

Initial-state geometry and multiplicity distributions in pp and $p\text{Pb}$ collisions

R. Terra¹, A. V. Giannini^{2,3} and F. S. Navarra¹

¹*Instituto de Física, Universidade de São Paulo, Rua do Matão, 1371, CEP 05508-090, São Paulo, SP, Brazil*

²*Faculdade de Ciências Exatas e Tecnologia, Universidade Federal da Grande Dourados, CEP 79804-970, Dourados, MS, Brazil*

³*Departamento de Física, Universidade do Estado de Santa Catarina, 89219-710 Joinville, SC, Brazil*



(Received 18 November 2025; accepted 23 March 2026; published 6 May 2026)

This work investigates the possibility of accessing the initial geometric shape of the proton in proton-proton and proton-nucleus collisions at the LHC. In particular, we look for manifestations of the configuration in which the proton is made of three quarks linked by a Y-shape gluon string, called baryon junction. This initial state spatial configuration has been used in the past to describe data on baryon rapidity distributions, diffractive J/ψ production and multiplicity distributions in pp collisions. In spite of its success in explaining the data, the evidence of the baryon junction still needs confirmation. Further studies will be undertaken at the electron-ion collider. In this work we study multiplicity distributions measured in pp and $p\text{Pb}$ collisions. Different initial state geometries are used as input in a Monte Carlo event generator which implements the k_T -factorization formalism of the color glass condensate with Kharzeev-Levin-Nardi unintegrated gluon distributions. The results show that the data on multiplicity distributions are good enough to discriminate between different initial state geometries. Moreover, they indicate that it is crucial to take into account the intrinsic fluctuations of the saturation scale.

DOI: [10.1103/z423-ch4g](https://doi.org/10.1103/z423-ch4g)

I. INTRODUCTION

A. The baryon junction

Despite all the progress, we are still far from a complete picture of the nucleon. Determining the geometric structure of the proton is one of the great challenges in high energy physics. It is well established that the proton consists of quarks and gluons, collectively called partons and it is important to understand how these partons are distributed within the proton.

Long ago [1], Rossi and Veneziano suggested that the nucleon is made by three quarks located in the vertices of a Y-shaped gluon string, or baryon junction (BJ). The intermediate (Fermat) point was introduced in order to guarantee the gauge invariance of the baryon wave function. This picture received support later, when Y-shaped flux tubes were found in quenched lattice quantum chromodynamics (QCD) simulations [2–5] of a system of three static quarks.

The BJ picture of the nucleon was brought to the context of high energy proton-proton (pp) collisions

by D. Kharzeev in [6], where he suggested that the BJ would be responsible for a new mechanism of baryon stopping. In the standard picture of leading baryon formation [7], the valence quarks recombine after the collision, forming a baryon at large rapidities. In contrast, in [6] it was conjectured that the baryon junction breaking could lead to a low rapidity (“stopped”) baryon plus three mesons with large rapidities. The idea was implemented in analytical models [8] and also in the HIJING/B Monte-Carlo event generator [9]. The predictions of these models were in agreement with the available experimental data [10–14].

Later the BJ was used in [15] as initial condition in the study of coherent and incoherent diffractive production of J/ψ in electron-proton scattering with the IP-Sat and IP-Glasma models. The authors compared the results obtained with the BJ (which they called “stringy” proton) with the results obtained with three constituent quarks of finite size as initial condition. The conclusion was that, even though using the BJ initial condition leads to good agreement with data, the precise shape of the proton could not be constrained by incoherent diffractive J/ψ production. They also concluded that the data were well reproduced when they considered event-by-event geometric fluctuations, as well as the intrinsic saturation scale fluctuations as done in [16].

Another evidence of the baryon-junction was reported in [17], where the authors found that the BJ is able to

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP³.

explain the accelerated growth of the J/ψ yield observed in high multiplicity events in pp collisions.

All these phenomenological successes are not enough to unambiguously prove the existence of the baryon junction because all the measurements quoted above could still be explained with other models. Besides, one could argue that, increasing the resolution scale (Q^2) or boosting the nucleon to higher energies (decreasing x) the simple geometrical picture implied by the BJ is blurred by the profusion of produced partons during the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi and/or Balitsky-Fadin-Kuraev-Lipatov evolution.

Fortunately, there are good reasons to believe that the Y-shape configuration is not washed out and manifests itself even in high energy collisions. Indeed, there are indications that the geometric organization of matter persists even at LHC energies. In [18] it was shown that exclusive vector meson production is sensitive to the geometric deformation of the target. The authors concluded that the nuclear geometry (and its deformations) is not washed out by small- x evolution and hence the deformations previously inferred from low-energy experiments will be visible in high-energy collisions. This suggests the Y shape of the proton could survive the quantum evolution and manifest itself in high energy pp collisions. In view of the uncertainties involving this issue, more calculations and more data are needed. Along this direction, there is an interesting proposal to look for the BJ at the forthcoming electron-ion collider [19].

B. The baryon junction and multiplicity distributions

In this work we look for manifestations of the baryon junction in multiplicity distributions of charged particles measured at the LHC. Multiplicity is a global observable that allows us to characterize events in all colliding systems and has been studied as a tool to understand multiparticle production processes. From the experimental point of view, charged-particle multiplicity is one of the simplest observables and it allows us to gain insights on the production mechanisms [20,21].

In high energy proton (nucleus)-proton (nucleus) collisions, particles are produced in different ways. In an early stage of the collision there is a perturbative parton cascade process which is governed by the evolution equations of QCD. During a pp collision partons are produced in inelastic parton-parton collisions and later the partons are converted into hadrons with additional particle production. Here the main mechanism is nonperturbative: string formation and decays.

Fluctuations in the number of produced particles, which show up on the measured multiplicity distribution, can have several causes. One of them is certainly geometrical, i.e., it comes from the variation of the distribution of the colliding matter in space from event to event. Particle production and especially the multiplicity distribution is very sensitive to initial spatial distribution of matter and hence is a good place to test the baryon junction conjecture.

From the point of view of modeling, the initial state geometry can be naturally implemented in the partonic version of the Glauber model [22–24]. Typically one chooses the initial spatial density of matter and, with it, computes the thickness function, which appears in the expressions of the cross sections. In this work we will simulate the collisions in the MC-Kharzeev-Levin-Nardi (KLN) event generator [25]. This Monte Carlo simulator implements the k_T -factorization formalism of the color glass condensate (CGC) using the KLN unintegrated gluon distribution [26,27]. For a nucleon with a given initial spatial configuration, this simulator allows one to compute gluon production. Conversions to final-state hadrons are taken into account via the parton-hadron duality, meaning that the particle production at parton and hadron level only differ by a constant factor, also employed in previous phenomenological studies involving p_T -integrated observables calculated via the k_T -factorization formalism [25–31]. With these tools we can use the baryon junction initial conditions and calculate the final multiplicity distribution, looking for its manifestations.

The effect of the initial state geometry and its fluctuations on multiplicity distributions in pp collisions was investigated in [16]. In that work the authors studied multiplicity distributions in the dense-dense limit of the CGC, exploring the nature of high multiplicity events in pp collisions. They considered the standard fluctuations of the collision geometry together with color charge fluctuations and could obtain a reasonable description of data, excluding the highest multiplicity events. Additional sources of saturation scale fluctuations were required to explain the rare events which give rise to the tails of the multiplicity distribution in pp collisions. These fluctuations are a consequence of stochastic splitting of dipoles that are not accounted for in the conventional frameworks of CGC [32–35], and are described via a log-normal distribution. The width of the Gaussian distribution σ , as constrained by the data, appears to be energy independent, suggesting that these fluctuations have a non-perturbative origin. In [16] the authors did not address the specific case of the baryon junction. More recently, the effect of BJ initial conditions in multiplicity distributions in pp collisions was studied in [36]. The authors found that the BJ initial conditions provide a reasonable description of the data, but they were not able to account for the very large multiplicity events.

In this work we use the MC-KLN simulator implementing BJ initial configurations, as in [15], with intrinsic fluctuations, as in [16], and apply it to the multiplicity distributions, as in [36]. Here we focus on pp and pPb collisions, the second not addressed in any of these studies, and compare our results to very recent data [37].

II. FORMALISM

In this section we present the ingredients of our calculation. Our strategy is to use a well tested particle

production model and vary only the initial conditions, covering the most accepted matter spatial distributions, and try to determine which of them gives a better description of data. Ideally, this procedure would rule out some initial conditions and favor others. However, given the complexity of the calculation, it is difficult to exclude configurations. Nevertheless, the results give us hints and useful insights.

A. Initial conditions for the nucleon

We start by introducing four types of nucleon matter spatial distributions used in our simulations: hard-sphere, Gaussian, analytical baryon junction (from now on called BJ1) and a slightly different and purely numerical baryon junction (from now on called BJ2). Examples of the thickness function (z-integrated density) of each configuration are shown on the left panels of Figs. 1 and 2. The configurations are described below.

1. Hard-sphere

The hard-sphere is based on the collision criterion of distances between two nucleon centers [29]. Explicitly, there is a collision when the distance between the projectile (proj) and the target (targ) is

$$d^2 = (x_{\text{proj}} - x_{\text{targ}})^2 + (y_{\text{proj}} - y_{\text{targ}})^2 \leq \frac{\sigma_0}{\pi}, \quad (1)$$

where σ_0 is the transverse area of the large x partons of a nucleon. As in Ref. [29], we choose the parameter σ_0 to be equal to 42 mb. This collision criterion implies that each nucleon is geometrically described by a filled disk, with thickness given by

$$T(\mathbf{r}_\perp) = \frac{\Theta(\sqrt{\sigma_0/\pi} - r_\perp)}{\sigma_0}, \quad (2)$$

where $\Theta(x)$ is the usual Heaviside step function.

2. Gaussian nucleon

A way to soften the sharp edges of the hard-sphere model while still not including subnucleonic degrees of freedom and also maintaining the overall smooth shape of the proton is to consider a Gaussian nucleon. In this case, the nucleon transverse profile can be calculated analytically and is given by

$$T_N(\mathbf{r}_\perp) = \frac{e^{-\frac{r_\perp^2}{2B_p}}}{2\pi B_p}, \quad (3)$$

with B_p being the nucleon width, which is related to the root mean square radius of the proton as $r_p = \sqrt{2B_p}$. Older fits to HERA data [38], as well as a recent Bayesian analysis based solely on information from $e + p$ collisions

from HERA [39], show that $B_p \approx 4 \text{ GeV}^{-2}$, while a recent Bayesian analysis of combined data on diffractive J/Ψ photoproduction in $\gamma + p$ and $\gamma + Pb$ collisions yields a smaller value [40].

The probability of having a collision at nucleon level between a projectile and a target separated by an impact parameter b is now related to their overlap function, $T_{NN'}$, which can be calculated analytically in this model,

$$\begin{aligned} P(b) &= 1 - \exp(-\sigma_{\text{eff}}^{\text{eff}} T_{NN'}(b)), \\ T_{NN'}(b) &= \int T_N(x - b/2, y) T_{N'}(x + b/2, y) dx dy \\ &= \frac{e^{-b^2/4B_p}}{4\pi B_p}, \end{aligned} \quad (4)$$

where σ_{eff} is fixed so that the integral of $P(b)$ over the impact parameter yields the inelastic nucleon-nucleon cross section at a given energy. The thickness and overlap functions from Eqs. (3) and (4) enter as inputs in the MC-KLN.

3. Baryon junction BJ1

The BJ1 is an analytical Ansatz for the characteristic “Y” shape of the proton. The three effective quarks are disposed in the vertices of an equilateral triangle and they have coordinates $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$. To these quarks we add three gluon distributions with peaks at half of the distances between the nucleon center and the quark peaks. The resulting thickness is explicitly given by

$$T_N(x, y) = T_N^q(x, y) + T_N^g(x, y), \quad (5)$$

with the quark and gluon hot-spots, respectively, given by

$$\begin{aligned} T_N^q(x, y) &= \frac{(1-\kappa)}{6\pi B_h^2} \sum_{i=1}^3 \exp \\ &\times \left(-\frac{(x - (x_p + x_i))^2 + (y - (y_p + y_i))^2}{2B_h^2} \right), \end{aligned} \quad (6)$$

and

$$\begin{aligned} T_N^g(x, y) &= \frac{\kappa}{6\pi B_h^2} \sum_{i=1}^3 \exp \\ &\times \left(-\frac{(x - (x_p + \frac{x_i}{2}))^2 + (y - (y_p + \frac{y_i}{2}))^2}{2B_h^2} \right), \end{aligned} \quad (7)$$

In the formulas above, $\mathbf{r}_p = (x_p, y_p)$ is the coordinate of the center of the nucleon in the transverse plane, $\kappa = 0.5$ is the gluon fraction and $B_h = 0.17 \text{ fm}$ is the hot-spot size. All these parameters were fixed in previous works [15,17,36,41,42].

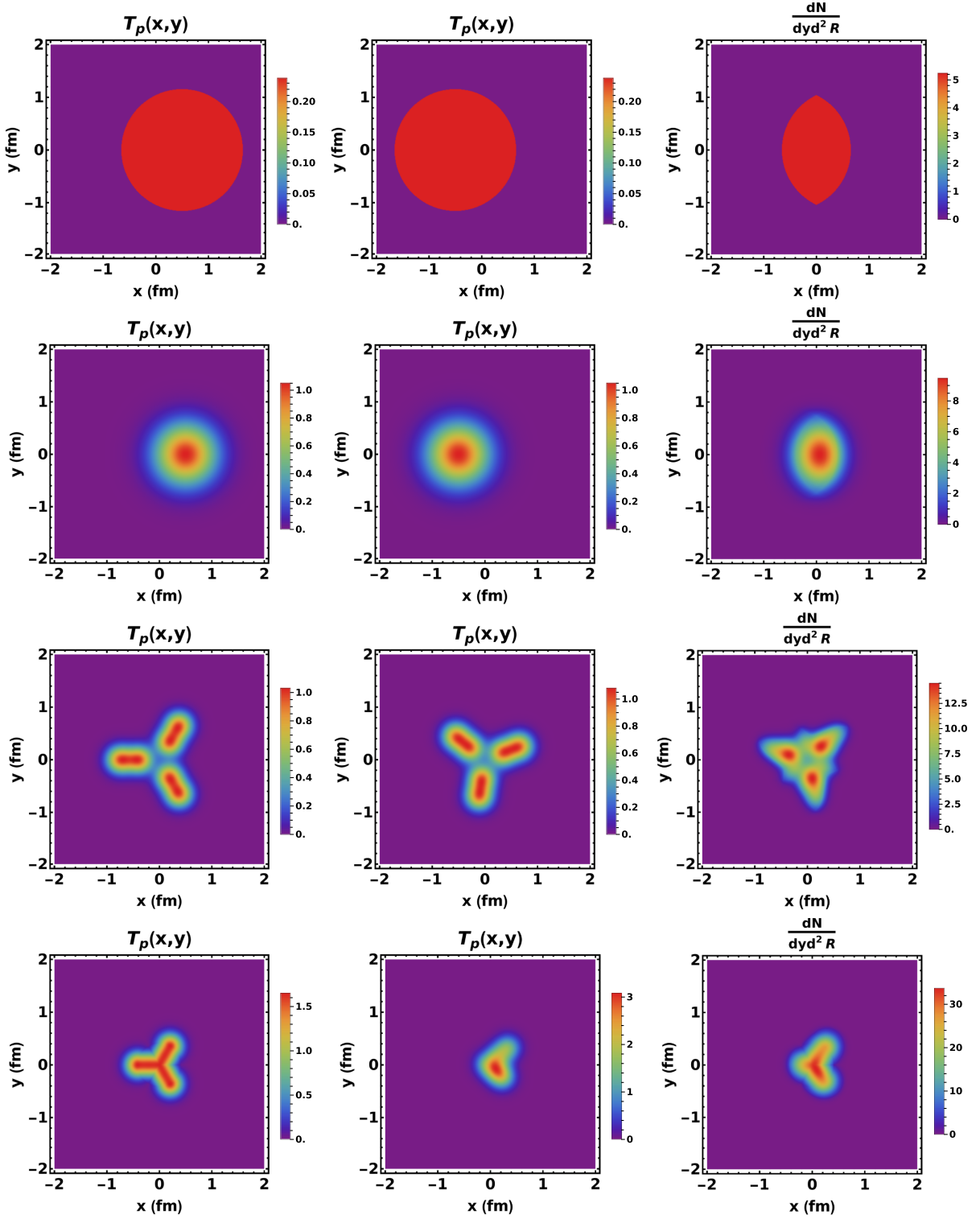


FIG. 1. Proton-proton collisions. Left: projectile proton thickness. Middle: target proton thickness. Right: $dN/dy d^2r$ of the produced gluons at midrapidity. The rows represent different density distributions of the nucleons. From top to bottom: hard-sphere, Gaussian, BJ1, BJ2.

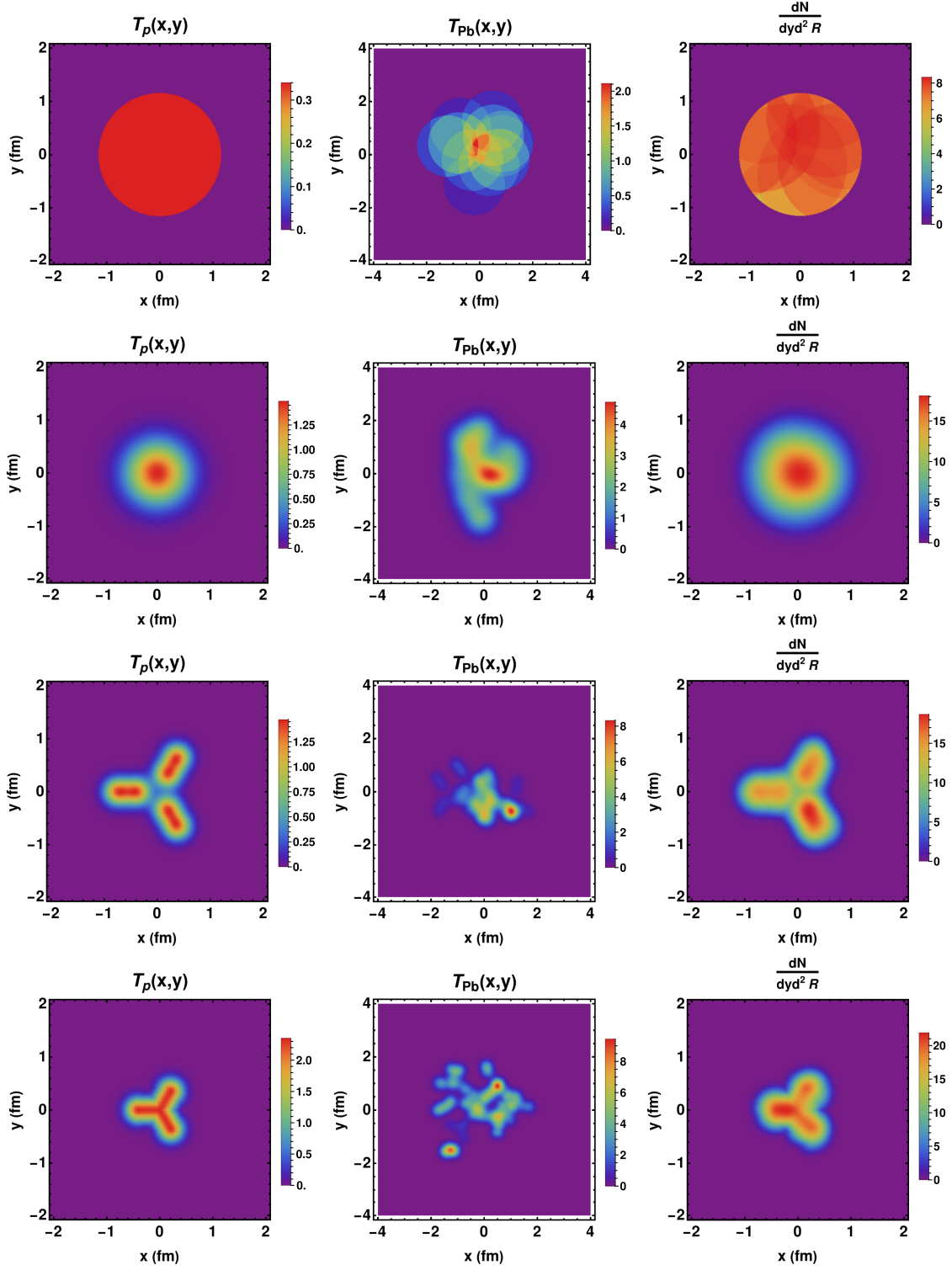


FIG. 2. Proton-lead collisions. Left: proton thickness. Middle: lead participant thickness. Right: $dN/dy d^2\mathbf{r}$ of the produced gluons at midrapidity. The rows represent different density distributions of the nucleons. From top to bottom: hard-sphere, Gaussian, BJ1, BJ2.

Each quark is positioned according to

$$\mathbf{r}_i = \frac{d}{2} (\cos(\phi_i + \alpha) \sin(\theta), \sin(\phi_i + \alpha) \sin(\theta), \cos(\theta)), \quad (8)$$

with $\phi_1 = \pi/3$, $\phi_2 = -\pi/3$, $\phi_3 = -\pi$ and $d = 1.5$ fm. The α parameter is randomly chosen in the range $[0, 2\pi]$, giving rotational freedom to the entire structure in the transverse plane, and the azimuthal θ is randomly chosen in

the range $[0, \pi]$. The respective overlap is calculated as

$$T_{NN'}(b) = \int T_N(x - b/2, y) T_{N'}(x + b/2, y) dx dy \\ = T_{NN'}^{qq}(b) + T_{NN'}^{gg}(b) + T_{NN'}^{qg}(b) + T_{NN'}^{gq}(b), \quad (9)$$

and hence it can be split into quark-quark, gluon-gluon, gluon-quark, and quark-gluon overlaps. Explicitly, each term is given by

$$T_{NN'}^{qq}(b) = \int T_N^q(x - b/2, y) T_{N'}^q(x + b/2, y) dx dy \\ = \frac{(1 - \kappa)^2}{36\pi B_h^2} \sum_{i=1}^3 \sum_{j=1}^3 \exp\left(-\frac{(b + (x_p + x_i) - (x_a + x_j))^2 + ((y_p + y_i) - (y_a + y_j))^2}{4B_h^2}\right), \quad (10)$$

$$T_{NN'}^{gg}(b) = \int T_N^g(x - b/2, y) T_{N'}^g(x + b/2, y) dx dy \\ = \frac{\kappa^2}{36\pi B_h^2} \sum_{i=1}^3 \sum_{j=1}^3 \exp\left(-\frac{(b + (x_p + x_i/2) - (x_a + x_j/2))^2 + ((y_p + y_i/2) - (y_a + y_j/2))^2}{4B_h^2}\right), \quad (11)$$

$$T_{NN'}^{qg}(b) = \int T_N^q(x - b/2, y) T_{N'}^g(x + b/2, y) dx dy \\ = \frac{\kappa(1 - \kappa)}{36\pi B_h^2} \sum_{i=1}^3 \sum_{j=1}^3 \exp\left(-\frac{(b + (x_p + x_i/2) - (x_a + x_j))^2 + ((y_p + y_i/2) - (y_a + y_j))^2}{8B_h^2}\right), \quad (12)$$

$$T_{NN'}^{gq}(b) = \int T_N^g(x - b/2, y) T_{N'}^q(x + b/2, y) dx dy \\ = \frac{\kappa(1 - \kappa)}{36\pi B_h^2} \sum_{i=1}^3 \sum_{j=1}^3 \exp\left(-\frac{(b + (x_p + x_i) - (x_a + x_j/2))^2 + ((y_p + y_i) - (y_a + y_j/2))^2}{8B_h^2}\right). \quad (13)$$

In the formulas above we introduced the position of the center of the target in the transverse plane, $\mathbf{r}_a = (x_a, y_a)$, and the peaks of the quark distribution in the target nucleon, $j = 1, 2, 3$, as in Eq. (8). We use the analytical thickness [Eq. (5)] and overlap [Eq. (9)] as inputs in the MC-KLN.

4. Baryon junction BJ2

The baryon junction BJ2 is a numerical Ansatz for the Y-like structure, as employed in [15]. In this approach we draw the position of the three quarks in the 3D space according to the distribution

$$P(\mathbf{r}) = \frac{e^{-\frac{r^2}{2B_r}}}{(2\pi B_r)^{3/2}}, \quad (14)$$

with $B_r = 0.41^2$ fm² as the characteristic size of the proton. After that we calculate the central point (Fermat point) as

the point that simultaneously minimizes the distance to the three quarks. In the last step the gluon flux tubes are filled by Gaussians with characteristic squared radius $B_l = 0.15^2$ fm² and the distribution is normalized to 1. The parameters were taken from [15] and the overlap of the string junction will be approximated by the overlap of the Gaussian nucleon of Eq. (4), but with constant B_r .

B. Initial condition for the nucleus

The initial condition for the lead nucleus will be defined by randomly positioning 208 nucleons inside the nucleus according to the Woods-Saxon profile

$$\rho(r) = \rho_0 \frac{1 + w(r/R)^2}{1 + \exp(\frac{r-R}{a})}, \quad (15)$$

where ρ_0 is the nucleon density in the center of the nucleus, R is the radius of the nucleus, w represents deviations from

the sphere format, and a is the skin depth. Some usual values for these parameters can be found in [43]. We perform our simulation by sorting 208 nucleon center positions, $\mathbf{r}_n$, and then inserting the chosen nucleon distribution around it. In the end, the Pb thickness is

$$T_{Pb}(x, y) = \sum_{n=1}^{208} T_n(x, y), \quad (16)$$

with T_n as the thickness of the chosen type of nucleon. Examples of the (participant) thickness for the lead nucleus are shown in the second column of Fig. 2.

C. k_T -factorization

In order to compute the rapidity and transverse momentum distributions of the particles produced in the final state, we use the k_T -factorization. For details we refer the reader to Refs. [44–46].

The single-inclusive (small- x) gluon production cross section is given by

$$\begin{aligned} \frac{d\sigma^{A+B \rightarrow g}}{dy d^2 p_T d^2 \mathbf{r}_\perp} &= K \frac{2}{C_F p_T^2} \int^{p_T} \frac{d^2 k_T}{4} \int d^2 b \alpha_s(Q) \phi_p \\ &\times \left(\frac{|p_T + k_T|}{2}, x_1; b \right) \phi_T \\ &\times \left(\frac{|p_T - k_T|}{2}, x_2; R - b \right). \end{aligned} \quad (17)$$

In our notation K is a multiplicative constant introduced to account for higher-order corrections. In practical terms, it will be fixed by requiring that the mean number of charged particles is reproduced. p_T is the transverse momentum of the produced gluon, k_T is the intrinsic transverse momentum and $C_F = (N_c^2 - 1)/2N_c$ is the color factor. The $2 \rightarrow 1$ kinematics implies that

$$x_{1,2} = \frac{p_T}{\sqrt{s}} e^{\pm y}. \quad (18)$$

The unintegrated gluon distribution function (UGD) gives the number of gluons per x , per transverse momentum unit. In this work we use the KLN UGD, which is given by

$$\phi_{\text{KLN}}(\mathbf{k}, x) = \frac{2C_F(1-x)^4}{3\pi^2 \alpha_s} \begin{cases} 1, & \text{if } k \leq Q_s \\ (Q_s/k)^2, & \text{if } k > Q_s \end{cases} \quad (19)$$

The UGD is evaluated at a saturation scale given by [47]

$$Q_s^2(x, \mathbf{r}_\perp) = T(\mathbf{r}_\perp) \frac{2 \text{ GeV}^2}{1.53 \text{ fm}^{-2}} \left(\frac{0.01}{x} \right)^{\bar{\lambda}}, \quad (20)$$

where $T(\mathbf{r}_\perp)$ is the projectile or target thickness and $\bar{\lambda} = 0.23$.

The MC-KLN convolutes the initial conditions of p and Pb with the k_T -factorization formula of Eq. (17) by calculating $dN/dy d^2 \mathbf{r}_\perp$ of gluons at a given rapidity in the overlap region. One example of each initial state geometry for pPb collisions is shown in Fig. 2, for the same seed. On the left column we present the proton thickness, on the middle column we show the thickness of the lead nucleus and on the right column we show the convolution of the initial conditions and Eq. (17) at midrapidity.

In Fig. 2 we can observe that, remarkably, the produced gluon distribution on the right column inherits the proton shape, which makes pPb collisions good systems to probe the proton initial geometry. This happens because the proton size, which is about 1 fm, is much smaller than the lead radius (6.62 fm). Thus, the overlap region will invariably present the proton shape. On the other hand, in pp collisions, since the projectile and target are comparable in size, only fragments of the proton structure will overlap. More specifically, as represented in Fig. 1, for hard-sphere and Gaussian initial conditions the convolution will only inherit the proton structure for $b = 0$ events, and for the baryon junctions, only for specific angle combinations. Since these conditions are rare, in general this system become less sensitive to the initial proton geometry than pPb .

Finally, in order to calculate the final charged particle distribution, we assume parton-hadron duality. This means that the number of detected hadrons is assumed to be proportional to the number of produced gluons. We also use the $y \rightarrow \eta$ Jacobian

$$\frac{\cosh \eta}{\sqrt{\cosh^2 \eta + m_\pi^2/k^2}} \quad (21)$$

to obtain the pseudorapidity distribution. In the above expression $m_\pi \approx 140$ MeV is the pion mass. We integrate the distribution (17) in $d^2 p_T$ and in the transverse plane to obtain the particle production at a given (pseudo) rapidity value.

D. Intrinsic fluctuations

So far we have considered only geometrical fluctuations, i.e., those arising from changing positions of the initial quarks and gluons in the transverse plane. In order to describe high-multiplicity events, we must include dynamical fluctuations in the saturation scale [16] of a given nucleon in the collision. These fluctuations are characterized by deviations of the saturation scale from its mean value and they follow the distribution

$$P(\ln(Q^2/\langle Q^2 \rangle)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\ln^2(Q(\mathbf{s}_\perp)^2/\langle Q(\mathbf{s}_\perp)^2 \rangle)}{2\sigma^2}\right), \quad (22)$$

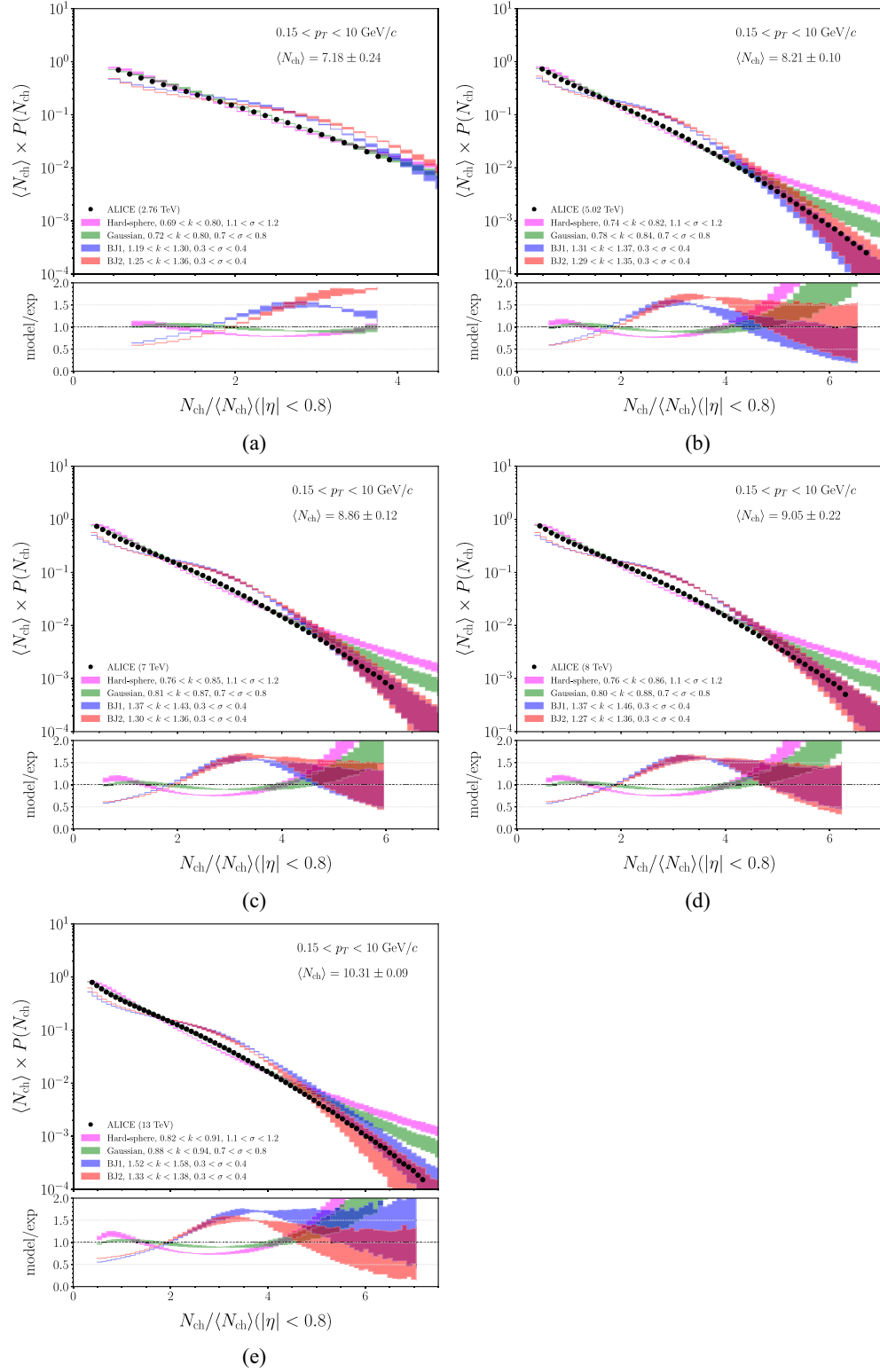


FIG. 3. KNO plot for pp multiplicity distribution at (a) $\sqrt{s} = 2.76$ TeV, (b) $\sqrt{s} = 5.02$ TeV, (c) $\sqrt{s} = 7$ TeV, (d) $\sqrt{s} = 8$ TeV and (e) $\sqrt{s} = 13$ TeV. The four different lines correspond to the four different initial geometries. The experimental data (black points) are from [37].

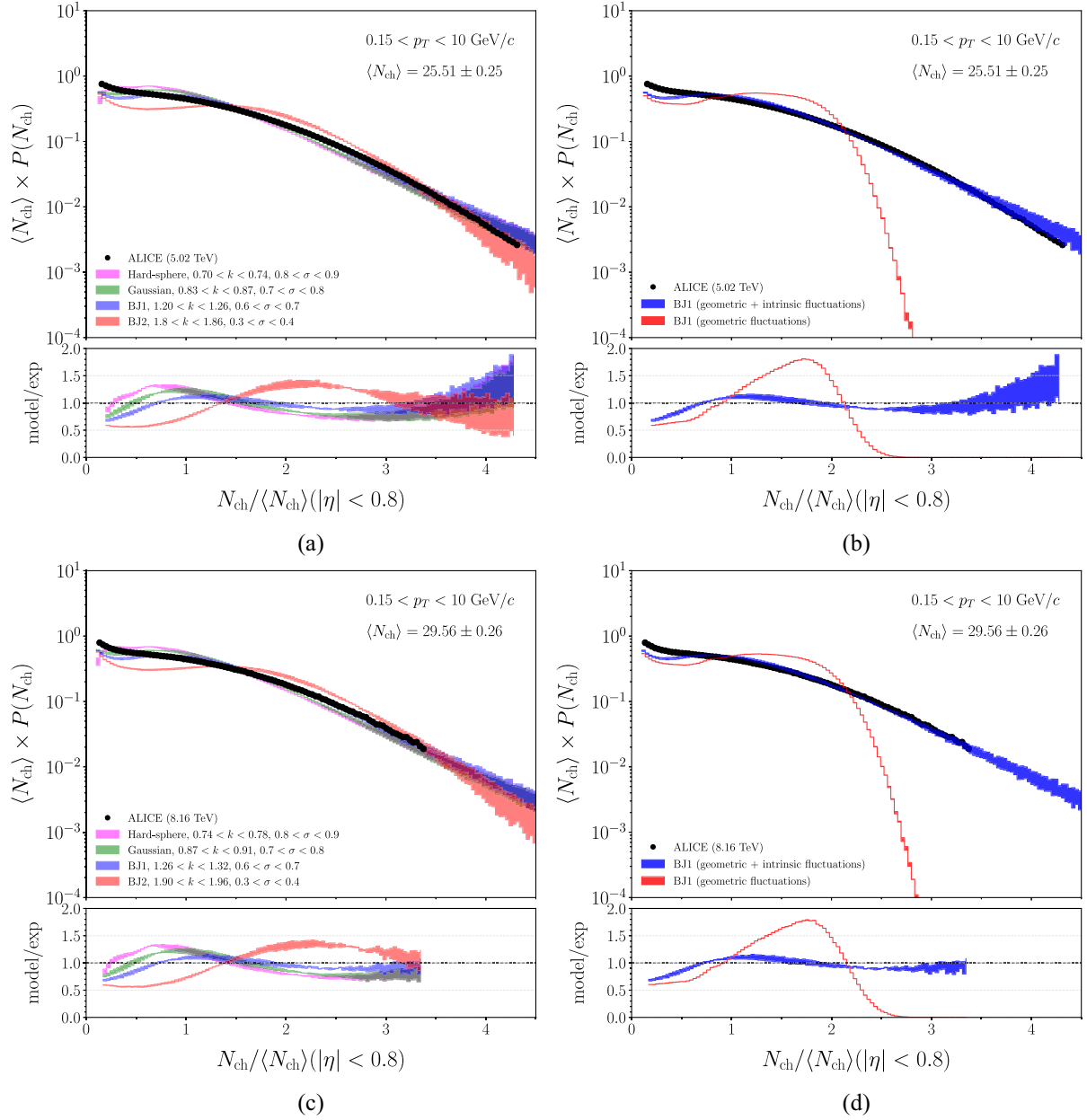


FIG. 4. KNO plot for p Pb multiplicity distribution at (a) $\sqrt{s} = 5.02$ TeV and (c) $\sqrt{s} = 8.16$ TeV. The four different lines correspond to the four different initial geometries. Comparison of BJ2 with and without intrinsic fluctuations at (b) $\sqrt{s} = 5.02$ TeV and (d) $\sqrt{s} = 8.16$ TeV. The experimental data (black points) are from [37].

with σ as a free parameter to be adjusted. These fluctuations are such that large deviations from the averaged value are rare, which implies that high-multiplicity events are rare too.

III. RESULTS

We will now address the recent ALICE data [37] on charged particle multiplicity distributions measured in pp collisions at $\sqrt{s} = 2.76, 5.02, 7, 8, 13$ TeV and p Pb collisions at $\sqrt{s} = 5.02, 8.16$ TeV, in the range $0.15 < p_T < 10$ GeV/c and $-0.8 < \eta < 0.8$.

Integrating Eq. (17) in p_T , in y (or η) and in $r_\perp$ in an event-by-event basis, we construct the probability $P(N_{\text{ch}})$ of producing N_{ch} charged particles. The results for pp and p Pb collisions are shown in Figs. 3 and 4, respectively. In the figures K is the multiplicative constant of Eq. (17). For fixed values of K and σ we integrate the k_T factorization formula and obtain values of N_{ch} . Accepting all the values of K and σ which yield values of N_{ch} which are consistent with the experimental values (within the errors) and which are quoted in the figures, leads to sets of values which define the areas shown in the figures.

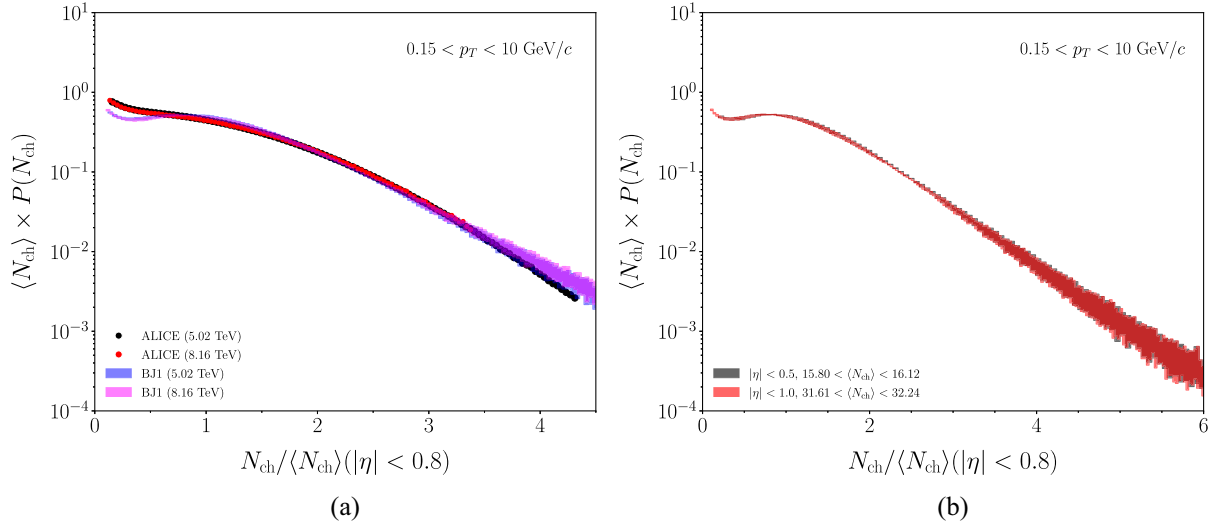


FIG. 5. (a) KNO plot for $\sqrt{s} = 5.02$ and 8.16 TeV. The experimental data are from [37]; (b) Predictions for p Pb multiplicity distributions at $\sqrt{s} = 5.02$ TeV for two rapidity intervals $|\eta| < 0.5$ and $|\eta| < 1.0$ presented in the KNO form.

The k_T factorization formalism is appropriate to describe inelastic nondiffractive events which yield any number of produced particles above a certain value, which we choose to be $N_{\text{ch}} = 3$. In terms of the KNO variable, this means something like $N_{\text{ch}}/\langle N_{\text{ch}} \rangle \geq 0.5$. Another limitation of our formalism is that we did not include the contribution from the valence quarks. They are supposed to be important only in the large rapidity regions, where one can use the hybrid formalism [48]. At the low rapidities considered here their contribution should be less important but still needed for a better agreement with data.

Results for $P(N_{\text{ch}})$ are often presented in terms of the so called Koba-Nielsen-Olsen (KNO) scaling [49] variables. KNO scaling means that the multiplicity distribution, $P(N_{\text{ch}})$, becomes energy independent at asymptotically high energies when we multiply it by the mean value $\langle N_{\text{ch}} \rangle$ and plot it as a function of the scaling variable $z = N_{\text{ch}}/\langle N_{\text{ch}} \rangle$. This feature has been experimentally observed at the LHC for narrow pseudorapidity ($|\eta| < 0.5$) intervals [50]; at the same time, the violation of KNO scaling has been reported for larger pseudorapidity intervals.

In order to reconcile the CGC formalism with KNO scaling, it has been shown that the theory must be nearly Gaussian (corresponding to the strong color field limit) and running coupling effects must be included as well [51]. Such result was only verified for pp collisions up to ~ 3.5 times the average multiplicity. Small violations of the KNO scaling have been reported for p Pb collisions in the region where $N_{\text{ch}} \lesssim 3$, where additional sources of fluctuations (geometric) are present [31]. Here, we extend the investigation to larger multiplicities. It is worth noting that KNO scaling might need to be modified for very large multiplicity events [52]. In spite of all these efforts, a clear connection between KNO scaling and QCD is not yet

established. For complementary references on this subject we refer the reader to [53–55].

The first measurements of multiplicity distribution in p Pb collisions [37] were taken at two different energies. It is possible to make a first check of the validity of the KNO, which, as can be seen in Fig. 5(a), seems to be valid. In that figure we see that both experimental data and the MC-KLN model with BJ1 initial geometry exhibit KNO scaling.

Using the BJ1 initial condition we estimate the multiplicity distributions for different pseudorapidity bins. In Fig. 5(b) we show our predictions for the pseudorapidity ranges $|\eta| < 0.5$ and $|\eta| < 1.0$. We also provide the mean value estimate and observe that KNO scaling is still valid for these pseudorapidity intervals.

Looking at Fig. 3 we observe that data from pp collisions are better described by Gaussian initial conditions, except for the region $z = N_{\text{ch}}/\langle N_{\text{ch}} \rangle \geq 4$, where BJ1 and BJ2 initial conditions show a better agreement with the high multiplicity data. From Fig. 4 we see that p Pb data are better described by BJ1. A remarkable feature is that the results from BJ1 and BJ2 are close to each other (within uncertainties) in pp collisions but are quite different in p Pb collisions. In the low multiplicity region almost all curves slightly underpredict the data. In Figs. 4(b) and 4(d) we see the crucial role played by intrinsic fluctuations at large values of N_{ch} .

To summarize, we would like to make the following remarks: first, from Figs. 1 and 2 we conclude that p Pb collisions are better places to study the initial proton geometry than pp collisions; second, our model can be used to calculate the measured multiplicity distributions and the results are sensitive to the initial state inputs. Further constraining the model parameters and reducing the theoretical error, the results would be different enough to be discriminated by the data, which would “prefer” one initial

condition. For the moment, in p Pb and ultracentral pp collisions ($z \geq 4$) the data favor the baryon junction models. Moreover, in p Pb collisions it is possible to discriminate between different baryon junction models and BJ1 is clearly favored. This is a consequence of the high sensitivity of p Pb to the proton initial condition.

IV. CONCLUSION

This work is part of a broader project of looking for experimental manifestations of the baryon junction configuration of the proton. Here we investigated proton-proton and proton-nucleus collisions at the LHC. More specifically, we calculated the multiplicity distributions measured in pp and p Pb collisions. We used different initial state geometries as input in the MC-KLN Monte Carlo event generator which implements the k_T -factorization formalism of the CGC with KLN unintegrated gluon distributions. We found that in pp collisions the Gaussian spatial configuration is favored, except in the region $z \geq 4$, where BJ1 and BJ2 give a better description of the data. In p Pb collisions the data are better explained by BJ1 initial conditions, provided that fluctuations in the saturation scale (also called “intrinsic fluctuations”) are included.

The existence of the baryon junction still needs confirmation. Further studies will be undertaken at the electron-ion collider.

ACKNOWLEDGMENTS

We are grateful to M. Munhoz, H. Martins-Fontes, Y. Lima and M. V. T. Machado for useful discussions. This study was financed, in part, by the São Paulo Research Foundation (FAPESP), Brasil. Process Number 2024/06652-4. This work was financed by the Brazilian funding agencies CNPq, FAPESP, CAPES, FAPERJ, and by the INCT-FNA. This work has been done as a part of the Project INCT-Física Nuclear e Aplicações, Projeto No. 464898/2014-5. The authors acknowledge the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil), through the ambassador program (UFGD), subprojects FCNAE and SADFT for providing HPC resources of the SDumont supercomputer.

DATA AVAILABILITY

The data are not publicly available. The data are available from the authors upon reasonable request.

-
- [1] G. C. Rossi and G. Veneziano, *Phys. Rep.* **63**, 153 (1980).
 - [2] T. T. Takahashi, H. Matsufuru, Y. Nemoto, and H. Suganuma, *Phys. Rev. Lett.* **86**, 18 (2001).
 - [3] H. Suganuma, T. T. Takahashi, F. Okiharu, and H. Ichie, *AIP Conf. Proc.* **756**, 123 (2005).
 - [4] F. Bissey, F. G. Cao, A. R. Kitson, A. I. Signal, D. B. Leinweber, B. G. Lasscock, and A. G. Williams, *Phys. Rev. D* **76**, 114512 (2007).
 - [5] N. Sakumichi and H. Suganuma, *Phys. Rev. D* **92**, 034511 (2015).
 - [6] D. Kharzeev, *Phys. Lett. B* **378**, 238 (1996).
 - [7] G. N. Fowler, F. S. Navarra, M. Plumer, A. Vourdas, R. M. Weiner, and G. Wilk, *Phys. Rev. C* **40**, 1219 (1989).
 - [8] F. W. Bopp and M. Shabelski, *Eur. Phys. J. A* **28**, 237 (2006).
 - [9] V. Topor Pop, M. Gyulassy, J. Barrette, C. Gale, X. N. Wang, and N. Xu, *Phys. Rev. C* **70**, 064906 (2004).
 - [10] I. G. Bearden *et al.* (BRAHMS Collaboration), *Phys. Rev. Lett.* **93**, 102301 (2004).
 - [11] B. I. Abelev *et al.* (STAR Collaboration), *Phys. Rev. C* **79**, 034909 (2009).
 - [12] L. Adamczyk *et al.* (STAR Collaboration), *Phys. Rev. C* **96**, 044904 (2017).
 - [13] K. Aamodt *et al.* (ALICE Collaboration), *Phys. Rev. Lett.* **105**, 072002 (2010).
 - [14] E. Abbas *et al.* (ALICE Collaboration), *Eur. Phys. J. C* **73**, 2496 (2013).
 - [15] H. Mäntysaari and B. Schenke, *Phys. Rev. D* **94**, 034042 (2016).
 - [16] L. McLerran and P. Tribedy, *Nucl. Phys. A* **945**, 216 (2016).
 - [17] R. Terra and F. S. Navarra, *Phys. Rev. D* **108**, 054002 (2023).
 - [18] H. Mäntysaari, B. Schenke, C. Shen, and W. Zhao, *Phys. Rev. Lett.* **131**, 062301 (2023).
 - [19] J. D. Brandenburg, N. Lewis, P. Tribedy, and Z. Xu, *Eur. Phys. J. C* **84**, 590 (2024).
 - [20] For a review see: W. Kittel and E. A. De Wolf, *Soft Multihadron Dynamics* (World Scientific, Singapore, 2005).
 - [21] J. F. Grosse-Oetringhaus and K. Reygers, *J. Phys. G* **37**, 083001 (2010).
 - [22] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, *Annu. Rev. Nucl. Part. Sci.* **57**, 205 (2007).
 - [23] C. Loizides, *Phys. Rev. C* **94**, 024914 (2016).
 - [24] C. Loizides, J. Kamin, and D. d’Enterria, *Phys. Rev. C* **99**, 019901 (2019).
 - [25] A. Dumitru, A. V. Giannini, M. Luzum, and Y. Nara, *Phys. Lett. B* **784**, 417 (2018).
 - [26] D. Kharzeev, E. Levin, and M. Nardi, *Phys. Rev. C* **71**, 054903 (2005).
 - [27] D. Kharzeev, E. Levin, and M. Nardi, *Nucl. Phys. A* **747**, 609 (2005).
 - [28] A. Dumitru, D. E. Kharzeev, E. M. Levin, and Y. Nara, *Phys. Rev. C* **85**, 044920 (2012).

- [29] J. L. Albacete, A. Dumitru, H. Fujii, and Y. Nara, *Nucl. Phys.* **A897**, 1 (2013).
- [30] F. O. Durães, A. V. Giannini, V. P. Goncalves, and F. S. Navarra, *Phys. Rev. D* **94**, 054023 (2016).
- [31] A. Dumitru and Y. Nara, *Phys. Rev. C* **85**, 034907 (2012).
- [32] E. Iancu, A. H. Mueller, and S. Munier, *Phys. Lett. B* **606**, 342 (2005).
- [33] S. Munier and R. B. Peschanski, *Phys. Rev. D* **69**, 034008 (2004).
- [34] A. H. Mueller, *Nucl. Phys.* **B415**, 373 (1994).
- [35] E. Iancu and A. H. Mueller, *Nucl. Phys.* **A730**, 494 (2004).
- [36] S. Deb, G. Sarwar, D. Thakur, P. Subramani, R. Sahoo, and J. e. Alam, *Phys. Rev. D* **101**, 014004 (2020).
- [37] S. Acharya *et al.* (ALICE Collaboration), *Phys. Lett. B* **853**, 138700 (2024).
- [38] A. H. Rezaeian, M. Siddikov, M. Van de Klundert, and R. Venugopalan, *Phys. Rev. D* **87**, 034002 (2013).
- [39] H. Mäntysaari, B. Schenke, C. Shen, and W. Zhao, *Phys. Lett. B* **833**, 137348 (2022).
- [40] H. Mäntysaari, H. Roch, F. Salazar, B. Schenke, C. Shen, and W. Zhao, *Phys. Rev. D* **113**, 014038 (2026).
- [41] R. Terra, Charmonium Production in High Multiplicity pp and pPb Collisions and the Structure of the Proton, MSc thesis, University of Sao Paulo, 2024, [arXiv:2404.07743](https://arxiv.org/abs/2404.07743).
- [42] P. Kubiczek and S. D. Glazek, *Lith. J. Phys.* **55**, 155 (2015).
- [43] H. De Vries, C. W. De Jager, and C. De Vries, *Atom. Data Nucl. Data Tables* **36**, 495 (1987).
- [44] W. A. Horowitz and Y. V. Kovchegov, *Nucl. Phys.* **A849**, 72 (2011).
- [45] Y. V. Kovchegov and H. Weigert, *Nucl. Phys.* **A807**, 158 (2008).
- [46] J. L. Albacete, A. Dumitru, H. Fujii, and Y. Nara, *Nucl. Phys.* **A897**, 1 (2013).
- [47] J. L. Albacete, A. Dumitru, and Y. Nara, *J. Phys. Conf. Ser.* **316**, 012011 (2011).
- [48] A. Dumitru, A. Hayashigaki, and J. Jalilian-Marian, *Nucl. Phys.* **A765**, 464 (2006).
- [49] Z. Koba, H. B. Nielsen, and P. Olesen, *Nucl. Phys.* **B40**, 317 (1972).
- [50] V. Khachatryan *et al.* (CMS Collaboration), *J. High Energy Phys.* **01** (2011) 079.
- [51] A. Dumitru and E. Petreska, [arXiv:1209.4105](https://arxiv.org/abs/1209.4105).
- [52] L. S. Moriggi, F. S. Navarra, and M. V. T. Machado, *Phys. Rev. D* **112**, 074019 (2025);
- [53] H. R. Martins-Fontes and F. S. Navarra, *Phys. Rev. D* **112**, 094045 (2025); *MDPI Phys.* **7**, 57 (2025);
- [54] G. R. Germano, F. S. Navarra, G. Wilk, and Z. Włodarczyk, *Phys. Rev. D* **110**, 034026 (2024).
- [55] G. R. Germano and F. S. Navarra, *Phys. Rev. D* **105**, 014005 (2022).