Physics Letters B 833 (2022) 137348

Contents lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Bayesian inference of the fluctuating proton shape

Heikki Mäntysaari [a,b,*], Björn Schenke [c], Chun Shen [d,e], Wenbin Zhao [d]

[a] Department of Physics, University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland
[b] Helsinki Institute of Physics, P.O. Box 64, 00014 University of Helsinki, Finland
[c] Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA
[d] Department of Physics and Astronomy, Wayne State University, Detroit, MI 48201, USA
[e] RIKEN BNL Research Center, Brookhaven National Laboratory, Upton, NY 11973, USA

ARTICLE INFO

Article history:
Received 15 February 2022
Received in revised form 8 July 2022
Accepted 23 July 2022
Available online 29 July 2022
Editor: J.-P. Blaizot

ABSTRACT

Using Bayesian inference, we determine probabilistic constraints on the parameters describing the fluctuating structure of protons at high energy. We employ the color glass condensate framework supplemented with a model for the spatial structure of the proton, along with experimental data from the ZEUS and H1 Collaborations on coherent and incoherent diffractive J/ψ production in e+p collisions at HERA. This data is found to constrain most model parameters well. This work sets the stage for future global analyses, including experimental data from e+p, p+p, and p+A collisions, to constrain the fluctuating structure of nucleons along with properties of the final state.

1. Introduction

Extracting the multi-dimensional structure of protons and nuclei is one of the main goals of future Deep Inelastic Scattering (DIS) facilities such as the Electron-Ion Collider [1,2], LHeC/FCC-he [3] and EicC [4]. Exclusive processes like J/ψ production are especially powerful probes of the hadron structure at a small longitudinal momentum fraction x for two reasons. First, the exclusive nature of the process requires at lowest order at least two gluons to be exchanged with the target, rendering the cross section approximately proportional to the squared gluon distribution [5]. Additionally, only in exclusive processes is it possible to measure the total momentum transfer, which is Fourier conjugate to the impact parameter and thereby provides access to the transverse geometry.

Understanding the proton structure, including its event-by-event fluctuations [6], is of fundamental interest. Additionally, knowledge of the spatial structure of the colliding objects in hadronic and heavy-ion collisions is required in order to construct realistic initial conditions that can be coupled to relativistic hydrodynamic simulations to describe the space-time evolution of the produced Quark-Gluon Plasma (QGP). Besides heavy-ion collisions, collective phenomena that can be interpreted as signatures of QGP production have been seen in small systems such as proton/deuteron/^{3}He - nucleus, proton-proton, and even photon-nucleus collisions, see [7] for a recent review. In such small collision systems with a proton projectile, the detailed fluctuation spectrum of the proton geometry is particularly important to determine if QGP is indeed produced.

It is possible to constrain the proton structure from hadronic collisions by performing a statistical analysis to extract both the transport coefficients describing the matter produced in proton-lead collisions (see e.g. Refs. [8–11]), as well as the proton's fluctuating geometry, by comparing with the LHC data as in Ref. [12]. Another approach, which we take in this work, is to use exclusive DIS data from HERA, especially exclusive vector meson production [13–17], as a complementary input to constrain the proton shape fluctuations, as initially suggested in Ref. [18]. In the future, the Electron-Ion Collider will provide a vast amount of precise vector meson production data with proton and nuclear targets that will provide further constraints on e.g. momentum fraction $x_{\mathbb{P}}$ and the nuclear mass number A dependence. Additionally, Ultra Peripheral Collisions [19,20] at RHIC and at the LHC provide access to very high energy photoproduction processes and to effects of a nuclear environment on nucleon substructure fluctuations at high energy [21,22].

In this work, we go beyond previous studies [18,23], where model parameters were constrained "by eye", and perform a Bayesian analysis to extract in a statistically rigorous manner the non-perturbative parameter values allowed by the HERA data, and construct initial conditions for hadronic collisions that are compatible with the experimental DIS data.

This paper is organized as follows. In Section 2 we review the calculation of coherent and incoherent exclusive vector meson pro-

* Corresponding author.
E-mail address: heikki.mantysaari@jyu.fi (H. Mäntysaari).

https://doi.org/10.1016/j.physletb.2022.137348

duction in the dipole picture and discuss the various aspects of our model for the proton target. In Section 3 we explain the procedure for our Bayesian analysis. We present results in Section 4 and conclude in Section 5.

2. Vector meson production at high energy

In this work we calculate vector meson production in a framework similar to the one used in Refs. [18,23] (see also Refs. [24–26]), and for completeness briefly review the calculation in this section.

At high energies, DIS processes can be conveniently described in the dipole picture in the rest frame of the target proton, and interaction with the target color field is described in the Color Glass Condensate (CGC) framework [27]. In the proton rest frame, the lifetime of a fluctuation of the incoming virtual photon into a quark-antiquark dipole is much longer than the characteristic timescale of the dipole-target interaction. Consequently, the scattering amplitude can be factorized into a convolution of photon and vector meson wave functions and the dipole-target interaction. The scattering amplitude for exclusive vector meson V production can then be written as [28,29]

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

Here $\mathbf{r}_\perp$ is the transverse size of the $q\bar{q}$ dipole, $\mathbf{b}_\perp$ is the impact parameter measured relative to the proton center, and Q^2 is the photon virtuality. The fraction of the large photon plus momentum carried by the quark is given by z, and $\boldsymbol{\Delta}_\perp$ is the transverse momentum transfer. Note that at high energies we can employ the eikonal approximation and assume that the quark transverse coordinates are fixed during the propagation through the target color field.

The $\gamma^*\to q\bar{q}$ splitting is described by the virtual photon light front wave function Ψ_γ, which can be computed from QED [30]. The vector meson wave function is non-perturbative, and in this work, we use the so-called Boosted Gaussian parametrization from [28], where the model parameters are constrained by experimental data on the vector meson decay width. We note that there are multiple vector meson wave functions proposed in the literature (see e.g. Refs. [31–33]). Different wave functions mostly affect the overall normalization of the J/ψ production cross section, and have a much smaller effect on the $|t|$ spectra, which we are most interested here [21,28,31]. Consequently, our results will depend only weakly on the specific wave function choice (except for the parameter controlling the overall proton density).

Equation (1) is a leading order result for the vector meson production in the CGC framework (note that multiple scattering effects are resummed in the dipole amplitude N_Ω). This framework is applicable in the high energy limit where the parton densities in the proton are very large and the DGLAP scale evolution can be neglected (see also e.g. Refs. [34,35] for a complementary approach based on collinear factorization where one also can study vector meson production as a function of rapidity gap size). Currently, there is rapid progress in the field toward next-to-leading order (NLO) accuracy. In particular, the cross section for the production of longitudinally polarized heavy vector mesons are now available [36] at next-to-leading order, as well as the virtual photon light front wave function [37] and small-x evolution equation [38].

However, the NLO calculations are not yet at the level where they can be consistently used in phenomenological applications (in particular the cross section for transversely polarized heavy vector meson production is still missing). As the purpose of this work is to demonstrate the potential of Bayesian analyses to systematically extract non-perturbative parameters describing the proton event-by-event fluctuating structure, we do not expect the NLO contributions to have a large effect on the results we obtain using our leading order setup.

The coherent cross section, corresponding to the process where the target proton remains in the same quantum state, can be obtained by averaging over the target color charge configurations Ω at the amplitude level [39]:

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

Subtracting the coherent contribution from the total diffractive vector meson production cross section we obtain the cross section for incoherent vector meson production, in which case the final state of the target is different from the initial state [6,40,41]. Experimentally, this corresponds to processes where the target proton (or nucleus) dissociates, but the rapidity gap between the produced vector meson and the target remnants remains. The incoherent cross section can then be written as a variance

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

Dependence on the small-x structure of the target proton is included in the dipole amplitude $N_\Omega(\mathbf{r}_\perp,\mathbf{b}_\perp,x_{\mathbb{P}})$, which, for a given target color charge configuration Ω, can be written as

$$N_\Omega(\mathbf{r}_\perp,\mathbf{b}_\perp,x_{\mathbb{P}}) = 1-\frac{1}{N_c}\mathrm{tr}\left[V\left(\mathbf{b}_\perp+\frac{\mathbf{r}_\perp}{2}\right)V^\dagger\left(\mathbf{b}_\perp-\frac{\mathbf{r}_\perp}{2}\right)\right]. \quad (4)$$

Here $V(\mathbf{x}_\perp)$ represents a Wilson line, which describes the color rotation of a quark state when it propagates through the target field (given the target color field configuration Ω) at transverse coordinate $\mathbf{x}_\perp$. We suppressed the dependence of V on

$$x_{\mathbb{P}} \approx \frac{M_V^2+Q^2}{W^2+Q^2}, \quad (5)$$

which is the fraction of the target longitudinal momentum transferred to the meson with mass M_V in the frame where the target has a large momentum. In principle $x_{\mathbb{P}}$ also depends on the squared momentum transfer t and on the proton mass, but at $W^2\gg|t|,m_N^2$ (where m_N is the invariant mass of the proton) this dependence is negligible. The photon-proton center-of-mass energy is denoted by W.

The Wilson lines are obtained in the same way as done in the IP-Glasma calculation [42] used e.g. in Ref. [43]. The color charges ρ^a are first determined using the McLerran-Venugopalan [44] model, assuming that color charges are local Gaussian random variables with expectation value zero and variance

$$g^2\left\langle\rho^a(x^-,\mathbf{x}_\perp)\rho^b(y^-,\mathbf{y}_\perp)\right\rangle = g^4\lambda_A(x^-)\delta^{ab} \times\delta^{(2)}(\mathbf{x}_\perp-\mathbf{y}_\perp)\delta(x^--y^-). \quad (6)$$

Here the color charge density is $\mu^2=\int\mathrm{d}x^-\lambda_A(x^-)$, and it is related to the local saturation scale $Q_s(\mathbf{x}_\perp)$ determined from the IPsat parametrization [45] fitted to HERA data [46]. In our Bayesian analysis, the ratio

$$\frac{Q_s(\mathbf{x}_\perp)}{g^2\mu}, \quad (7)$$

is a free parameter, controlling the overall proton density (see also Ref. [47] for a detailed study of this ratio). The Wilson lines $V(\mathbf{x}_\perp)$ can be obtained by solving the Yang-Mills equations, and the result reads

$$V(\mathbf{x}_\perp) = \mathrm{P}_- \left\{ \exp\left(-ig \int_{-\infty}^{\infty} \mathrm{d}z^- \frac{\rho^a(x^-, \mathbf{x}_\perp) t^a}{\nabla^2 - m^2} \right) \right\}, \tag{8}$$

where P_- represents path ordering in the x^- direction. Here, we introduced the infrared regulator m, which is needed to avoid the emergence of unphysical Coulomb tails, and will be another free parameter in the Bayesian analysis.

In the IPsat parametrization the saturation scale $Q_s^2(\mathbf{b}_\perp)$ is directly proportional to the local density $T_p(\mathbf{b}_\perp)$. We introduce an event-by-event fluctuating density by writing the density profile following Refs. [18,23] as:

$$T_p(\mathbf{b}_\perp) = \frac{1}{N_q} \sum_{i=1}^{N_q} p_i T_q(\mathbf{b}_\perp - \mathbf{b}_{\perp,i}), \tag{9}$$

where

$$T_q(\mathbf{b}_\perp) = \frac{1}{2\pi B_q} e^{-\mathbf{b}_\perp^2/(2B_q)}, \tag{10}$$

and the coefficient p_i allows for different normalizations for individual hot spots, to be discussed below. Our prescription corresponds to having N_q hot spots with hot spot width B_q (note that the hot spot transverse root mean square (RMS) radius in $\sqrt{2B_q}$). The hot spot positions $\mathbf{b}_{\perp,i}$ are sampled from a two-dimensional Gaussian distribution whose width is denoted by B_{qc}, and the center-of-mass is shifted to the origin in the end.

Repulsive short-range correlations between the hot spots may explain the hollowness effect and negative correlation between the v_2 and v_3 flow harmonics observed in the highest multiplicity proton-proton collisions [48,49]. In order to study if exclusive vector meson production in DIS can be used to probe such repulsive correlations, we also introduce an additional model parameter $d_{q,\mathrm{Min}}$ which controls the minimum three-dimensional distance required between any two hot spots.[1] We checked that for a large number of hot spots $N_q = 10$ in a typical nucleon of size $B_{qc} = 4.2$ GeV^{-2}, the model parameter $d_{q,\mathrm{Min}}$ remains effective more than 90% for $d_{q,\mathrm{Min}} \leq 0.4$ fm, meaning that in 10% of the sampled configurations the distance requirement cannot be fulfilled.

Finally, we include saturation scale fluctuations by allowing the local density of each hot spot to fluctuate independently, following again Refs. [18,23] (see also Ref. [51]). These fluctuations are implemented by sampling the coefficients p_i in Eq. (9) from the log-normal distribution

$$P(\ln p_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\ln^2 p_i}{2\sigma^2} \right]. \tag{11}$$

The sampled p_i are at the end normalized by the expectation value of the distribution $E[p_i] = e^{\sigma^2/2}$ in order to keep the average density unmodified. The magnitude of density fluctuations is controlled by the parameter σ.

[1] To implement the minimal distance we follow [50], first sampling 3D distributions and if necessary resampling the solid angle until the requirement posed by $d_{q,\mathrm{Min}}$ is satisfied.

3. Bayesian analysis setup

Bayesian Inference 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 $\mathbf{y}_{\mathrm{exp}}$ [52]. Bayes' theorem provides the posterior distribution of model parameters as

$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{exp}}) \propto \mathcal{P}(\mathbf{y}_{\mathrm{exp}}|\boldsymbol{\theta}) \mathcal{P}(\boldsymbol{\theta}). \tag{12}$$

Here $\mathcal{P}(\mathbf{y}_{\mathrm{exp}}|\boldsymbol{\theta})$ is the likelihood for model results with parameter $\boldsymbol{\theta}$ to agree with the experimental data. We choose a multivariate normal distribution for the logarithm of the likelihood with $\Delta\mathbf{y}(\boldsymbol{\theta}) = \mathbf{y}(\boldsymbol{\theta}) - \mathbf{y}_{\mathrm{exp}}$ [53],

$$\ln[\mathcal{P}(\mathbf{y}_{\mathrm{exp}}|\boldsymbol{\theta})] = -\frac{1}{2} \Delta\mathbf{y}(\boldsymbol{\theta})^T \Sigma^{-1} \Delta\mathbf{y}(\boldsymbol{\theta}) - \frac{1}{2} \ln[(2\pi)^n \det \Sigma]. \tag{13}$$

Here n is the number of experimental data points and $\Sigma \equiv \Sigma_{\mathrm{exp}} + \Sigma_{\mathrm{model}}$ is the $n \times n$ covariance matrix, which encodes experimental and model uncertainties. In the current analysis, we assume no correlation among experimental errors of the n observables $\{\sigma_i\}$. So the covariance matrix for experimental uncertainty takes a diagonal form,

$$\Sigma_{\mathrm{exp}} = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_n^2). \tag{14}$$

Our model uncertainties Σ_{model} are estimated using the covariance matrix from the trained Gaussian Process (GP) Emulators [54].

We employ GP emulators [53] for our model and couple them with the Monte-Carlo Markov Chain (MCMC) method to efficiently explore the model parameter space [55,56]. The HERA measurements can be represented by five Principal Components (PC) with a residual variance of less than 0.01%, meaning that 99.99% of the variation of all studied observables within the prior parameter range are captured by the five principal components. Our GP emulators are trained to fit these five PCs with 1,000 training simulations in the model parameter space. In each model parameter point, we generate 3,000 configurations to compute the coherent and incoherent cross sections. The relative statistical errors are within 5%.

All model parameters and their prior ranges (and Maximum a Posterior values that are discussed in Sec. 4) are summarized in Table 1. We treat the parameter N_q as a continuous real number. The fractional part of N_q is treated as a probability to sample either $\lceil N_q \rceil$ or $\lfloor N_q \rfloor$ partons inside protons. The same approach was recently used in Ref. [57]. The experimental data included in the Bayesian analysis is the H1 data on coherent and incoherent J/ψ production cross section measured at $W = 75$ GeV [13]. The incoherent data is included in the $|t|$ range $0 < |t| < 2.5$ GeV^2. We note that there is incoherent data at higher $|t|$ also (studied in a similar context in Ref. [24]), but the highest $|t|$ points are not included in our analysis for two reasons. First, as we determine Wilson lines at fixed $x_{\mathbb{P}}$, as discussed in Sec. 2, and do not include the full $x_{\mathbb{P}}$ evolution, we neglect $|t|$ dependence in $x_{\mathbb{P}}$ (see Eq. (5)). Additionally, at large $|t|$ other effects such as DGLAP evolution [58–60] may become important.

4. Results

For two distinct scenarios, the first with N_q fixed to 3 (upper right and red lines), the second with N_q a free parameter (lower left and blue lines), the posterior distribution of model parameters is shown in Fig. 1. Particularly for the case of $N_q = 3$, most of the model parameters are tightly constrained by the H1 data included

Table 1
Summary of model parameters, their prior ranges, and constrained maximum likelihood values with uncertainty estimates in 90% credible intervals.

Parameter	Description	Prior range	MAP (variable N_q)	MAP ($N_q \equiv 3$)
m [GeV]	Infrared regulator	[0.05, 2]	$0.506^{+1.12}_{-0.356}$	$0.246^{+0.162}_{-0.103}$
B_{qc} [GeV^{-2}]	Proton size	[1, 10]	$4.02^{+1.73}_{-0.728}$	$4.45^{+0.801}_{-0.803}$
B_q [GeV^{-2}]	Hot spot size	[0.1, 3]	$0.474^{+0.434}_{-0.286}$	$0.346^{+0.282}_{-0.202}$
σ	Magnitude of Q_s fluctuations	[0, 1.5]	$0.833^{+0.194}_{-0.441}$	$0.563^{+0.143}_{-0.141}$
$Q_s/(g^2\mu)$	Ratio of color charge density and saturation scale	[0.2, 1.5]	$0.598^{+0.230}_{-0.264}$	$0.747^{+0.0704}_{-0.0930}$
$d_{q,\mathrm{Min}}$ [fm]	Minimum 3D distance between hot spots	[0, 0.5]	$0.257^{+0.221}_{-0.231}$	$0.254^{+0.222}_{-0.229}$
N_q	Number of hot spots	[1, 10]	$6.79^{+2.93}_{-4.83}$	3

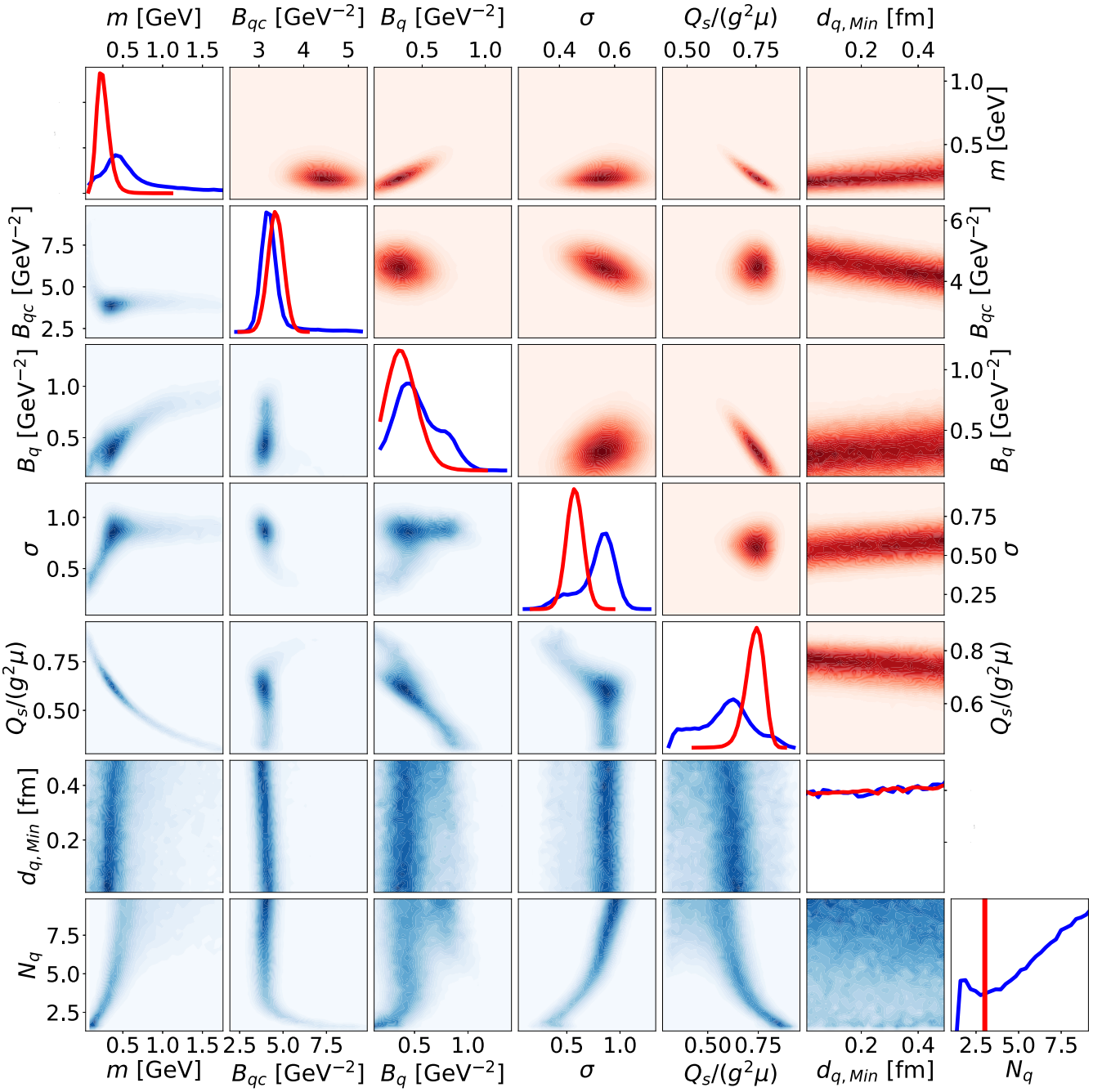

Fig. 1. Bayesian posterior distributions of the model parameters. The diagonal panels show the probability distributions for individual parameters, and off-diagonal panels illustrate their pairwise correlations. Upper right panels and red lines for $N_q = 3$, lower left panels and blue lines for variable N_q.

in the Bayesian analysis. This is due to the fact that different regions of the dataset are sensitive to different model parameters.

First, the infrared regulator m suppresses long-distance Coulomb tails, and as such, it controls the shape of the proton at large distances. This part of the proton geometry is probed by coherent diffraction at low $|t| \lesssim 0.2$ GeV^2 [23]: impact parameter and momentum transfers are Fourier conjugates, and consequently the low-$|t|$ region is sensitive to large distances and vice versa. On the other hand, the actual proton size controlled mostly by B_{qc} determines the overall slope of the coherent spectrum. The hot spot size B_q then determines the slope of the incoherent cross section in the $|t| \gtrsim 1$ GeV^2 region: as shown in Ref. [61] the slope of the incoherent spectra at high $|t|$ is given by the size of the smallest fluctuating constituent.

The overall normalization is determined by the $Q_s/(g^2\mu)$ parameter and the magnitude of the cross sections constrains that, however, it can not be determined very precisely from our analysis as it is strongly correlated with many other parameters, particularly for the case that N_q is a free parameter. Here we note that there is some model uncertainty related to the non-perturbative vector meson wave functions, and different phenomenological parametrizations can result in cross sections that differ by $\sim 20\%$, see e.g. [21,28,31]. In phenomenological analyses, the so-called skewness correction [28] is sometimes included, which can also have a numerically significant effect on the cross section. In this work we include an approximative 42% skewedness correction following Ref. [23]. The small real part correction discussed e.g. in Ref. [28] is neglected.

One can see that when leaving N_q variable, its value can not be constrained in our analysis, except that $N_q \geq 2$ is required in order to get geometry fluctuations that are necessary to describe the incoherent HERA data. At first one would expect the configurations with large N_q to be so smooth that event-by-event fluctuations would not be enough to result in a large enough incoherent cross section. However, we note that there is a strong positive correlation between N_q and σ, which means that large N_q goes along with large hot spot density fluctuations. In this situation one can not really interpret N_q as the number of hot spots, but one has to consider effective hot spots, that are generated dynamically from the sum of the N_q constituents that are each strongly fluctuating in magnitude. This effective hot spot number will generally be smaller than N_q. For a quantitative analysis a hot spot finding prescription, similar to jet clustering algorithms, would be needed. Additional constraints for σ originate from the incoherent cross section at small $|t|$, which is sensitive to the fluctuations at long-distance scales, i.e., overall density fluctuations [18,23].

The minimum distance required between the hot spots in three dimensions, $d_{q,\mathrm{Min}}$, can not be constrained at all in our analysis. This parameter is also only very weakly correlated with the other parameters, emphasizing its limited effect on the results. This means that the J/ψ production data allows, but does not require, repulsive short-range correlations used e.g. in Refs. [48,49]. We note that for large N_q it is not always possible to fulfill the minimum distance requirement, reducing the effective value of $d_{q,\mathrm{Min}}$ that is actually employed.

Let us next study correlations between the model parameters in more detail. First, we observe that there is a clear positive correlation between the infrared regulator m and the hot spot size B_q. This can be understood, as increasing m suppresses color fields at large distances and as such results in smaller hot spots. Interestingly, we find no clear correlation between the overall proton size parameter B_{qc} and the infrared regulator m. In fact, the proton size B_{qc} shows no clear correlation with any of the other model parameters. The infrared regulator m is strongly anti-correlated with $Q_s/(g^2\mu)$. This is again easy to understand: large m results in a reduction of the normalization of the cross section, which has to be compensated by using a larger color charge density which requires smaller $Q_s/(g^2\mu)$.

We also find a strong negative correlation between the number of hot spots N_q and the proton density $Q_s/(g^2\mu)$. This correlation can be understood by considering the effect of these parameters on the incoherent cross section. When the number of hot spots with the same size increases, the incoherent slope is not significantly affected but the normalization goes down as there are smaller fluctuations in the scattering amplitude due to the smoother proton profile. This would result in the incoherent cross section being underestimated, which has to be compensated by smaller $Q_s/(g^2\mu)$.

The negative correlation between the hot spot size B_q and the proton density $Q_s/(g^2\mu)$ can also be understood. First, we note that large hot spots (large B_q) require a larger infrared regulator m which effectively makes the hot spots smaller (notice a positive correlation between B_q and m). Then, to compensate for the effect of the larger infrared regulator, which suppresses the cross section, a smaller $Q_s/(g^2\mu)$ is needed to increase the cross section again.

The posterior distribution of parameters B_{qc} and B_q, that determine the proton size, which can be quantified e.g. by the two dimensional RMS radius, $r_{\mathrm{rms}} = \sqrt{2(B_{\mathrm{qc}} + B_q)}$, are shown in the second and third diagonal panels of Fig. 1. We find that they are sharply peaked, in particular for the case of fixed $N_q = 3$. The extracted gluonic proton two-dimensional radius from this Bayesian analysis is $r_{\mathrm{rms}} = 0.591^{+0.130}_{-0.071}$ fm ($0.610^{+0.065}_{-0.068}$ fm) for variable N_q ($N_q \equiv 3$) with uncertainty estimates in 90% credible intervals.

We further included the possibility of fluctuating values of B_q and N_q, and studied the dependence of the observables on the respective variances. We found that the considered experimental data could not constrain these parameters and decided not to include them in the presented analysis.

Next, we demonstrate explicitly that sampling parameter values from the posterior distribution lead to a good description of the HERA data. The main result of the Bayesian analysis is the posterior distribution shown in Fig. 1, but it is also possible to find the so-called Maximum a Posteriori (MAP) parameter set, which is the mode of the posterior distribution. Because our prior parameter distributions are uniform, the MAP parameters shown in Table 1 maximize the likelihood function and provide the best fit to the experimental data. We note that statistically the expectation value for any observable is not obtained using the MAP parameters. Instead, one should calculate observables using averages over parameter samples obtained from the posterior distribution.

Comparison to the HERA coherent and incoherent J/ψ production data measured at $W = 75$ GeV [13] and used in the Bayesian analysis to constrain the model parameters are shown in panel (a) of Fig. 2. The spectra calculated by averaging the results computed using different parametrizations sampled from the posterior distribution indeed provide an excellent description of the data. We also show the statistical uncertainty, obtained by calculating the one standard deviation interval shown as red (coherent) and blue (incoherent) bands.

Next, we study compatibility with experimental data not included in the Bayesian analysis. We do not include full small-x evolution e.g. by means of the JIMWLK equation (see [62]) in this work. Consequently, the only center-of-mass energy W (or momentum fraction $x_{\mathbb{P}}$) dependence comes from the $x_{\mathbb{P}}$ dependence of the saturation scale Q_s determined from the IPsat fit. This almost only affects the overall normalization of the calculated spectra, and misses important physical effects such as the growth of the proton with decreasing x. Consequently, we do not expect that within our setup we can describe J/ψ production at different center-of-mass energies simultaneously. Nevertheless, the geometry evolution between $W = 75$ GeV and $W = 100$ GeV should be weak enough to make predictions for the higher center-of-mass energy.

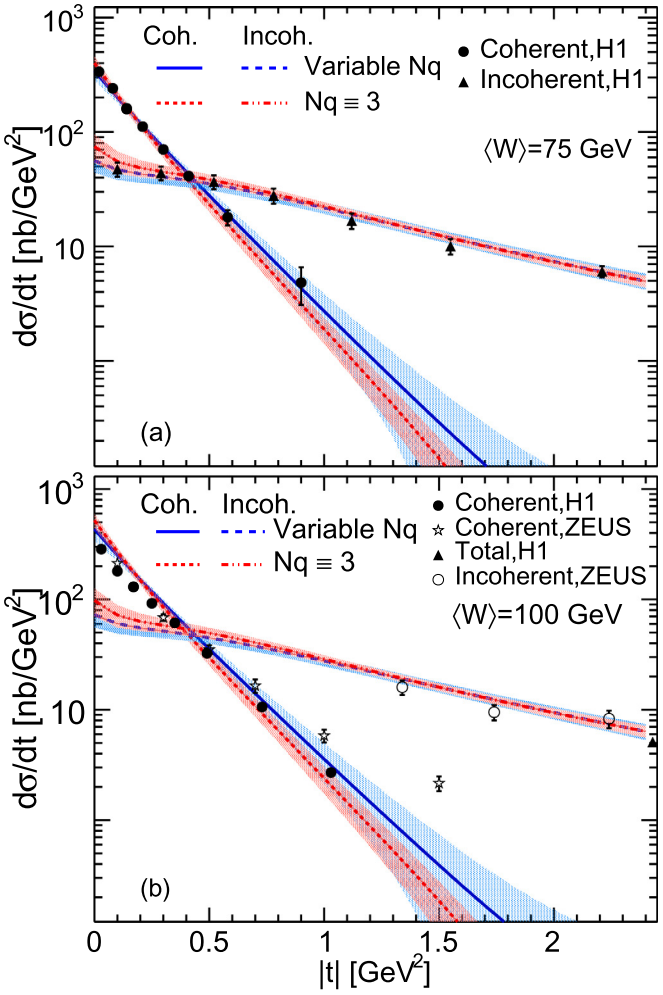

Fig. 2. Coherent and incoherent J/ψ photoproduction cross sections at $\langle W \rangle = 75$ GeV (a) and $\langle W \rangle = 100$ GeV (b), calculated by averaging over many parameter sets sampled from the posterior distribution, which was determined using the $\langle W \rangle = 75$ GeV HERA data [13]. We compare the case with variable N_q to that with fixed $N_q = 3$, and to experimental data at $\langle W \rangle = 75$ GeV [13] and $\langle W \rangle = 100$ GeV [14–17], respectively. The bands show the one standard deviation uncertainty.

The calculated J/ψ spectra at $W = 100$ GeV are shown in panel (b) of Fig. 2 and compared with the data from H1 and ZEUS collaborations [14,16,17,63]. Again, we show the model prediction as the average over many samples of the posterior distribution and provide one standard deviation bands. The results are very similar to the $W = 75$ GeV case studied above, and the HERA data is well described (except the coherent cross section at low $|t|$, and high $|t|$ in the case of the ZEUS measurement), and variation between the different parametrizations sampled from the posterior distribution is small. Similarity to the $W = 75$ GeV case is not surprising, as we use exactly the same fluctuating geometry, with the only difference being slightly larger Q_s values, extracted from the IPsat parametrization.

5. Conclusions

We have performed a statistically rigorous Bayesian analysis to extract posterior likelihood distributions for the non-perturbative parameters describing the event-by-event fluctuating proton geometry as constrained by exclusive J/ψ production data from HERA. We presented a comparison of the $|t|$-dependent coherent and incoherent cross sections, obtained from an average over many parameter samples from the posterior distribution, to the experimental data.

Generally, model parametrizations sampled from the determined posterior distribution can be used to systematically take into account uncertainties in the proton geometry as constrained by the HERA DIS data when calculating any other observable that depends on the proton geometry, such as flow observables in high-multiplicity proton-proton and proton-nucleus collisions. To enable such studies, we provide 1,000 parametrizations sampled from the determined posterior distributions in the supplemental material.

Because the model parameters are sensitive to different aspects of the coherent and incoherent vector meson production spectra, most of them are well constrained by the Bayesian analysis. The only exceptions are the minimum distance between the hot spots (repulsive short-range correlations), and the number of hot spots N_q, which cannot be well constrained by the considered HERA data. However, we note that although the analysis suggests that large N_q is compatible with the HERA data, for $N_q \gtrsim 5$ one should no longer simply interpret N_q as the number of actual hot spots. In this regime, a good description of the data requires large fluctuations of individual hot spots' densities, and the sampled 'hot spots' can also overlap significantly. Thus, the effective number of hot spots is significantly smaller than the parameter N_q might imply.

The HERA data used in this work probes the proton structure at $x \sim 10^{-3}$. The energy (Bjorken-x) dependence can be included in terms of JIMWLK evolution as e.g. in Refs. [64,65]. In the future we plan to extend our framework by including the full JIMWLK evolution and vector meson production data at different center-of-mass energies from HERA [13,63,66] and from the ultra peripheral proton-lead collisions measured at the LHC [67–69], allowing us to extract also the Bjorken-x dependence of the fluctuating proton geometry.

In addition to constraining the energy dependence, performing global analyses including both exclusive vector meson production data from HERA and flow harmonics from the proton-proton and proton-lead collisions measured at the LHC (including a model calculation along the lines of [70]) would allow for a powerful global analysis of the fluctuating nucleon substructure and properties of the final state. More differential DIS measurements from the future EIC such as dijet [71] or lepton-meson angular correlations [43] can also provide further constraints and can in principle be included in our framework in a straightforward manner.

The numerical framework for our physics models and the Bayesian analysis package are publicly available on Github [72,73]. Our Bayesian analysis code is developed based on the open-source numerical package by the Duke group [10]. To help visualize how observables depend on the model parameters, we provide an interactive web page with the trained GP emulators [74], where one can also find the posterior samples included in the supplemental material.

The presented structure of the incoherent cross section in the CGC approach is notably different from that obtained in the collinear factorization approaches, where at $|t|$ greater than a few GeV^2 it is proportional to the (generalized) parton distribution functions [34,35]. However, there should be a kinematical domain where both approaches are justified, and properly understanding the mapping between the coordinate space description with event-by-event fluctuations and the momentum space collinear factorization approach would be extremely useful.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Heikki Mäntysaari reports financial support was provided

by Academy of Finland. Heikki Mäntysaari reports financial support was provided by Horizon 2020. Bjoern Schenke reports financial support was provided by US Department of Energy. Chun Shen reports financial support was provided by US Department of Energy. Wenbin Zhao reports financial support was provided by National Science Foundation.

Data availability

Determined posterior distributions are available in the supplementary material.

Acknowledgements

B.P.S. and C.S. are supported by the U.S. Department of Energy Office of Science, Office of Nuclear Physics, under DOE 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, and under the European Union's Horizon 2020 research and innovation programme by the European Research Council (ERC, grant agreement No. ERC-2018-ADG-835105 YoctoLHC) and by the STRONG-2020 project (grant agreement No. 824093). W.B.Z. is supported by the National Science Foundation (NSF) under grant numbers ACI-2004571 within the framework of the XSCAPE project of the JETSCAPE collaboration. The content of this article does not reflect the official opinion of the European Union and responsibility for the information and views expressed therein lies entirely with the authors. This research was done using resources provided by the Open Science Grid (OSG) [75,76], which is supported by the National Science Foundation award #2030508.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.physletb.2022.137348.

References

[1] R. Abdul Khalek, et al., Science requirements and detector concepts for the electron-ion collider: EIC yellow report, arXiv:2103.05419 [physics.ins-det].
[2] E.C. Aschenauer, S. Fazio, J.H. Lee, H. Mäntysaari, B.S. Page, B. Schenke, T. Ullrich, R. Venugopalan, P. Zurita, The electron–ion collider: assessing the energy dependence of key measurements, Rep. Prog. Phys. 82 (2) (2019) 024301, arXiv:1708.01527 [nucl-ex].
[3] LHeC, FCC-he Study Group collaboration, P. Agostini, et al., The large hadron-electron collider at the HL-LHC, J. Phys. G 48 (11) (2021) 110501, arXiv:2007.14491 [hep-ex].
[4] D.P. Anderle, et al., Electron-ion collider in China, Front. Phys. (Beijing) 16 (6) (2021) 64701, arXiv:2102.09222 [nucl-ex].
[5] M.G. Ryskin, Diffractive J/ψ electroproduction in LLA QCD, Z. Phys. C 57 (1993) 89.
[6] H. Mäntysaari, Review of proton and nuclear shape fluctuations at high energy, Rep. Prog. Phys. 83 (8) (2020), arXiv:2001.10705 [hep-ph].
[7] B. Schenke, The smallest fluid on Earth, Rep. Prog. Phys. 84 (8) (2021), arXiv:2102.11189 [nucl-th].
[8] JETSCAPE collaboration, D. Everett, et al., Multisystem Bayesian constraints on the transport coefficients of QCD matter, Phys. Rev. C 103 (5) (2021) 054904, arXiv:2011.01430 [hep-ph].
[9] JETSCAPE collaboration, D. Everett, et al., Phenomenological constraints on the transport properties of QCD matter with data-driven model averaging, Phys. Rev. Lett. 126 (24) (2021) 242301, arXiv:2010.03928 [hep-ph].
[10] J.E. Bernhard, J.S. Moreland, S.A. Bass, Bayesian estimation of the specific shear and bulk viscosity of quark–gluon plasma, Nat. Phys. 15 (11) (2019) 1113.
[11] J.E. Parkkila, A. Onnerstad, F. Taghavi, C. Mordasini, A. Bilandzic, D.J. Kim, New constraints for QCD matter from improved Bayesian parameter estimation in heavy-ion collisions at LHC, arXiv:2111.08145 [hep-ph].
[12] J.S. Moreland, J.E. Bernhard, S.A. Bass, Bayesian calibration of a hybrid nuclear collision model using p-Pb and Pb-Pb data at energies available at the CERN large hadron collider, Phys. Rev. C 101 (2) (2020) 024911, arXiv:1808.02106 [nucl-th].
[13] H1 collaboration, C. Alexa, et al., Elastic and proton-dissociative photoproduction of J/ψ mesons at HERA, Eur. Phys. J. C 73 (6) (2013) 2466, arXiv:1304.5162 [hep-ex].
[14] H1 collaboration, A. Aktas, et al., Elastic J/ψ production at HERA, Eur. Phys. J. C 46 (2006) 585, arXiv:hep-ex/0510016.
[15] ZEUS collaboration, S. Chekanov, et al., Exclusive photoproduction of J/ψ mesons at HERA, Eur. Phys. J. C 24 (2002) 345, arXiv:hep-ex/0201043.
[16] ZEUS collaboration, S. Chekanov, et al., Measurement of proton dissociative diffractive photoproduction of vector mesons at large momentum transfer at HERA, Eur. Phys. J. C 26 (2003) 389, arXiv:hep-ex/0205081.
[17] H1 collaboration, A. Aktas, et al., Diffractive photoproduction of J/ψ mesons with large momentum transfer at HERA, Phys. Lett. B 568 (2003) 205, arXiv:hep-ex/0306013.
[18] H. Mäntysaari, B. Schenke, Evidence of strong proton shape fluctuations from incoherent diffraction, Phys. Rev. Lett. 117 (5) (2016) 052301, arXiv:1603.04349 [hep-ph].
[19] S.R. Klein, H. Mäntysaari, Imaging the nucleus with high-energy photons, Nat. Rev. Phys. 1 (11) (2019) 662, arXiv:1910.10858 [hep-ex].
[20] C.A. Bertulani, S.R. Klein, J. Nystrand, Physics of ultra-peripheral nuclear collisions, Annu. Rev. Nucl. Part. Sci. 55 (2005) 271, arXiv:nucl-ex/0502005.
[21] H. Mäntysaari, B. Schenke, Probing subnucleon scale fluctuations in ultraperipheral heavy ion collisions, Phys. Lett. B 772 (2017) 832, arXiv:1703.09256 [hep-ph].
[22] B. Sambasivam, T. Toll, T. Ullrich, Investigating saturation effects in ultraperipheral collisions at the LHC with the color dipole model, Phys. Lett. B 803 (2020) 135277, arXiv:1910.02899 [hep-ph].
[23] H. Mäntysaari, B. Schenke, Revealing proton shape fluctuations with incoherent diffraction at high energy, Phys. Rev. D 94 (3) (2016) 034042, arXiv:1607.01711 [hep-ph].
[24] A. Kumar, T. Toll, Investigating the structure of gluon fluctuations in the proton with incoherent diffraction at HERA, arXiv:2106.12855 [hep-ph].
[25] J. Cepila, J.G. Contreras, M. Krelina, J.D. Tapia Takaki, Mass dependence of vector meson photoproduction off protons and nuclei within the energy-dependent hot-spot model, Nucl. Phys. B 934 (2018) 330, arXiv:1804.05508 [hep-ph].
[26] M.C. Traini, J.-P. Blaizot, Diffractive incoherent vector meson production off protons: a quark model approach to gluon fluctuation effects, Eur. Phys. J. C 79 (4) (2019) 327, arXiv:1804.06110 [hep-ph].
[27] E. Iancu, R. Venugopalan, The color glass condensate and high-energy scattering in QCD, Quark Gluon Plasma 3 (2003) 249, arXiv:hep-ph/0303204 [hep-ph].
[28] H. Kowalski, L. Motyka, G. Watt, Exclusive diffractive processes at HERA within the dipole picture, Phys. Rev. D 74 (2006) 074016, arXiv:hep-ph/0606272.
[29] Y. Hatta, B.-W. Xiao, F. Yuan, Gluon tomography from deeply virtual compton scattering at small-x, Phys. Rev. D 95 (11) (2017) 114026, arXiv:1703.02085 [hep-ph].
[30] Y.V. Kovchegov, E. Levin, Quantum Chromodynamics at High Energy, vol. 33, Cambridge University Press, 8 2012.
[31] T. Lappi, H. Mäntysaari, J. Penttala, Relativistic corrections to the vector meson light front wave function, Phys. Rev. D 102 (5) (2020) 054020, arXiv:2006.02830 [hep-ph].
[32] Y. Li, P. Maris, J.P. Vary, Quarkonium as a relativistic bound state on the light front, Phys. Rev. D 96 (1) (2017) 016022, arXiv:1704.06968 [hep-ph].
[33] M. Li, Y. Li, G. Chen, T. Lappi, J.P. Vary, Light-front wavefunctions of mesons by design, arXiv:2111.07087 [hep-ph].
[34] J.R. Forshaw, M.G. Ryskin, Diffractive vector meson production at large momentum transfer, Z. Phys. C 68 (1995) 137, arXiv:hep-ph/9501376.
[35] M. Deák, A.M. Stašto, M. Strikman, High $|t|$ diffractive vector meson production at the EIC, Phys. Rev. D 103 (1) (2021) 014022, arXiv:2011.04711 [hep-ph].
[36] H. Mäntysaari, J. Penttala, Exclusive heavy vector meson production at next-to-leading order in the dipole picture, Phys. Lett. B 823 (2021) 136723, arXiv:2104.02349 [hep-ph].
[37] G. Beuf, T. Lappi, R. Paatelainen, Massive quarks at one loop in the dipole picture of deep inelastic scattering, arXiv:2112.03158 [hep-ph].
[38] I. Balitsky, G.A. Chirilli, Next-to-leading order evolution of color dipoles, Phys. Rev. D 77 (2008) 014019, arXiv:0710.4330 [hep-ph].
[39] M.L. Good, W.D. Walker, Diffraction dissociation of beam particles, Phys. Rev. 120 (1960) 1857.
[40] H.I. Miettinen, J. Pumplin, Diffraction scattering and the parton structure of hadrons, Phys. Rev. D 18 (1978) 1696.
[41] A. Caldwell, H. Kowalski, Investigating the gluonic structure of nuclei via J/ψ scattering, Phys. Rev. C 81 (2010) 025203, arXiv:0909.1254.
[42] B. Schenke, P. Tribedy, R. Venugopalan, Fluctuating glasma initial conditions and flow in heavy ion collisions, Phys. Rev. Lett. 108 (2012) 252301, arXiv:1202.6646 [nucl-th].
[43] H. Mäntysaari, K. Roy, F. Salazar, B. Schenke, Gluon imaging using azimuthal correlations in diffractive scattering at the electron-ion collider, Phys. Rev. D 103 (9) (2021) 094026, arXiv:2011.02464 [hep-ph].
[44] L.D. McLerran, R. Venugopalan, Computing quark and gluon distribution functions for very large nuclei, Phys. Rev. D 49 (1994) 2233, arXiv:hep-ph/9309289.
[45] H. Kowalski, D. Teaney, An impact parameter dipole saturation model, Phys. Rev. D 68 (2003) 114005, arXiv:hep-ph/0304189.

[46] A.H. Rezaeian, M. Siddikov, M. Van de Klundert, R. Venugopalan, Analysis of combined HERA data in the impact-parameter dependent saturation model, Phys. Rev. D 87 (3) (2013) 034002, arXiv:1212.2974 [hep-ph].
[47] T. Lappi, Wilson line correlator in the MV model: relating the glasma to deep inelastic scattering, Eur. Phys. J. C 55 (2008) 285, arXiv:0711.3039 [hep-ph].
[48] J.L. Albacete, A. Soto-Ontoso, Hot spots and the hollowness of proton–proton interactions at high energies, Phys. Lett. B 770 (2017) 149, arXiv:1605.09176 [hep-ph].
[49] J.L. Albacete, H. Petersen, A. Soto-Ontoso, Symmetric cumulants as a probe of the proton substructure at LHC energies, Phys. Lett. B 778 (2018) 128, arXiv:1707.05592 [hep-ph].
[50] J.S. Moreland, J.E. Bernhard, S.A. Bass, Alternative ansatz to wounded nucleon and binary collision scaling in high-energy nuclear collisions, Phys. Rev. C 92 (1) (2015) 011901, arXiv:1412.4708 [nucl-th].
[51] L. McLerran, P. Tribedy, Intrinsic fluctuations of the proton saturation momentum scale in high multiplicity p+p collisions, Nucl. Phys. A 945 (2016) 216, arXiv:1508.03292 [hep-ph].
[52] D. Sivia, J. Skilling, Data Analysis: a Bayesian Tutorial, OUP, Oxford, 2006.
[53] C.K. Williams, C.E. Rasmussen, Gaussian Processes for Machine Learning, vol. 2, MIT Press, Cambridge, MA, 2006.
[54] J.E. Bernhard, P.W. Marcy, C.E. Coleman-Smith, S. Huzurbazar, R.L. Wolpert, S.A. Bass, Quantifying properties of hot and dense QCD matter through systematic model-to-data comparison, Phys. Rev. C 91 (5) (2015) 054910, arXiv:1502.00339 [nucl-th].
[55] J. Goodman, J. Weare, Ensemble samplers with affine invariance, Commun. Appl. Math. Comput. Sci. 5 (1) (2010) 65.
[56] D. Foreman-Mackey, D.W. Hogg, D. Lang, J. Goodman, emcee: the mcmc hammer, Publ. Astron. Soc. Pac. 125 (925) (2013) 306.
[57] G. Nijs, W. van der Schee, Predictions and postdictions for relativistic lead and oxygen collisions with *Trajectum*, arXiv:2110.13153 [nucl-th].
[58] V.N. Gribov, L.N. Lipatov, Deep inelastic e p scattering in perturbation theory, Sov. J. Nucl. Phys. 15 (1972) 438.
[59] G. Altarelli, G. Parisi, Asymptotic freedom in parton language, Nucl. Phys. B 126 (1977) 298.
[60] Y.L. Dokshitzer, Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics, Sov. Phys. JETP 46 (1977) 641.
[61] T. Lappi, H. Mäntysaari, Incoherent diffractive J/ψ production in high energy nuclear DIS, Phys. Rev. C 83 (2011) 065202, arXiv:1011.1988 [hep-ph].
[62] A.H. Mueller, A simple derivation of the JIMWLK equation, Phys. Lett. B 523 (2001) 243, arXiv:hep-ph/0110169.
[63] ZEUS collaboration, S. Chekanov, et al., Exclusive photoproduction of J/ψ mesons at HERA, Eur. Phys. J. C 24 (2002) 345, arXiv:hep-ex/0201043.
[64] S. Schlichting, B. Schenke, The shape of the proton at high energies, Phys. Lett. B 739 (2014) 313, arXiv:1407.8458 [hep-ph].
[65] H. Mäntysaari, B. Schenke, Confronting impact parameter dependent JIMWLK evolution with HERA data, Phys. Rev. D 98 (3) (2018) 034013, arXiv:1806.06783 [hep-ph].
[66] H1 collaboration, A. Aktas, et al., Elastic J/ψ production at HERA, Eur. Phys. J. C 46 (2006) 585, arXiv:hep-ex/0510016.
[67] LHCb collaboration, R. Aaij, et al., Updated measurements of exclusive J/ψ and $\psi(2S)$ production cross-sections in pp collisions at $\sqrt{s} = 7$ TeV, J. Phys. G 41 (2014) 055002, arXiv:1401.3288 [hep-ex].
[68] ALICE collaboration, S. Acharya, et al., Energy dependence of exclusive J/ψ photoproduction off protons in ultra-peripheral p–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV, Eur. Phys. J. C 79 (5) (2019) 402, arXiv:1809.03235 [nucl-ex].
[69] LHCb collaboration, R. Aaij, et al., Central exclusive production of J/ψ and $\psi(2S)$ mesons in pp collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 10 (2018) 167, arXiv:1806.04079 [hep-ex].
[70] H. Mäntysaari, B. Schenke, C. Shen, P. Tribedy, Imprints of fluctuating proton shapes on flow in proton-lead collisions at the LHC, Phys. Lett. B 772 (2017) 681, arXiv:1705.03177 [nucl-th].
[71] H. Mäntysaari, N. Mueller, B. Schenke, Diffractive dijet production and Wigner distributions from the color glass condensate, Phys. Rev. D 99 (7) (2019), arXiv:1902.05087 [hep-ph].
[72] The overarching framework for IPGlasma + subnucleonic diffraction model is available at https://github.com/chunshen1987/IPGlasmaFramework (v1.0.0). It uses the open-source code packages IP-Glasma https://github.com/schenke/ipglasma and the program to compute exclusive J/ψ production cross-section https://github.com/hejajama/subnucleondiffraction.
[73] The numerical package for our Bayesian analysis is available at https://github.com/chunshen1987/bayesian_analysis/releases/tag/v1.0.0.
[74] An interactive emulator for our model is available at https://share.streamlit.io/chunshen1987/ipglasmadiffractionstreamlit/main/IPGlasmaDiffraction_app.py.
[75] R. Pordes, et al., The open science grid, J. Phys. Conf. Ser. 78 (2007) 012057.
[76] I. Sfiligoi, D.C. Bradley, B. Holzman, P. Mhashilkar, S. Padhi, F. Wurthwrin, The pilot way to grid resources using glideinWMS, WRI World Congr. 2 (2009) 428.