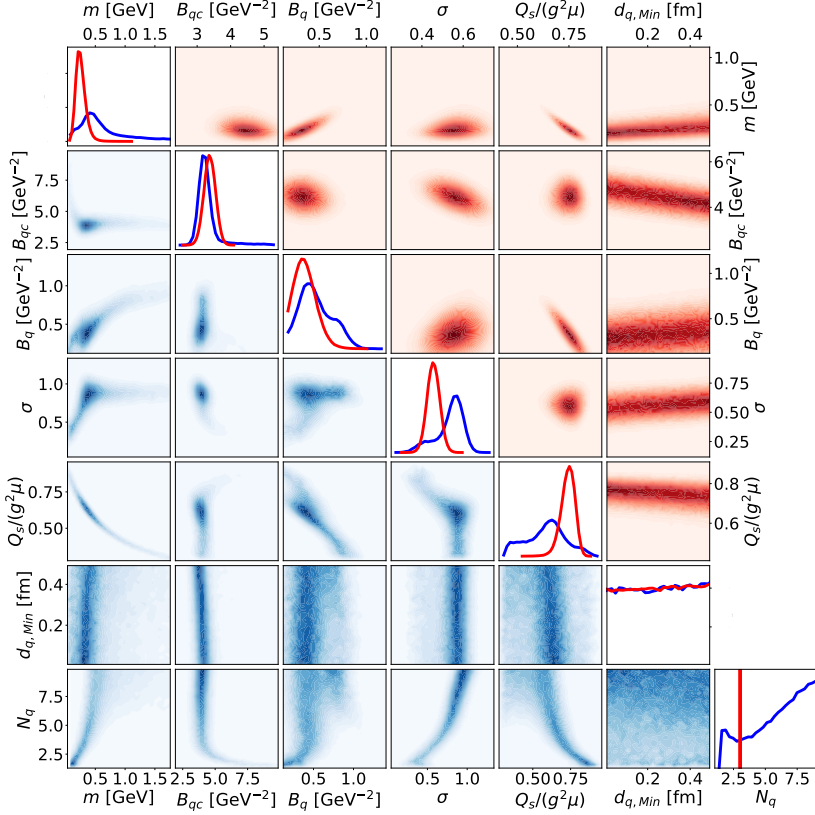


Fig. 1. (Color online) Posterior distribution of the model parameters.

Most model parameters can be constrained well, except the parameter $d_{q,min}$ describing the repulsive short-range correlations between the hot spots that were found in Ref. [9] to be necessary to describe high-multiplicity proton–proton collisions. This means that the J/ψ production data from HERA allows but does not require such repulsive correlations. Similarly, the number of hot spots is not constrained by the data. This can be understood by noticing that there is a strong positive correlation between the number of hot spots N_q and the hot-spot density fluctuations σ . With large N_q , there are also very large density fluctuations which means that only a few hot spots actually dominate. Additionally, with large N_q , the hot spots start to overlap which further reduces the “effective number of hot spots”.

With both variables N_q and $N_q = 3$, we get an equally good description of the HERA data, which implies that the HERA data alone does not completely constrain the fluctuating geometry. Additional constraints can be obtained from heavy-ion collisions. As a proof-of-concept, we take maximum likelihood parametrizations with $N_q = 3$ and $N_q = 9$, and use those to

construct an initial condition for Pb+Pb collisions at $\sqrt{s} = 5.02$ TeV. The initial condition and early evolution before the QGP phase is described using the IP-Glasma framework [7]. It is then coupled to MUSIC [10] hydrodynamical simulations of the plasma evolution, and to the UrQMD afterburner describing the more dilute hadronic phase [11] (see [12] for a description of the entire framework).

The multiplicity distribution in Pb+Pb collisions is shown in Fig. 2. Note that the multiplicities in the most central bin match by construction. We find that the ALICE data [13] prefers the $N_q = 3$ parametrization. Similarly the centrality dependence of the flow harmonics $v_2\{2\}$ and $v_2\{4\}$ shown in Fig. 3 and compared with the ALICE data [14] prefers this parametrization. These results clearly indicate that the LHC data can provide further constraints on the fluctuating shape of the nucleons.

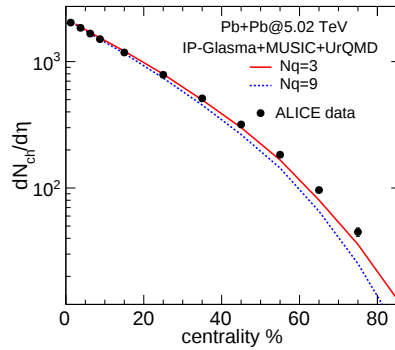


Fig. 2. Multiplicity distribution in Pb+Pb compared to the ALICE data.

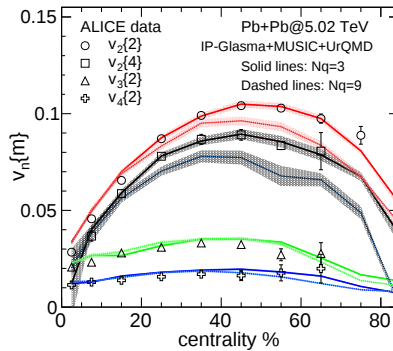


Fig. 3. Flow harmonics in Pb+Pb collisions compared to the ALICE data.

4. Conclusions

We have performed a Bayesian analysis to extract the posterior likelihood distribution for the non-perturbative parameters describing the event-

by-event fluctuating proton geometry using the HERA J/ψ production data. Most of the model parameters are well constrained by the data, except that the potential repulsive short-range correlations cannot be determined from this data. The obtained likelihood distribution can be used to systematically take into account uncertainties in the proton geometry when calculating any other observable that depends on the event-by-event fluctuating geometry. We have further demonstrated that complementary constraints can be obtained from simulations of heavy-ion collisions where the initial nucleon geometry affects the space-time evolution of the produced QGP.

B.P.S. and C.S. are supported under Department of Energy Contract No. DE-SC0012704 and Award No. DE-SC0021969, respectively. C.S. acknowledges a DOE Office of Science Early Career Award. H.M. is supported by the Academy of Finland, the Centre of Excellence in Quark Matter, and projects 338263 and 346567. W.B.Z. is supported by the National Science Foundation (NSF) under grant number ACI-2004571.

REFERENCES

- [1] R. Abdul Khalek *et al.*, [arXiv:2103.05419](#) [[physics.ins-det](#)].
- [2] H. Mäntysaari, B. Schenke, C. Shen, W. Zhao, *Phys. Lett. B* **833**, 137348 (2022), [arXiv:2202.01998](#) [[hep-ph](#)].
- [3] H1 Collaboration (C. Alexa *et al.*), *Eur. Phys. J. C* **73**, 2466 (2013), [arXiv:1304.5162](#) [[hep-ex](#)].
- [4] H. Mäntysaari, B. Schenke, *Phys. Rev. Lett.* **117**, 052301 (2016), [arXiv:1603.04349](#) [[hep-ph](#)].
- [5] H. Mäntysaari, *Rep. Prog. Phys.* **83**, 082201 (2020), [arXiv:2001.10705](#) [[hep-ph](#)].
- [6] H. Kowalski, L. Motyka, G. Watt, *Phys. Rev. D* **74**, 074016 (2006), [arXiv:hep-ph/0606272](#).
- [7] B. Schenke, P. Tribedy, R. Venugopalan, *Phys. Rev. Lett.* **108**, 252301 (2012), [arXiv:1202.6646](#) [[nucl-th](#)].
- [8] D. Sivia, J. Skilling, «Data Analysis: A Bayesian Tutorial», Oxford University Press, Oxford 2006.
- [9] J.L. Albacete, H. Petersen, A. Soto-Ontoso, *Phys. Lett. B* **778**, 128 (2018), [arXiv:1707.05592](#) [[hep-ph](#)].
- [10] B. Schenke, S. Jeon, C. Gale, *Phys. Rev. C* **82**, 014903 (2010), [arXiv:1004.1408](#) [[hep-ph](#)].
- [11] M. Bleicher *et al.*, *J. Phys. G: Nucl. Part. Phys.* **25**, 1895 (1999), [arXiv:hep-ph/9909407](#).
- [12] B. Schenke, C. Shen, P. Tribedy, *Phys. Rev. C* **102**, 044905 (2020), [arXiv:2005.14682](#) [[nucl-th](#)].
- [13] ALICE Collaboration (J. Adam *et al.*), *Phys. Rev. Lett.* **116**, 222302 (2016).
- [14] ALICE Collaboration (J. Adam *et al.*), *Phys. Rev. Lett.* **116**, 132302 (2016).