Dipole evolution: perspectives for collectivity and gamma*A collisions

Bierlich, Christian; Rasmussen, Christine O.

Published in:
Journal of High Energy Physics

DOI:
10.1007/JHEP10(2019)026

Publication date:
2019

Published for SISSA by Springer

Received: July 31, 2019
Revised: September 3, 2019
Accepted: September 15, 2019
Published: October 4, 2019

Dipole evolution: perspectives for collectivity and γ^*A collisions

Christian Bierlich[a,b] **and Christine O. Rasmussen**[a]

[a] *Dept. of Astronomy and Theoretical Physics, Sölvegatan 14A, S-223 62 Lund, Sweden*
[b] *Niels Bohr Institute, Blegdamsvej 17, 2100 København Ø, Denmark*

E-mail: bierlich@thep.lu.se, christine.rasmussen@thep.lu.se

Abstract: The transverse, spatial structure of protons is an area revealing fundamental properties of matter, and provides key input for deeper understanding of emerging collective phenomena in high energy collisions of protons, as well as collisions of heavy ions. In this paper eccentricities and eccentricity fluctuations are predicted using the dipole formulation of BFKL evolution. Furthermore, first steps are taken towards generation of fully exclusive final states of γ^*A collisions, by assessing the importance of colour fluctuations in the initial state. Such steps are crucial for the preparation of event generators for a future electron-ion collider. Due to the connection between an impact parameter picture of the proton structure, and cross sections of ep and pp collisions, the model parameters can be fully determined by fits to such quantities, leaving results as real predictions of the model.

Keywords: QCD Phenomenology

ArXiv ePrint: 1907.12871

https://doi.org/10.1007/JHEP10(2019)026

JHEP10(2019)026

Contents

JHEP10(2019)026

1 Introduction

In the research program at the Large Hadron Collider (LHC) and the Relativistic Heavy Ion Collider (RHIC), collisions of ultra relativistic heavy ions are hypothesized to result in the creation of a quark-gluon plasma (QGP) with partonic degrees of freedom. One of the main avenues for investigating and characterizing this plasma consists of measurements of azimuthal correlations between particle pairs separated in rapidity, connecting particle emission angles to the initial geometry of the collision. Non-trivial correlations reflecting collective properties were first observed in gold-gold and copper-copper collisions at RHIC [1], but has since been investigated also in lead-lead (PbPb) collisions at the LHC [2–4]. Such non-trivial azimuthal correlations had at that point already been hypothesized to be a signal for hydrodynamic behaviour [5], or, even earlier, to involve microscopic dynamics of overlapping "quark tubes" or strings [6].

Similar results have been obtained in smaller collision systems such as proton-lead (pPb) [7], deuteron-gold [8], and, perhaps most surprisingly, in proton-proton (pp) [9]. Attempts to observe similar behaviour in even smaller collision systems, e.g. e^+e^-, has, while carrying interesting prospects, so far not produced positive results [10]. Even though the discovery of collectivity in pp is almost ten years old, the origin of such correlations in small collision systems is still highly debated (see ref. [11] for a recent review), and its resolution is among the top priorities for the future heavy ion program at LHC [12]. One possibility is that the correlations in these small collision systems are due to coherence effects [13] or initial state correlations [14]. Another is a repetition of the argument from heavy ion collisions, where the observed collective behaviour is a hydrodynamic response to the initial partonic spatial configuration [15]. A picture where a hydrodynamic "core" coexists with a non-hydrodynamic "corona" has been shown by the EPOS model [16, 17] to provide good descriptions of collectivity even in small collision systems.

The possibility of a hydrodynamic (or in fact any other) response to an initial geometric configuration of partons, poses a challenge to the traditional strategies for pp event generators, such as Pythia 8 [18] or Herwig 7 [19], both based on perturbative QCD (pQCD) with no obvious way to extract a spatial configuration for which to calculate a response. Attempts to calculate such a structure [20–22] generally involve assuming a certain spatial distribution of partons in the proton and, using the eikonal approximation, then transferring this structure to a spatio-temporal structure of the multiple partonic interactions (MPIs). The immediate drawbacks of such an approach are that (a) such models will in general contain parameters which need to be fitted to the same type of particle correlations as they wish to predict, and (b) assuming a spatial distribution of partons in a proton will generally contain many *ad hoc* elements.

Even though the spatial distribution of partons in a proton cannot be assessed *ab initio*, the evolution of said distribution can be calculated perturbatively in the formalism of Mueller [23, 24]. At high energies, average properties will retain little dependence on the initial configuration, i.e. be mostly dependent on the evolution. Since the transverse substructure of the colliding protons (or virtual photons) can be linked to total or semi-inclusive cross sections, any model parameters can be tuned to such quantities, and leave

JHEP10(2019)026

any further estimation of collective effects as real predictions of the model. Attempts to predict the elliptic flow in pp collisions using an implementation of Mueller's model was provided in 2011 [25], showing $v_{2,3}$ comparable to values found from PbPb at RHIC and LHC energies.

This paper is concerned with presenting a new Monte Carlo implementation of Mueller's model, study its description of cross sections in pp and γ^*p collisions, in order to provide estimates on parton level geometries in pp, proton-ion (pA) and ion-ion (AA) collisions, linked to collective phenomena. Mueller's model has been implemented as a Monte Carlo several times before, as it is not only useful for calculating spatial distributions of partons, but in fact has much wider applications due to the equivalence of the Mueller formalism with B-JIMWLK (Balitsky, Jalilian-Marian, Iancu, McLerran, Weigert, Leonidov and Kovner) [26–33] evolution (see section 2). Such an implementation makes direct introduction of effects beyond the leading logarithmic approximation possible, e.g. conservation of energy and momentum without imposing kinematical constraints on the splitting kernel [34]. This makes the implementation attractive for estimation of basic quantities dominated by small x processes (e.g. cross sections) in cases where little guidance from data exists. In this paper (see section 7) we will also apply the formalism to extract Glauber-Gribov (GG) colour fluctuations in γ^*p collisions, in order to take the initial steps towards a generation of electron-ion (eA) collisions within the Angantyr framework [35, 36] — a possibility which is foreseen to aid the preparation of an eA program currently being planned [37].

Earlier implementations of the Mueller dipole model include the public Oedipus [38] and Dipsy [39, 40] codes, as well as a private implementation by Kovalenko et al. [41]. All implementations treat only gluons in the evolution, as will this work. The implementation in this paper is similar to the implementation in Dipsy in some respects, but differs in other, while bearing less resemblance to the other two. The key differences between most of the used approaches, lies in the treatment of effects beyond leading order. Worth mentioning already here, is the treatment of sub-leading N_c (number of colours) effects in the evolution, leading to saturation in the cascade. In Dipsy, this is addressed through so-called swing mechanisms [42, 43], which suppresses the contribution from large dipoles in dense environments by replacing them with small dipoles. In this paper we consider only sub-leading N_c effects in the collision frame, by including multiple interactions in a way consistent with unitarity. Thus we make *no attempt* at treating saturation in the cascade, as the focus is rather to study how well one can do with an approach that includes only a minimal set of sub-leading corrections. Effects included in this paper is energy-momentum conservation and recoil effects (which are beyond leading log) and confinement (which is a non-perturbative effect). This also separates our approach from the IP-Glasma approach [44], which includes gluon saturation effects in the initial configuration explicitly, and evolve using B-JIMWLK.

On a more technical note, the approach presented in this paper is implemented within the larger framework of the Pythia 8 Monte Carlo event generator. This first of all means that the implementation will become publicly available,[1] and to aid reproducibility and

[1]From a future version of Pythia, larger than version 8.300, yet to be determined. See https://home.thep.lu.se/Pythia for up-to-date information.

JHEP10(2019)026

transparency, a large part of the manuscript, as well as appendix A, are devoted to the details of the implemented model. Our approach is simplistic in the sense that only a minimal amount of corrections to Mueller's original model has been added, and where ambiguities have arisen, the simplest possible choice has been taken.

The structure of the paper is as follows: After this introduction, the pQCD model of Mueller is introduced. Then follows a description on how observables are calculated within the Good-Walker framework as well as a definition of the observables related to the substructure of protons. The next section describes the overall features of the Monte Carlo implementation, before we proceed to the results on cross sections, eccentricities and colour fluctuations in processes with incoming virtual photons. Lastly, a section is devoted to conclusions and forthcoming work.

2 Proton substructure evolution

In this section we will outline the theoretical basis of the initial state evolution approach used in subsequent sections, and briefly review its relation to other approaches. The theoretical basis is the well known dipole QCD model by Mueller et al. [23, 24].

2.1 Dipole evolution in impact parameter space

We consider in general a picture with a projectile with a dipole structure incident on a target. In the simplest case, the projectile is just a single dipole r_{12}, spanned between the coordinates $\vec{r}_1$ and $\vec{r}_2$, in impact parameter space. The probability at leading order for this dipole to branch when evolved in rapidity (y), is

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}y} = \mathrm{d}^2\vec{r}_3 \, \frac{N_c\alpha_s}{2\pi^2}\frac{r_{12}^2}{r_{13}^2 r_{23}^2} \equiv \mathrm{d}^2\vec{r}_3 \, \kappa_3. \tag{2.1}$$

Here $\vec{r}_3$ is the transverse coordinate of the emitted gluon and κ_3 is used as a short-hand for the splitting kernel. An observable O known initially, will after an infinitesimal interval $\mathrm{d}y$ have the expectation value (denoted by a bar), assuming unitarity:

$$\bar{O}(y+\mathrm{d}y) = \mathrm{d}y \int \mathrm{d}^2\vec{r}_3 \, \kappa_3 \left[O(r_{13}) \otimes O(r_{23})\right] + O(r_{12}) \left[1 - \mathrm{d}y \int \mathrm{d}^2\vec{r}_3 \, \kappa_3\right], \tag{2.2}$$

where $\otimes$ denotes the evaluation of the observable O in the two dipole system r_{13}, r_{23}. In the limit $\mathrm{d}y \to 0$ this becomes:

$$\frac{\partial\bar{O}}{\partial y} = \int \mathrm{d}^2\vec{r}_3 \, \kappa_3 \left[O(r_{13}) \otimes O(r_{23}) - O(r_{12})\right]. \tag{2.3}$$

Remarkably, eq. (2.3) allows for the evolution of any observable calculable in impact parameter space. In the case of S-matrices in impact parameter space, the evaluation in the two dipole system reduces to a normal product in the eikonal approximation. Thus $O(r_{13}) \otimes O(r_{23}) \to S(r_{13})S(r_{23})$. Changing to scattering amplitudes, T, by substituting $T \equiv 1 - S$, one obtains:

$$\frac{\partial\langle\overline{T}\rangle}{\partial y} = \int \mathrm{d}^2\vec{r}_3 \, \kappa_3 \left[\langle T_{13}\rangle + \langle T_{23}\rangle - \langle T_{12}\rangle - \langle T_{13}T_{23}\rangle\right]. \tag{2.4}$$

JHEP10(2019)026

This is the B-JIMWLK equation hierarchy [26–33] in impact parameter space, which, as shown already by Mueller [45], can be generated directly from eq. (2.1). Equation (2.4) includes a non-linear term, $\langle T_{13}T_{23}\rangle$, and the treatment of this term is defining for many of the various approaches dealing with initial state evolution at low x.

Removal of the non-linear term yields the BFKL (Balitsky, Fadin, Kuraev and Lipatov) equation [46, 47], which correctly sums all of the leading logarithms in energy (or, more precisely in rapidity $(\alpha_s \cdot y)^n$) to all orders. Other than simply neglecting it, the simplest treatment of the non-linear term is by a mean-field approach, where it factorizes as: $\langle T_{13}T_{23}\rangle \to \langle T_{13}\rangle\langle T_{23}\rangle$. This approximation yields the BK (Balitsky and Kovchegov) equation [26, 48].

2.2 The Mueller dipole model

Several approaches have been proposed to utilize the simple, but powerful evolution equation introduced in eq. (2.4). Here we will focus on the Mueller dipole model that neglects the non-linear term completely, but is particularly suitable for calculation of geometric quantities. Usually, eq. (2.4) is solved as an initial value problem: given a scattering matrix at small initial rapidity (y_0), it determines the resulting scattering matrix at any $y \geq y_0$. Note however, that eq. (2.3) is applicable for any type of observable calculable in impact-parameter space, notably observables linked to the geometry of the partonic initial state. As an example, consider the average vertex coordinate position, $\langle z\rangle$, where z is either the x or y coordinate of a dipole. For a single dipole $\langle z\rangle = (z_1 + z_2)/2$, for the two dipole system $\langle z\rangle = (z_1 + z_2 + z_3)/3$, where the two dipoles has a common point z_3, and directly:

$$\frac{\partial\langle\overline{z}\rangle}{\partial y} = \int \mathrm{d}^2\vec{r}_3\kappa_3 \left(\frac{1}{3}z_3 - \frac{1}{6}(z_1 + z_2)\right). \tag{2.5}$$

For more complicated geometric observables, such as eccentricity (see section 3.2), the analytic expressions become quite involved, and must be handled observable by observable. They are, however, quite easy to handle in a Monte Carlo, where any O can be evaluated event by event, and the expectation value extracted from a large statistics sample.

The starting point for the model, is the evolution of an Onium (or $\gamma^* \to \mathrm{q\overline{q}}$) state in transverse space and rapidity, following eq. (2.1). Instead of calculating average quantities directly from the evolution equation, Monte Carlo events are generated, by performing a probabilistic evolution of a given initial state, corresponding to a collision event performed by an experiment. The calculational details of performing such an evolution are deferred to section 4 and appendix A. It is, however, important to note here the approximation of this evolution, namely that all dipoles in the dipole-chain radiate independently, removing the non-linear effect from the cascade itself.

After a full evolution in rapidity, a single dipole will have evolved to a chain of dipoles, each of which are allowed to interact with dipoles from another evolved system through gluon exchange. The lowest order interaction between two dipoles, at amplitude level, is single gluon exchange, resulting in two gluon exchange at cross section level. This cross section can be related to the elastic amplitude (cf. section 3.1) through the optical theorem.

JHEP10(2019)026

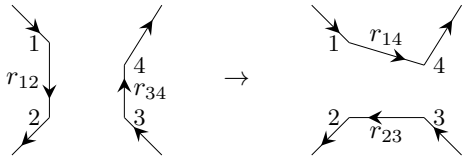


Figure 1. Schematic view of two colliding gluon dipoles. The initial dipoles denoted r_{12} and r_{34} are allowed to interact via two-gluon exchange. This results in the creation of two new dipoles, r_{14} and r_{23} and a connection of the two dipole chains. The lines r_{13} and r_{24} are not drawn, but enters in eq. (2.6).

The dipole-dipole cross section depends on the distances between the interacting dipoles (the enumeration of dipoles follows figure 1) as [49]:

$$\begin{aligned}\frac{\mathrm{d}\sigma_{\mathrm{dip}}}{\mathrm{d}^2\vec{b}} &= \frac{\alpha_s^2 C_F}{N_c}\log^2\left[\frac{r_{13}r_{24}}{r_{14}r_{23}}\right] \\ &\to \frac{\alpha_s^2}{2}\log^2\left[\frac{r_{13}r_{24}}{r_{14}r_{23}}\right] \equiv f_{ij},\end{aligned} \tag{2.6}$$

where the arrow indicates that the 't Hooft large-N_c limit[2] is taken to reach line 2 of eq. (2.6), which then defines f_{ij}. The 't Hooft large-N_c limit is taken in order to ensure consistency with the leading logarithmic approximation in the (BFKL) evolution. The distances r_{ij} are indicated in figure 1, except for r_{13} and r_{24}, the distances between (anti-)colour-(anti-)colour pairs (1,3) and (2,4). Where the dipole evolution comes with a factor of $\alpha_s N_c$, the dipole-dipole cross section is proportional to α_s^2. Thus in the 't Hooft limit where $\alpha_S \sim 1/N_c$, the dipole evolution is of order $N_c^0 \sim 1$, while the dipole-dipole interaction is of order $1/N_c^2$. This means the dipole-dipole interaction is formally N_c-suppressed compared to the dipole evolution [49].

A single collision can contain several dipole-dipole scatterings, equivalent to MPIs in a standard parton language. Assuming that the individual scatterings are uncorrelated, the contribution from each scattering exponentiates, resulting in the unitarized scattering amplitude for a single event (see section 3.1):

$$T(\vec{b}) = 1 - \exp\left(-\sum_{ij} f_{ij}\right). \tag{2.7}$$

As each f_{ij} comes with a factor of $1/N_c^2$, the unitarized scattering amplitude correctly resums $1/N_c^2$-suppressed terms in the interaction. In Regge terminology, each scattering f_{ij} can be viewed as a Pomeron exchange. The first term in the expansion corresponds to single Pomeron exchange, and the latter terms to multi-Pomeron exchanges. The unitarized scattering amplitude can thus also be viewed as a resummation of all possible Pomeron exchanges in the collision frame.

An expansion of the exponent in eq. (2.7) into a power series results in factors of $(\sum_{ij} f_{ij})^n$. To second order this results in a term quadratic in $\sum_{ij} f_{ij}$, which corresponds to

[2]The 't Hooft large-N_c limit is the limit where factors of $\alpha_S N_c$ are kept fixed while factors of $1/N_c^2$ are suppressed.

JHEP10(2019)026

the mean-field approximation of the non-linear term from eq. (2.4). As this non-linear term corresponds to saturation, we note that the dipole framework does not include saturation in the evolution, but, when using the unitarized scattering amplitude, non-linearity is included in the interaction frame, and only there.

2.3 Dipole evolution beyond leading order

Significant formal progress has been made in the pursuit of systematic next-to-leading order (NLO) in α_s corrections to the BK equation [50] and the full B-JIMWLK hierarchy [51–53]. Numerical studies of NLO BK [54] have, however, shown that the equation becomes unstable for some values of the initial conditions, making it yet unsuitable for a full Monte Carlo implementation. Recent work by Ducloué et al. [55] have shown that, for a specific choice of the initial scattering matrix, some problems of unphysical results can be overcome in the dilute-dense limit, by reformulating the NLO evolution equation w.r.t. rapidity of the dense target. This gives hope that a future improvement of the model, implemented as a Monte Carlo in this paper, could include formal improvements beyond leading order, but at this point it is not deemed feasible.

An approach for going beyond leading colour in the cascade, which is also suited for Monte Carlo implementation, is the so-called "swing" mechanism, introduced by Avsar et al. [43, 56]. This can be understood as an extension of the identification of multiple interactions in the collision frame with Pomeron loops, as presented in the previous section. Since loops cannot be formed during the BFKL-like evolution, only loops cut in the collision frame are included. The problem is then posed as equivalent to forming $1/N_c^2$ suppressed dipole configurations in the evolution, by allowing dipoles to reconnect in such a way that the formalism becomes frame independent. This is another viable path for future extensions beyond leading order. Further work on the formalism is needed, however. Currently only a $2 \to 2$ dipole swing has been thoroughly studied, which is not enough to make the formalism fully frame independent. Going beyond $2 \to 2$ is a full study by itself, and not considered in the present paper.

In this paper we instead choose to include corrections beyond (formal) leading-log arising from energy-momentum conservation. It is well known that the leading-log BFKL equation, derived in the high-energy limit, will get sizable corrections at collider energies [57]. From studies of the full next-to-leading log BFKL [58, 59], it is shown that contributions beyond leading log are very large, and a sizable amount are related to energy-momentum conservation [60]. In a Monte Carlo such corrections can be implemented directly, see details in appendix A. Related are non-eikonal corrections. Non-eikonal corrections arise due to the large but finite energy available during the cascade. In the CGC approach this can be understood as sub-leading effects to infinite Lorentz dilation of the projectile, which are troublesome but manageable analytically [61]. In a Monte Carlo implementation of the dipole model, the finite energy can be treated as recoil effects in the dipole splittings.

A non-perturbative effect from confinement is also included in our simulation. This must be done both in the cascade, where large dipoles must be suppressed, and in the interaction, where the range of the interaction must be limited to take confinement into account. Following ref. [42], this is done by replacing $1/p_g^2$ in the Coulomb propagator

JHEP10(2019)026

implicitly entering eq. (2.6) by $1/(p_g^2 + M_g^2)$, where M_g can be taken as a confinement scale, or a fictitious gluon mass. This changes the expressions for the splitting kernel and the dipole-dipole interaction probability, the full expressions are written in section 4.

3 From substructure to observables

The following section is dedicated to the introduction of the framework used for linking partonic substructure to physical observables, such as cross sections and flow coefficients. First, we describe the Good-Walker formalism for calculating cross sections for particles with an inner structure, from obtained scattering amplitudes, and secondly, the apparent scaling of flow coefficients with initial state eccentricity seen in heavy ion collisions is explained.

3.1 The Good-Walker formalism and cross sections

The Good-Walker formalism is a method of calculating cross sections of particles with a well-defined wave function. It includes a normalised and complete set of eigenstates $\{|\psi_i\rangle\}$ of the imaginary part of the scattering amplitude (neglecting the real part, which is vanishing at high energies), denoted $\hat{T}(\vec{b})$ (related to the $\hat{S}$-matrix through $\hat{T} \equiv 1 - \hat{S}$), with eigenvalues $\hat{T}(\vec{b})|\psi_i\rangle = T_i(\vec{b})|\psi_i\rangle$. These scattering states have equal quantum numbers, but differ in masses. The wave function of the incoming beams can thus be expressed in terms of the above eigenstates, and written in short-hand notation as

$$|\psi_I\rangle = |\psi_p, \psi_t\rangle = \sum_{p=1}^{N_p} \sum_{t=1}^{N_t} c_p c_t |\psi_p, \psi_t\rangle \tag{3.1}$$

with $|\psi_{p,t}\rangle$ denoting the projectile and target wave functions, respectively, and $c_{p,t}$ the expansion coefficients. The scattered wave function is found by operating with the transition matrix on the incoming wave function,

$$|\psi_S\rangle = \hat{T}(\vec{b})|\psi_I\rangle = \sum_{p,t} c_p c_t \; T_{p,t}(\vec{b})|\psi_p, \psi_t\rangle \,, \tag{3.2}$$

and from these definitions, the profile function for elastic scattering (at fixed Mandelstam s) can be defined:

$$\begin{aligned} \Gamma_{\text{el}}(\vec{b}) &= \langle \psi_S | \psi_I \rangle = \sum_{p,t} |c_p|^2 |c_t|^2 \; T_{p,t}(\vec{b}) \langle \psi_p, \psi_t | \psi_p, \psi_t \rangle \\ &= \sum_{p,t} |c_p|^2 |c_t|^2 \; T_{p,t}(\vec{b}) \equiv \langle T(\vec{b}) \rangle_{p,t} \,, \end{aligned} \tag{3.3}$$

where we have defined an average over projectile and target states in the last equality and suppressed indices on T inside the average (which is done in all the following, unless specifically noted otherwise). Thus we obtain the cross sections and elastic slope in the

JHEP10(2019)026

eikonal approximation (again also at fixed Mandelstam s),

$$\sigma_{\rm tot} = 2\int \mathrm{d}^2\vec{b}\Gamma(\vec{b}) = 2\int \mathrm{d}^2\vec{b}\ \langle T(\vec{b})\rangle_{p,t}, \tag{3.4}$$

$$\sigma_{\rm el} = \int \mathrm{d}^2\vec{b}|\Gamma(\vec{b})|^2 = \int \mathrm{d}^2\vec{b}\ \langle T(\vec{b})\rangle^2_{p,t}, \tag{3.5}$$

$$B_{\rm el} = \frac{\partial}{\partial t}\log\left(\frac{\mathrm{d}\sigma_{\rm el}}{\mathrm{d}t}\right)\Big|_{t=0} = \frac{\int \mathrm{d}^2\vec{b}\ b^2/2\ \langle T(\vec{b})\rangle_{p,t}}{\int \mathrm{d}^2\vec{b}\ \langle T(\vec{b})\rangle_{p,t}}\,. \tag{3.6}$$

In eqs. (3.4)–(3.6) we have implicitly assumed a particle wave function $\langle\psi|\psi\rangle = 1$. In cases where the wave function is not normalizable, one has to take into account the wave function in the above cross sections. This includes processes with photons, where the wave function is well-defined in pQCD for high virtualities. The total γ^*p cross section would thus require an additional integration over wave function parameters:

$$\sigma^{\gamma^*\mathrm{p}}(s) = \int_0^1 \mathrm{d}z \int_0^{r_{\rm max}} r\mathrm{d}r \int_0^{2\pi} \mathrm{d}\phi \left(|\psi_L(z,r)|^2 + |\psi_T(z,r)|^2\right)\sigma_{\rm tot}(z,\vec{r}), \tag{3.7}$$

with z the fractional momentum carried by the quark, r the distance between the quark and anti-quark, $\psi_{L,T}$ the longitudinal and transverse parts of the photon wave function and $\sigma(z,\vec{r})$ the dipole cross section calculated from the elastic profile function, eq. (3.4). The photon wave function implemented in our approach is given in eqs. (4.1)–(4.2) and the discussion for γ^*A is continued in section 7.

3.2 Eccentricity scaling of flow observables

Anisotropic flow is measured as momentum space anisotropies and quantified in flow coefficients (v_n), obtained by a Fourier expansion of the azimuthal (ϕ) spectrum:

$$\frac{\mathrm{d}N}{\mathrm{d}\phi} \propto 1 + 2\sum_n v_n \cos\left[n(\phi - \Psi_n)\right], \tag{3.8}$$

with Ψ_n the symmetry plane of the nth harmonic. For a hydrodynamical expansion, it has been shown that v_2 and v_3 are proportional to the initial state eccentricity in the corresponding harmonic, $v_n \propto \epsilon_n$, to a very good approximation [62], with the constant of proportionality depending on the properties of the QGP transporting the initial state anisotropy to the final state. A similar relation may be expected when a pressure gradient is obtained without a thermalized or hydrodynamized plasma [22, 63]. In the following, the eccentricities will therefore be taken as a proxy for flow observables, noting that the model imposed for the response may deviate from this linear scaling behaviour. In pp and pA collisions this type of behaviour becomes very apparent, due to the dominance of non-flow effects,[3] in particular at small event multiplicities. Non-flow mechanisms aside, it is clear that no matter what the actual response is, measurable observables will be affected by large deviations in predicted eccentricities.

[3]Including correlations from jets and due to particle decays.

JHEP10(2019)026

We follow the usual definition of the initial anisotropy or participant eccentricity [64, 65]:

$$\epsilon_n = \frac{\sqrt{\langle r^2 \cos(n\phi)\rangle^2 + \langle r^2 \sin(n\phi)\rangle^2}}{\langle r^2\rangle}. \tag{3.9}$$

Here r and ϕ are usual polar coordinates, with the origin shifted to the center of the distribution. From eq. (3.9), higher order cumulants can be calculated:

$$\epsilon_n^2\{2\} = \langle\epsilon_n^2\rangle, \tag{3.10}$$
$$\epsilon_n^4\{4\} = 2\langle\epsilon_n^2\rangle^2 - \langle\epsilon_n^4\rangle, \tag{3.11}$$
$$4\epsilon_n^6\{6\} = \langle\epsilon_n^6\rangle - 9\langle\epsilon_n^4\rangle\langle\epsilon_n^2\rangle + 12\langle\epsilon_n^2\rangle^3, \tag{3.12}$$
$$33\epsilon_n^8\{8\} = 144\langle\epsilon_n^2\rangle^4 + 18\langle\epsilon_n^4\rangle^2 + 16\langle\epsilon_n^6\rangle\langle\epsilon_n^2\rangle - 144\langle\epsilon_n^4\rangle\langle\epsilon_n^2\rangle^2 - \langle\epsilon_n^8\rangle. \tag{3.13}$$

In nuclear collisions, the normal participant nucleon eccentricity is used as a baseline. The notion of "participating" is, however, a model dependent statement. We use the definition from Angantyr [35, 36], which defines participating nucleons as either "inelastically" or "absorptively" (inelastic non-diffractively) wounded, see appendix B for a brief review. For pp collisions, and for fluctuations in nuclear collisions, we follow Avsar et al. [25], and define a participant *parton* eccentricity (though somewhat modified from the cited exploratory work). Assuming that the hydrodynamic evolution takes place at the end of the perturbative parton cascade, the participant parton eccentricity should be evaluated at this point in the evolution. In section 6.1 this participant parton eccentricity will be compared to a more purist initial state approach, where the final state parton cascade is not included. This is meant to inform a discussion about what the notion of an "initial state" really ought to entail.

Parton level eccentricities are, however, not infrared safe. Consider the simple example of a soft gluon emission at the same impact parameter point as its mother. Such an emission will double count this spatial point at parton level, but disappear after hadronization, which will place two such partons inside the same hadron. To improve this, all contributions are weighted by a factor $p_\perp/(p_\perp + p_{\perp\min})$, where $p_{\perp\min} = 0.1\,\text{GeV}$ ensures that considerably soft gluons will not double count.

Normalised symmetric cumulants will also be studied. Such quantities eliminate the dependence on the magnitude of the flow coefficients, and should thus remove the response factor between flow harmonics and eccentricities, and directly probe the substructure [66]. They are defined as:

$$NSC(n,m) = \frac{\langle v_n^2 v_m^2\rangle - \langle v_n^2\rangle\langle v_m^2\rangle}{\langle v_n^2\rangle\langle v_m^2\rangle} \approx \frac{\langle \epsilon_n^2 \epsilon_m^2\rangle - \langle\epsilon_n^2\rangle\langle\epsilon_m^2\rangle}{\langle\epsilon_n^2\rangle\langle\epsilon_m^2\rangle}, \tag{3.14}$$

where the last approximate equality indicates the removal of the response. Especially interesting for this study is $NSC(3,2)$, it being sensitive to initial-state fluctuations, namely the geometric correlation between ϵ_2 and ϵ_3, the elliptical and triangular parts of the Fourier expansion.

Finally it is noted that, since the model is implemented in a full event generator able to generate full final states for pp, pA and AA collisions, it is possible to investigate

JHEP10(2019)026

the event geometry as a function of final state multiplicity with the same acceptance as the experiment.

4 Monte Carlo implementation

In this section, the Monte Carlo implementation of Mueller's model is briefly described. The full details are given in appendix A. First, the details of the various initial states are described, then some assumptions on the cascade and the interaction are described, and lastly, some geometric properties of the evolution are presented.

4.1 The initial states

The new implementation is applicable for both virtual photon and proton beams. A photon state is represented by a single dipole, with a wave function given as,

$$|\psi_L(z,r)|^2 = \frac{6\alpha_{\text{em}}}{\pi^2}\sum_q e_q^2 Q^2 z^2(1-z)^2 K_0^2\left(\sqrt{z(1-z)}Qr\right) \tag{4.1}$$

$$|\psi_T(z,r)|^2 = \frac{3\alpha_{\text{em}}}{2\pi^2}\sum_q e_q^2 Q^2\left[z^2+(1-z)^2\right] z(1-z) K_1^2\left(\sqrt{z(1-z)}Qr\right), \tag{4.2}$$

where we include the three lightest (massless) quarks. Here z is the longitudinal momentum fraction carried by the quark, $(1-z)$ the longitudinal momentum fraction carried by the anti-quark, r is the distance between them, Q^2 the photon virtuality and K_i the modified Bessel functions. For protons, the wave function is not known. Instead it is represented by three dipoles in an equilateral triangle configuration and normalized to unity, shown previously to give the best description of data [56]. The lengths of the initial dipoles are allowed to fluctuate on an event-by-event basis, chosen from a Gaussian distribution with mean r_0, and width r_{w}.

4.2 The dipole evolution

To implement eq. (2.1) as a parton shower, it is modified by a Sudakov factor:

$$\begin{aligned}\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}y\ \mathrm{d}^2\vec{r}_3} &= \frac{N_c\alpha_s}{2\pi^2}\frac{r_{12}^2}{r_{13}^2 r_{23}^2}\exp\left(-\int_{y_{\min}}^{y}\mathrm{d}y\mathrm{d}^2\vec{r}_3\ \frac{N_c\alpha_s}{2\pi^2}\frac{r_{12}^2}{r_{13}^2 r_{23}^2}\right)\\ &\equiv \frac{N_c\alpha_s}{2\pi^2}\frac{r_{12}^2}{r_{13}^2 r_{23}^2}\Delta(y_{\min},y),\end{aligned} \tag{4.3}$$

allowing for a trial emission from each dipole in the cascade. The strategy of "the winner takes it all" is then employed, such that only the trial emission with the lowest rapidity is chosen as a true branching. This lowest rapidity then becomes the minimal rapidity in the next (trial) emission. The process is reiterated until none of the trial emissions are below a maximal rapidity, governed by the energy of the collision,

$$\text{pp}: \quad y_{\max}^{\text{p}} = \log\left(\frac{\sqrt{s}}{m_p}\right), \tag{4.4}$$

$$\gamma^*\text{p}: \quad y_{\max}^{\text{p},\gamma^*} = \log\left(\frac{W}{m_0}\right), \tag{4.5}$$

JHEP10(2019)026

Parameter	Value	Meaning
r_0 [fm]	1.	Mean of normally distributed initial dipole sizes
r_w [fm]	0.	Width of normally distributed initial dipole sizes
r_max [fm]	1.	Maximally allowed dipole size (confined evolution only)
α_s	0.21	Value of fixed strong coupling

Table 1. The input parameters used in this section.

with m_p the proton mass, m_0 a reference scale set to 1 GeV, and $W, \sqrt{s}$ the γ^*p and pp center-of-mass (CM) energies, respectively. Note that eq. (4.5) is an approximation to the actual rapidity available for the dipole formed by a virtual photon. The "true" rapidity is not well-defined for virtual photons, as it depends not only on W, but also on Q^2 and momentum fractions carried by the quark and anti-quark ends of the dipole. This introduces different rapidity ranges available for either end of the dipole, complicating the evolution further. Equation (4.5) was chosen as the simplest possible rapidity range.

If confinement is taken into account (as described in section 2.3), the evolution equation is modified accordingly:

$$\frac{\mathrm{d}P}{\mathrm{d}y\ \mathrm{d}^2\vec{r}} = \frac{N_c\alpha_s}{2\pi^2}\frac{1}{r_\mathrm{max}^2}\left[\frac{\vec{r}_{13}}{|r_{13}|}K_1(|r_{13}|/r_\mathrm{max}) - \frac{\vec{r}_{23}}{|r_{23}|}K_1(|r_{23}|/r_\mathrm{max})\right]^2 \Delta(y_\mathrm{min}, y), \tag{4.6}$$

with K_1 the modified Bessel functions of the first kind and r_max a maximal radius of the initial dipole, left as a tunable parameter.

4.3 Geometric properties of the dipole evolution

Given a specific parameter set, table 1, the probability distribution in rapidity for the first emission, $\mathrm{d}P/\mathrm{d}y$, is shown in figure 2. This distribution has a mean at around two units of rapidity. Thus, on average, a new emission is assigned a rapidity of roughly two units larger than the mother (or emitter) dipole. It is worth noting that the inclusion of confinement effects slightly increases the mean as compared to the unconfined distribution. This is caused by the additional suppression of large dipoles, requiring large dipoles to be discarded in the evolution and an emission at a larger rapidity to be tried.

In each step of the dipole evolution a mother dipole is split into two daughters. Figure 3 shows the distribution in sizes of the smaller and larger dipole, scaled w.r.t. their mothers' size for three different evolutions ($y_\mathrm{max} = 4, 8, 12$). Here, it is evident that on average the larger dipole retains the size of the mother, while the distribution of the smaller is much broader. At lower y_max there is a bump in the distribution at around 30–40% of the mothers size, while this bump is less pronounced at larger maximal rapidity.

Figure 4 shows the corresponding average and standard deviation in the lengths of all the daughter dipoles scaled w.r.t. the length of their mothers', as a function of maximal rapidity of the evolution. As stated above, it is clear that while the larger of the two daughter dipoles can be taken to be identical to the mother dipole, the size of the smaller dipole has larger fluctuations. The average size of the smaller dipole is, however, fixed at roughly a third of the mother dipole for all y_max.

JHEP10(2019)026

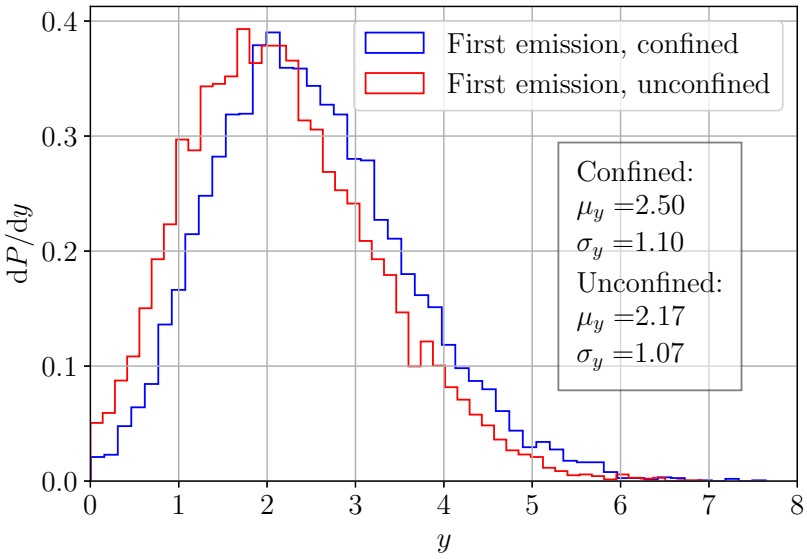


Figure 2. The probability distribution in rapidity for unconfined (a) and confined (b) dipole-evolution. The box shows the average and spread of the distributions.

After a full evolution, an initial proton consisting of three dipoles will have evolved to a larger set of dipoles of mostly smaller sizes than the initial dipoles, cf. figure 5. From these two figures it is evident that the effect of confinement plays a large role in the evolution, effectively reducing the number of large dipoles in the final configuration. Thus, as $\sigma_{\text{dip}} \sim r^2$, confinement is expected to play a large role when evaluating the cross sections. Confinement also introduces more activity — or hot spots — around the endpoints of the dipoles.

4.4 Pascal approximation for dipole evolution

The full dipole evolution can be approximated based on the geometric observation above. On average one dipole splitting happens per two units of rapidity, and the lengths of the two resulting dipoles after the splitting, are approximately equal to and one third the size of the mother dipole respectively. This behaviour is tabulated in table 2 for four generations of evolution. Similar results have been observed within the Dipsy framework, although their dipole swing slightly increases the size of the smaller dipole in a branching [67] to half of the size of the mother dipole.

The number of dipoles in table 2 follows the coefficients of the binomial theorem, with the number in column n row k being equal to $\binom{n}{k}$, and can thus be arranged to form Pascal's triangle. The total number of dipoles after a given number of generation, as well as the number of dipoles of a certain size, can be quickly approximated this way. Knowing the positions of the initial dipoles and the emitted dipole sizes, positions of all dipoles can also be inferred.

To exploit further these simple relations in the dipole evolution, we have created an alternative toy-model denoted the Pascal approximation. Here, the step size in rapidity (Δy) and the size of the smaller dipole ($r_s = f r_m$) in a branching are implemented as

JHEP10(2019)026

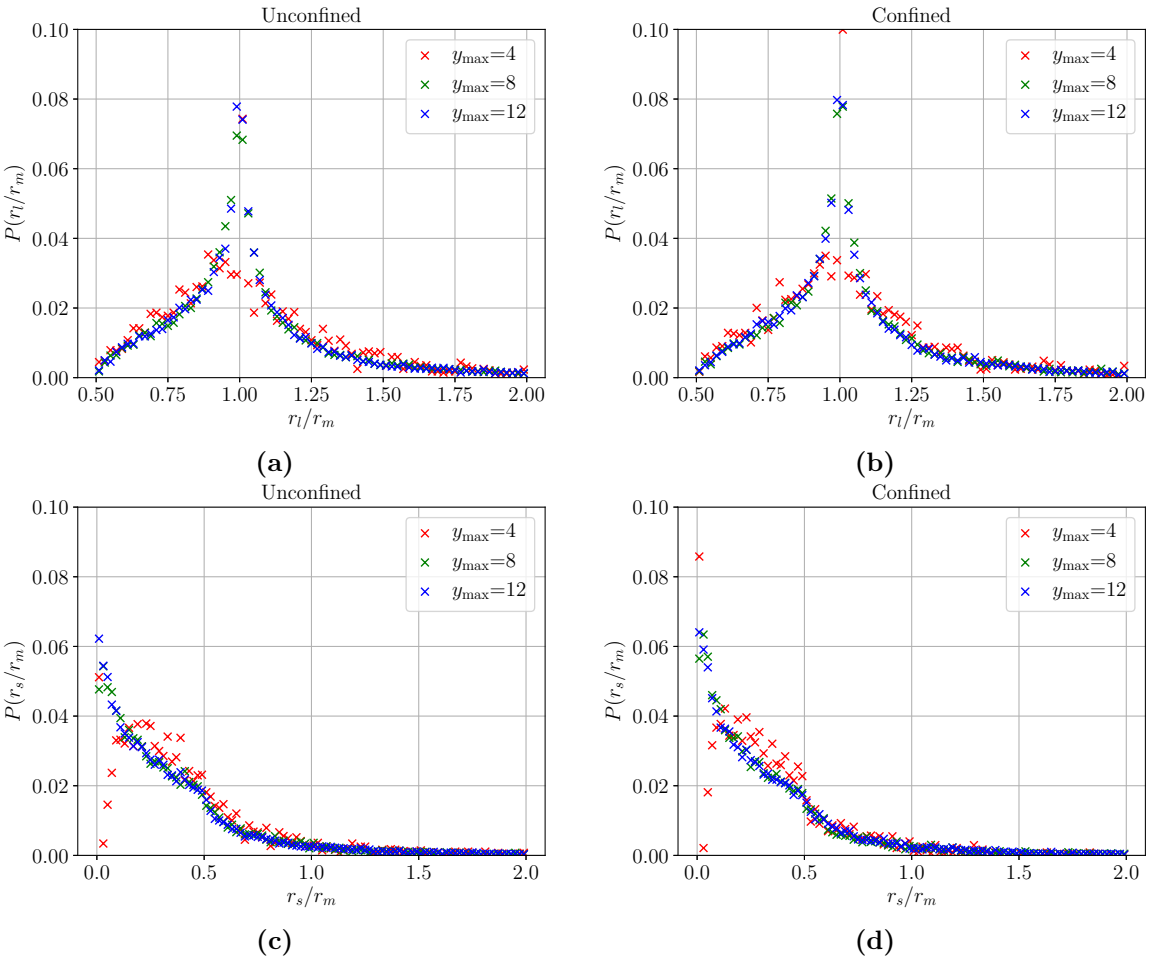


Figure 3. The scaled lengths of the daughter dipoles w.r.t. the mother dipole as a function of maximal rapidity for unconfined (a,c) and confined (b,d) dipole-evolution. Figures (a,b) shows the larger of the daughter dipoles, while figures (c,d) shows the smaller of the daughter dipoles. The parameters used in the dipole evolution are the same as presented in table 1.

	y = 0	y = 2	y = 4	y = 6	y = 8
r	1	1	1	1	1
$r/3$	0	1	2	3	4
$r/9$	0	0	1	3	6
$r/27$	0	0	0	1	4
$r/81$	0	0	0	0	1
N_{dip}	1	2	4	8	16

Table 2. Approximate behaviour of dipole evolution for four generations of dipoles. The number of dipoles in column n row k is equal to the binomial coefficient $\binom{n}{k}$.

JHEP10(2019)026

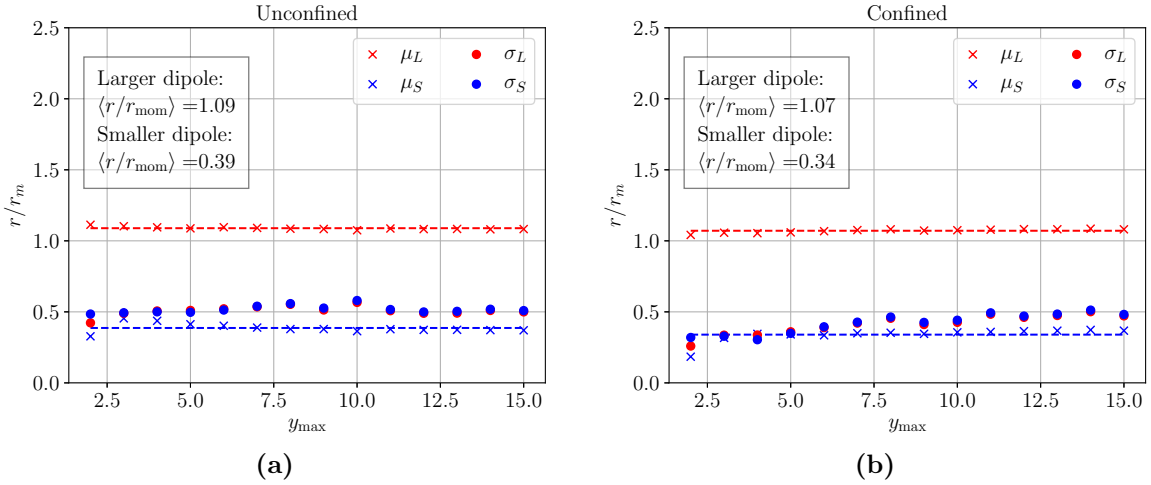


Figure 4. The scaled lengths of the daughter dipoles w.r.t. the mother dipole as a function of maximal rapidity for unconfined (a) and confined (b) dipole-evolution. The parameters used in the dipole evolution are the same as presented in table 1.

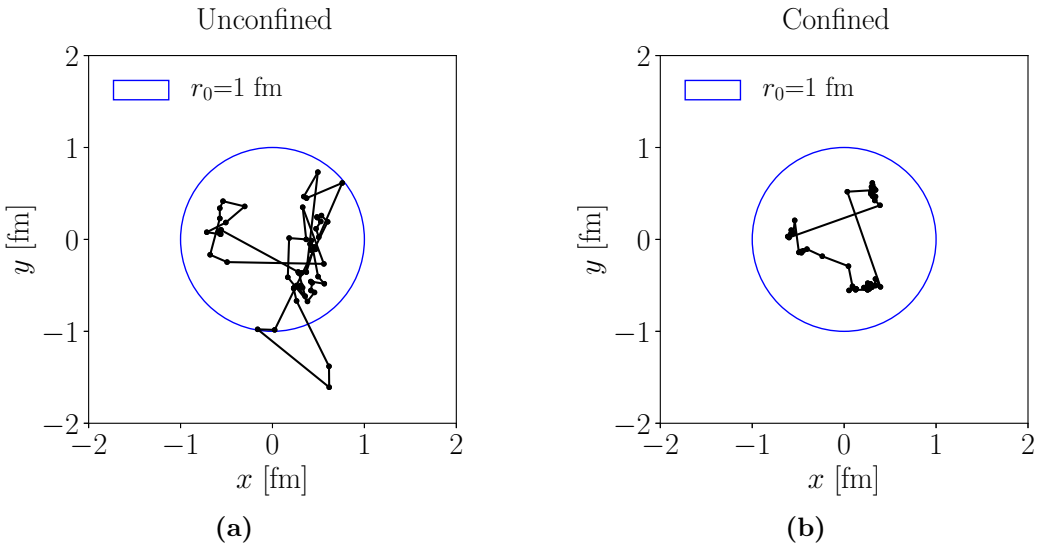
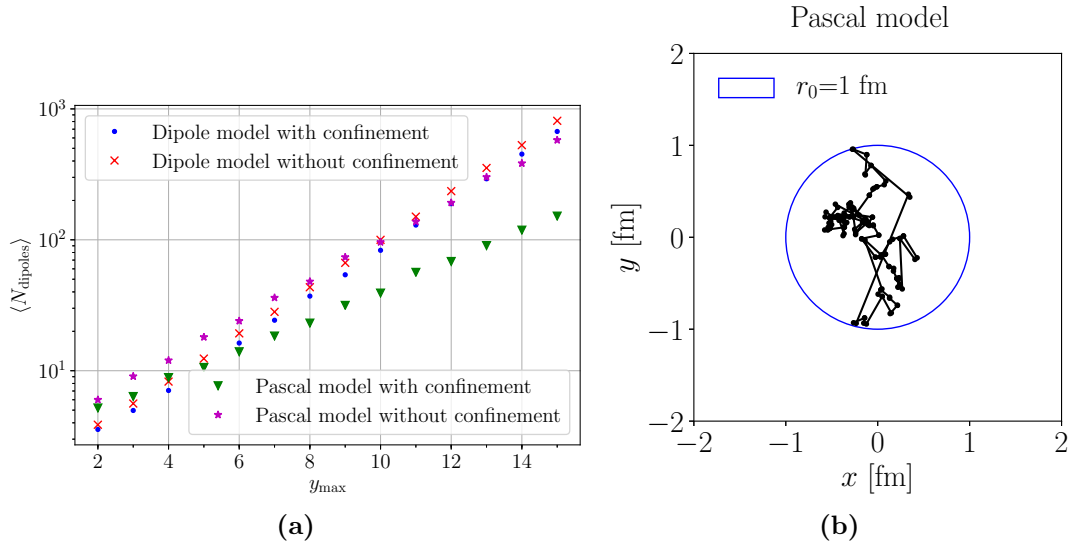


Figure 5. An initial state proton consisting of three dipoles in an equilateral triangle configuration after a full evolution at 7 TeV (corresponding to $y_{\max} = 8.86$). (a) has been evolved without confinement, while (b) has been evolved with confinement. The parameters of table 1 have also been applied in this evolution.

tunable parameters, with r_m the size of the mother dipole and f a tunable fraction. The number of steps taken in total is calculated from the step size, $N_{\text{steps}} = y^b_{\max}/\Delta y$ with $b = \text{p}, \gamma^*$ and $y_{\max}$ given in eqs. (4.4)–(4.5). To mimic the recoil effects in the full dipole evolution, a Gaussian smearing of the daughter lengths is introduced with mean $\mu = r_m, r_s$ for the larger and smaller daughters, respectively. Knowing the lengths of the mother and daughter dipoles, they are placed in transverse space by calculating the angles of the triangle spanned by the endpoints of the three (connected) dipoles.

JHEP10(2019)026



(a) **(b)**

Figure 6. (a) The average number of dipoles inside a proton after a full evolution to maximal rapidity y. (b) An initial state proton consisting of three dipoles in an equilateral triangle configuration after a full evolution with the Pascal model at 7 TeV (corresponding to $y_{\text{max}} = 8.86$).

This simple approximation is useful for introducing toy-models for sub-leading effects, such as confinement and saturation, as basic quantities like total number of dipoles after a given evolution, can be calculated analytically. A crude model for confinement is introduced by requiring the length of the emitted dipoles to not exceed a tunable maximally allowed cutoff, r_{max}. If this occurs, the branching is discarded and the next step in rapidity is tried. Once the full evolution has occurred, each of the dipoles are allowed to interact using the dipole-dipole scattering amplitude given in eq. (2.6) (or eq. (4.7) for the confined version).

In figure 6(a) the average number of dipoles in a single proton after a full evolution to maximal rapidity y_{max} is shown. The same parameters as given in table 1 are used, while the parameters $f, \Delta y$ are extracted from figures 2 and 3. The unconfined Pascal model follow the same trend as the full dipole model, albeit having a slightly larger average number of dipoles at small maximal rapidity. It is evident that the confined Pascal model has a different slope than the full dipole model and the unconfined Pascal model, with the effect of the crude model for confinement clearly seen at large y_{max}. Here, the confined Pascal model results in a smaller average number of dipoles as compared to the full dipole model.

In figure 6 (b) the dipole configuration of an evolved proton for $y_{\text{max}} = 8.86$ is shown. This has more features in common with full, unconfined dipole evolution than full, confined dipole evolution, with dipoles being more randomly distributed, than focused around hot spots.

A practical advantage of the Pascal approximation, besides being a toy model for testing models for sub-leading effects, is its computational speed. For simple cascade-quantities like numbers of dipoles, results can be calculated analytically. With inclusion of geometry, event-by-event results can be generated approximately a factor of 1000 faster,

JHEP10(2019)026

for large maximal rapidity ($y_{\max} \geq 10$). It thus serves as a decent replacement for the full dipole evolution model for calculation of cascade properties with limited computational resources. For full calculations of amplitudes and cross sections, the efficiency gain is not nearly as large, as in that case the bottleneck is the calculation of all f_{ij} in eq. (2.6).

4.5 Dipole-dipole interactions

The dipole-dipole interactions are defined to occur at rapidity zero and given by eq. (2.6). If confinement is introduced in the splitting kernel (eq. (4.6)), one also has to change the interaction probability in order to make the event generation consistent. This modifies eq. (2.6) to

$$f_{ij} = \frac{\alpha_s^2}{2}\left[K_0\left(\frac{r_{13}}{r_{\max}}\right) + K_0\left(\frac{r_{24}}{r_{\max}}\right) - K_0\left(\frac{r_{14}}{r_{\max}}\right) - K_0\left(\frac{r_{23}}{r_{\max}}\right)\right]^2, \tag{4.7}$$

where K_0 is the modified Bessel function and $r_{\max}$ the maximally allowed size for a dipole in the evolution.

The choice of collision frame, however, is not trivial. Obviously, no observables should depend on the frame-choice of the collision. In practice, the choice does matter, as no sub-leading color corrections are included in the dipole evolution. Previous studies have shown that for symmetric systems, e.g. pp collisions, the optimal frame choice is the center-of-mass (CM) frame [68]. This is also utilized in our approach, cf. eq. (4.4), where both beams are evolved the same distance in rapidity. In asymmetric systems, such as γ^*p or pA systems, the CM frame lies more towards the heavier of the two objects, and it has been previously argued that the optimal frame here would be the rest frame of the heavier beams [68]. This, however, is not the choice we have taken. The maximal rapidity chosen in eq. (4.5) is found in what we call the center-of-rapidity frame. Here, both beams are evolved the same distance in rapidity, similarly to what is chosen in symmetric collision systems. As already stated, this work does not attempt to include sub-leading colour effects in the evolution, thus frame-independence is not possible to obtain. Hence the simplest choice has been made to use the same frame for all systems, i.e. the center-of-rapidity frame given in eqs. (4.4)–(4.5).

4.6 Assigning spatial vertices to MPIs

In order to utilize the formalism developed so far in real pp, pA and AA events, the dipole cascade is matched to the Pythia 8 MPI model [69]. This allows for evaluation of geometric initial state quantities, such as eccentricities (see section 3.2), at fixed number of charged hadrons in the final state, using a similar definition of charged particles as the experiments. The Pythia 8 MPI model considers pp collisions, treating all partonic sub-collisions as separate $2 \to 2$ QCD scatterings, which are uncorrelated up to momentum conservation. Other factors present in the MPI model is a rescaling of the parton density between each scattering, preservation of valence quark content and a sophisticated treatment of beam remnants [70].

In the MPI framework, the sub-collisions and their kinematics are selected using the normal $2 \to 2$ QCD cross section. But since this cross section diverges at low $p_\perp$, the

JHEP10(2019)026

expression is regulated using a parameter, $p_{\perp 0}$:

$$\frac{\mathrm{d}\sigma_{2\to 2}}{\mathrm{d}p_\perp^2} \propto \frac{\alpha_s^2(p_\perp^2)}{p_\perp^4} \to \frac{\alpha_s^2(p_\perp^2 + p_{\perp 0}^2)}{(p_\perp^2 + p_{\perp 0}^2)^2}. \tag{4.8}$$

For matching of vertices to each individual partonic sub-collision, it is also useful to note that MPIs are generated in decreasing order of $p_\perp$, starting from a (process-dependent) maximal scale. This decreasing order is generated from a Sudakov-like expression of the form:

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}p_{\perp i}} = \frac{1}{\sigma_{\mathrm{ND}}} \frac{\mathrm{d}\sigma}{\mathrm{d}p_{\perp i}} \exp\left[-\int_{p_{\perp i}}^{p_{\perp i-1}} \frac{1}{\sigma_{\mathrm{ND}}} \frac{\mathrm{d}\sigma}{\mathrm{d}p'_\perp} \mathrm{d}p'_\perp \right], \tag{4.9}$$

with $\mathrm{d}\sigma/\mathrm{d}p_{\perp i}$ given by eq. (4.8). The impact-parameter of the collision is also taken into account in the evolution by connecting the average number of MPIs to the overlap $\mathcal{O}(b)$ of the two colliding protons. This introduces additional factors of $\mathcal{O}(b)/\langle\mathcal{O}(b)\rangle$ in eq. (4.9) along with the need to select the impact parameter consistently.[4] Furthermore, new partons are generated by initial- and final-state radiation.

Recently, a method of assigning space-time information to the MPIs in Pythia 8 was introduced [22]. Here, the transverse coordinates are sampled from a two-dimensional Gaussian distribution defined by the overlap of the mass distributions of the two colliding protons. The width of the Gaussian is a free parameter (which should not be too far from the proton radius) and a mean equal to the impact parameter chosen in the MPI framework. Initial- and final-state radiation are then treated as small displacements of the selected anchor points of the MPIs. This introduces and additional smearing of an MPI vertex whenever a parton is radiated off from the partons involved in the MPI. The smearing is done using another Gaussian with a width of $\sigma_\perp/\mathrm{p}_\perp$, where $\sigma_\perp$ is a parameter with default value 0.1 GeV·fm.

In this work, we utilize the dipole framework to generate the space-time vertices, instead of the default two-dimensional Gaussian distribution. We currently do not use the dipole model to generate the $p_\perp$-spectrum for the MPIs, but retain the $p_\perp$-distribution obtained internally with Pythia 8. This means that the dipole model is only used to obtain information on the spatial location of the MPIs. Using the dipole framework to generate space-time vertices requires (as with the Gaussian model) some assumptions, as this matching can not be derived from first principles. In order to obtain a reasonable matching, the following is noted:

- Each branch of the (projectile) dipole cascade can be identified as a virtual emission, which goes on shell if, and only if, it collides with a corresponding virtual emission from the target.

- Each proton–proton collision has many *potential* sub-collisions between all combinations of virtual emissions. We order the sub-collisions in terms of contribution to total cross section, thus the MPI with largest $p_\perp$ is identified with the dipole-dipole scattering with the largest f_{ij}.

[4] Here $\langle\mathcal{O}\rangle \equiv \int \mathrm{d}^2\vec{b}\mathcal{O}(b) / \int \mathrm{d}^2\vec{b}\,[1 - \exp(-k\mathcal{O}(b))]$, where k is constrained by the ratio of the dampened $2 \to 2$ QCD cross section in equation (4.8) to the total non-diffractive cross section.

JHEP10(2019)026

The concrete matching is done by first generating two dipole cascades, and allowing them to collide with the same impact parameter used in the generation of the MPIs. This produces a list of possible dipole-dipole collisions, each with an interaction strength f_{ij}. As the MPIs are generated (from hardest to softest), they are each assigned a vertex sampled from this list with a weight equal to f_{ij}, normalised to the summed dipole-dipole interaction strength (and *not* the unitarized interaction). The vertex is simply given as the mean of the transverse coordinates of the dipoles in the interaction. Once a set of dipoles have been assigned to an MPI, they are both flagged as used, and cannot initially be re-used to ensure a reasonable spread. In cases where the list runs out of interactions containing only unused dipoles, the dipoles are allowed to re-interact, though not with the same dipole as initially.

As opposed to the default model, vertices are now selected from a distribution which event-by-event is asymmetric, and contains "hot spots" with large activity, as shown in figure 5 (b) for the full evolution including confinement effects.

The matching of largest f_{ij} to hardest MPI requires further discussion, as one could argue the opposite. The dipole-dipole scattering amplitude is driven by the distances between the endpoints of the interacting dipoles, as indicated in figure 1. One can argue that small f_{ij} corresponds to small distances, which in turn corresponds to large $p_\perp$ of the gluons emitted in the interaction. Hard MPIs would with this reasoning correspond to small f_{ij}. This is indeed the choice made in the Dipsy event generator for exclusive final states [40], but opposite to the choice made above. We do, however, also note that the exclusive final states generated by Dipsy describes $p_\perp$ spectra of charged particles poorly, in particular the high-$p_\perp$ part of the distributions vastly overshoots data. We therefore currently refrain from associating the dipole sizes directly to the $p_\perp$ of emerging partons, but rather give larger attention to the cross section. We note that large f_{ij} interactions dominate the cross sections. A guiding principle is therefore to ensure that such interactions are *always* identified with an MPI, by assigning it first.

There are several possible future improvements of the matching technique. A small improvement of the existing technique, could include to also identify initial state radiation with emissions going on shell, and assign them vertices from the cascade as such. Going beyond improvement of matching techniques, would be a full re-evaluation of the MPI model, with the dipole cascade and interactions as a starting point. The consequences of such an approach could be studied in a toy-model where the $p_\perp$ obtained with the dipole formulation could be utilized instead of the $p_\perp$'s obtained within the Pythia 8 model. It would then be possible to study if this method gives rise to similar problems as the Dipsy MPI description has in the high $p_\perp$ tail.

Instead of creating a completely new model like Dipsy, it should be possible to use the dipole model to improve the existing Pythia 8 model. A first step would be to replace σ_{ND} in eq. (4.9) with a dynamically calculated cross section, event-by-event. Secondly, the parameter $p_{\perp 0}$ in eq. (4.8) could be re-evaluated in terms of the dipole model. The physical interpretation of $p_{\perp 0}$ in the MPI model, is that of a colour screening scale. The perturbative treatment of eq. (4.9) would naively break down at some minimal scale $\sim \hbar/r_p \sim \Lambda_{QCD}$, where r_p is the (colour screening) size of a proton, left as a free parameter. In the dipole

JHEP10(2019)026

model, this colour screening length could be identified as either the transverse size of the cascade after the evolution, or the length of the largest colour connected dipole chain. In that way the energy dependence of $p_{\perp 0}$ would also come for free, instead of having to assume a power-law dependence, as is the default assumption in Pythia 8.

4.6.1 Heavy ion collisions

The method described above can be directly applied to heavy ion collisions as they are modelled in the Angantyr framework for heavy ion collisions in Pythia 8 (see appendix B for a brief review, and refs. [35, 36] for a full description). In the Angantyr model, sub-collisions are chosen using a Glauber-like approach. Sub-collisions are in turn associated with one out of several types of pp events, depending on the properties of the sub-collision. Since all these events are generated using the MPI model described above, the generalization is only a matter of generating vertices for each sub-collision in its local coordinate system, and then moving them to the global coordinate system defined by the Glauber calculation.

5 Results I — comparing cross sections

In this section we present results on integrated cross sections for pp and γ^*p collisions. For pp we present results for both the dipole evolution model and for the Pascal model, while for γ^*p we focus our attention on the dipole model. The main purpose of this section is tuning: the model parameters have to be estimated by comparisons with data, preferably data that we do not aim to make predictions for in later sections.

It is thus not the aim of this section to be able to describe the cross sections perfectly — but more generally, to get an overall agreement between model and data, especially at LHC energies, where we aim to make predictions for the substructure observables.

More dedicated models are available to describe the cross sections at all energies, from the GeV range to the TeV range, results of which are shown alongside results from the dipole model in the pp section. The most widespread model is based on the 1992 total cross section fit by Donnachie and Landshoff (DL) [71] and the models for elastic and diffractive cross sections by Schuler and Sjöstrand (SaS) [72]. Another, more recent model by Appleby et al. (ABMST) [73] is more complex than SaS, and able to describe latest LHC data better. The models are both implemented in Pythia 8, with some additions to the original models [74]. In this paper we compare to the original models, and not those adapted to Pythia 8.

5.1 Results for γ^*p

We begin with the results on photon-proton total cross sections. Here, we compare the dipole evolution model to data obtained from H1 [75] at different energies and virtualities. We note that the photon wave function implemented only includes the three lightest quarks, and none of the vector meson states present at low Q^2. Thus we expect the results to be less precise at low virtualities, where the probability for the photon to fluctuate into a hadronic state becomes non-negligible. Similarly, the masses of the quarks should be taken

JHEP10(2019)026

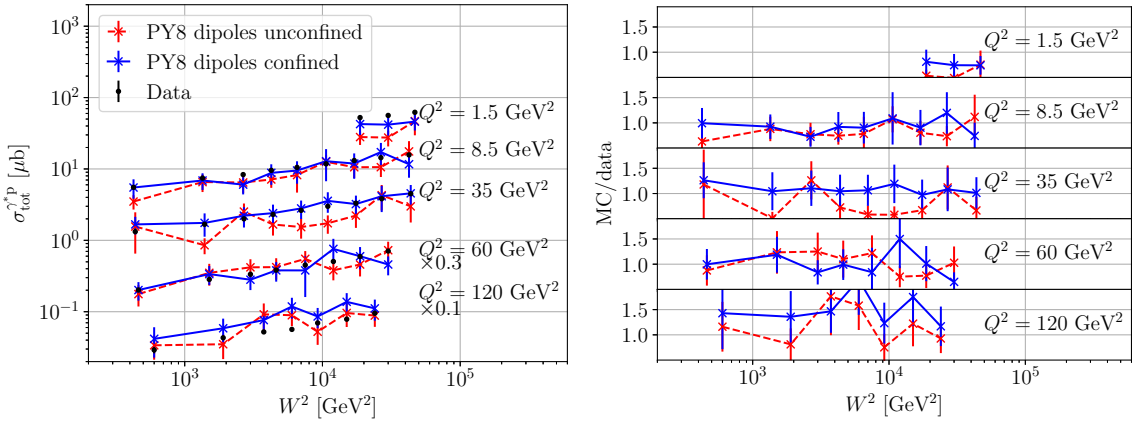


Figure 7. The total photon-proton cross section, $\sigma_{\text{tot}}^{\gamma^*\text{p}}$, as a function of squared photon-proton center-of-mass energy, W^2, for several virtualities (a). Note that the distributions for the two highest virtualities ($Q^2 = 60, 120\,\text{GeV}^2$) have been scaled with a factor of 0.3, 0.1, respectively, for better visibility. (b) shows the ratio MC/data as a function of squared center-of-mass energy, W^2, for the five different virtualities.

into account if the argument of the Bessel functions become close to the squared quark masses, i.e. if

$$z(1-z)Q^2 \simeq m_q^2 \tag{5.1}$$

occurring in the limits $z \to 0, 1$ or if Q^2 small. The contribution from c-quarks are neglected for simplicity, the uncertainty arising from this approximation is discussed at the end of the section.

The H1 data presents results on the proton structure function $F_2(x, Q^2)$ at a large range of virtualities and energies. This is translated into a photon-proton total cross section as follows:

$$\sigma_{\text{tot}}^{\gamma^*\text{p}} = \frac{4\pi^2 \alpha_{\text{em}} \hbar^2 c^2}{Q^2} F_2(x, Q^2) \tag{5.2}$$

with the CM energy given as $W^2 = Q^2(1-x)/x$ and $\hbar c$ a unit conversion factor.

It is evident from figure 7 that we undershoot data at low Q^2. At intermediate virtualities the model does a fairly good job, while at the highest virtuality probed the prediction overshoots data with roughly 50%. In order to quantify the performance of the models, a χ^2 test has been performed, taking into account the errors of the measurement:

$$\chi^2 = \sum_{i=W^2} \frac{(D[W^2] - M[W^2])^2}{\sigma_{D[W^2]}^2 + \sigma_{M[W^2]}^2} \tag{5.3}$$

with D denoting the cross section measured in data at a given squared energy W^2, M the model prediction for that squared energy and $\sigma_{D,M}^2$ the variance of the data and model predictions, respectively.

JHEP10(2019)026

Parameter	γ^*p unconfined	confined
r_0 [fm]	1.08	1.15
$r_{\max}$ [fm]	—	3.50
r_w [fm]	0.10	0.10
$r_{\max}^{\gamma^*}$ [fm]	2.07	2.56
α_s	0.21	0.22
χ^2/N_{dof} (shown Q^2 values)	2.41	0.57
χ^2/N_{dof} (full H1 data set)	2.99	1.98

Table 3. The parameter values obtained when tuning to the $\sigma_{\text{tot}}^{\gamma^*\text{p}}$ data set and the χ^2 obtained for the two models.

The model has been tuned with the PROFESSOR2 framework [76], and the parameters are shown along with the χ^2/N_{dof} in table 3. The parameters of the tune are reasonable, giving a initial dipole size roughly of order 1 fm with a width of the Gaussian fluctuations at around 0.1 fm. Adding confinement allows for a slightly larger initial dipole size, as the largest dipoles in the evolution will be suppressed as compared to the unconfined model, while also the upper integration limit on the photon is allowed to increase when turning on confinement. The width of the fluctuations and the strong coupling appear not to be affected by the confinement effect. Taking the full H1 data set into account, the confined model gives a reasonable χ^2/N_{dof}, and performs slightly better than the unconfined model.

Since the charm contribution to the γ^*p cross section has been neglected, an assessment of the uncertainty arising from this approximation should be made. Adding massless charm quarks shifts the total γ^*p cross section upwards by 67%, estimated by the ratio:

$$\frac{e_u^2 + e_d^2 + e_s^2 + e_c^2}{e_u^2 + e_d^2 + e_s^2} - 1 = 4/6. \tag{5.4}$$

This rise in cross section can be tuned away in a way similar to the procedure described above. Adding quark masses (lighter quark masses neglected) reduces the contribution. The reduction is larger for smaller Q^2. The quantitative effect of adding masses was studied in ref. [67]. For small Q^2 the decrease compared to the massless charm case is $\sim 15\%$ and for large Q^2 the decrease is $\sim 40\%$. Both represent un-tuned values. A conservative, un-tuned estimate of the uncertainty in figure 7 from neglecting (massive) charm quarks is therefore up to $\sim 25\%$. Retuning will allow for shifting the cross section upwards in the low Q^2 region where the values in figure 7 undershoots, improving the overall agreement.

5.2 Results for pp

In figure 8 we show the total cross section as a function of CM collision energy for both the full dipole model (a) and the Pascal model (b). Both figures show results using the confined (solid blue lines) and unconfined (dashed red lines) models as well as the ABMST (solid green lines) and SaS+DL model (solid magenta lines). It is evident that the full dipole

JHEP10(2019)026

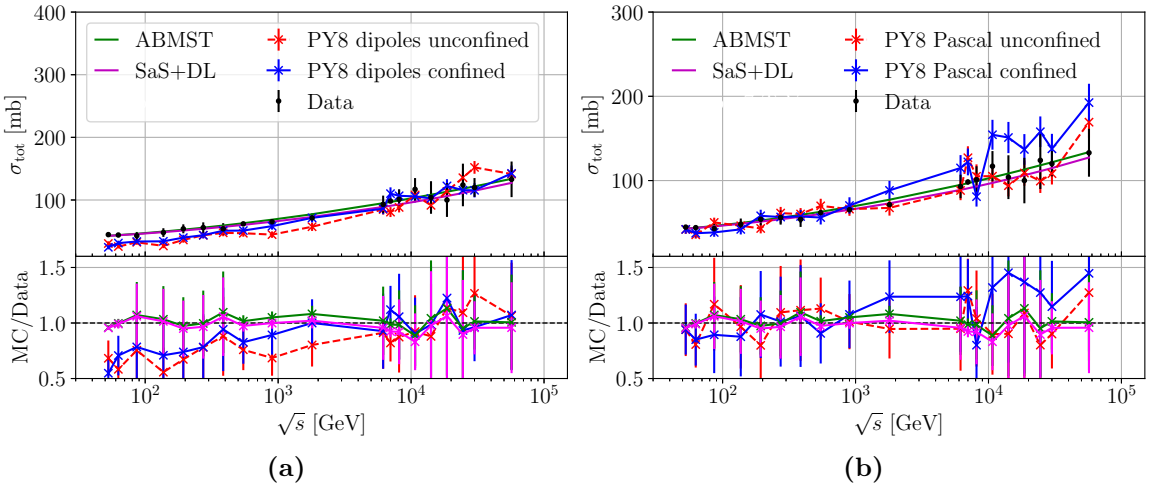


Figure 8. The total pp cross section as a function of $\sqrt{s}$ for the dipole (a) and Pascal (b) models. Both show the confined and unconfined versions in solid blue and dashed red lines, respectively. Both figures show the ABMST (solid green lines) and SaS+DL (solid magenta lines) for comparison.

model undershoots data at low $\sqrt{s}$, whereas it agrees with data at roughly $\sqrt{s} \geq 10^2$ GeV, with the confined model having a smaller χ^2/N_{dof} (cf. table 4) than the unconfined model using only this data set. The Pascal model, figure 8 (b), shows an overall shift towards higher cross sections as compared to the full dipole model, thus describing the lower energies well while slightly overshooting the higher energies. With only this data set, both Pascal models have a lower χ^2/N_{dof} than the dipole models. As explained in section 4.4, the key difference between the models, is the treatment of confinement as a hard cutoff. In both figures, it is evident that both the SaS+DL and ABMST models perform better, not surprising as these models have been created to reproduce (a subset of) this data.

In figure 9 we show the elastic pp cross section for the full dipole model (a) and the Pascal model (b). Neither of the dipole models are able to describe this cross section, being roughly 50% below data in the entire energy range, except for the very last bins, i.e. at LHC energies. The Pascal model, however, agrees with data at lower energies better than the full model. Also here, the two dedicated models describe the elastic data better than the dipole and Pascal models, with the SaS+DL model deviating from the data at LHC energies, while ABMST describes data in the entire energy range. This is partly due to a modification of the elastic slope in the SaS model, and partly due to the additional trajectories included in the ABMST model: where SaS+DL only contains a single Pomeron in the description of the elastic cross section, ABMST has two — along with additional terms not dominating at these energies. This of course introduces more freedom to the model, thus a better agreement with data at high energies.

The last result is the elastic slope at $t = 0$, shown in figure 10. Second to the total cross section, this is the most important distribution for us to describe, as this is sensitive to the internal structure of the proton, i.e. the actual impact parameter value used in the calculation, while e.g. the elastic cross section is only sensitive to the average impact parameter. Figure 10 again shows the results for the full dipole model (a) and the Pascal

JHEP10(2019)026

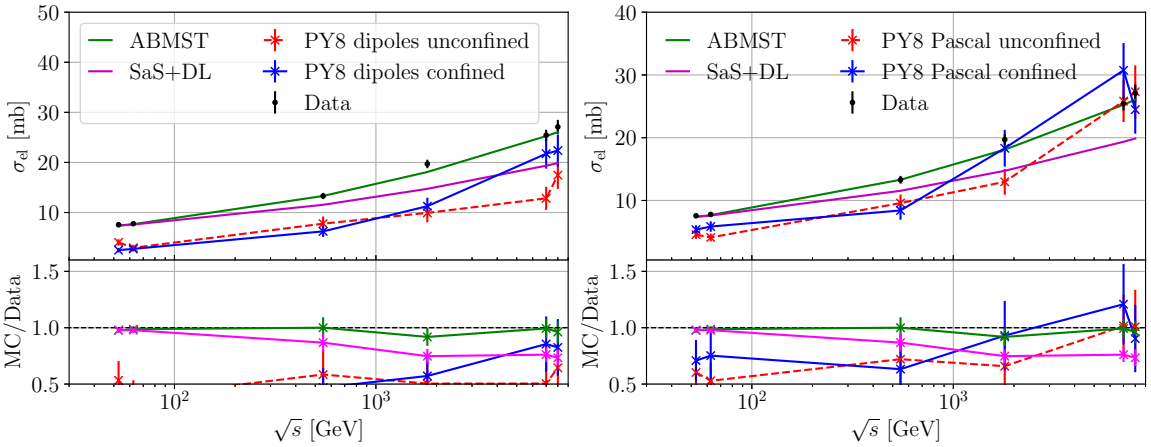


Figure 9. The elastic pp cross section as a function of $\sqrt{s}$ for the dipole (a) and Pascal (b) models. Both show the confined and unconfined versions in solid blue and dashed red lines, respectively. Both figures show the ABMST (solid green lines) and SaS+DL (solid magenta lines) for comparison.

model (b). Here, both models are undershooting the data by roughly 50% in the entire energy range, except that the dipole model is able to describe data in the very last bin. Also here, the ABMST and SaS+DL models predictions are closer to the data than the dipole and Pascal models.

We expect that the introduction of a running strong coupling will aid in the description of the data. This introduction appears in two places: in the dipole evolution and in the dipole-dipole scattering. A larger strong coupling in the evolution decreases the average step size in rapidity and increases the average size of the emitted dipoles, thus allowing for a larger number of larger dipoles at the end of the evolution. This, along with the increased dipole-dipole scattering cross section with increased strong coupling, would essentially increase all the cross sections, and thus also the elastic slope. The scale choice in such a running coupling would not be obvious, however, and we thus postpone the inclusion of a running coupling to future work.

The combined results on σ_{el}, σ_{tot} and the elastic slope deserves a further comment. From the optical theorem the differential elastic cross section is:

$$\frac{\mathrm{d}\sigma_{\text{el}}}{\mathrm{d}t} = \frac{\sigma_{\text{tot}}^2}{16\pi}(1+\rho^2). \tag{5.5}$$

Neglecting the real part of the amplitude puts $\rho = 0$. The left hand side is often approximated by an exponential: $\mathrm{d}\sigma_{\text{el}}/\mathrm{d}t = \exp(B_{\text{el}} \cdot t)$, giving $\sigma_{\text{el}} = \sigma_{\text{tot}}^2/(16\pi B_{\text{el}})$ when integrated over t. The results in figures 8, 9 and 10 are not in agreement with this simple proposition. This can either mean that the exponential approximation is not a good one (which is manifestly true for large $|t|$), or that the shape of $T(b)$ in our model simply is too narrow, such that the elastic cross section and hence B_{el} is not well described. If the latter is the case, a solution to the problem could be to include more dipoles than three in the initial proton state. This would increase the total and elastic cross sections at low $\sqrt{s}$,

JHEP10(2019)026

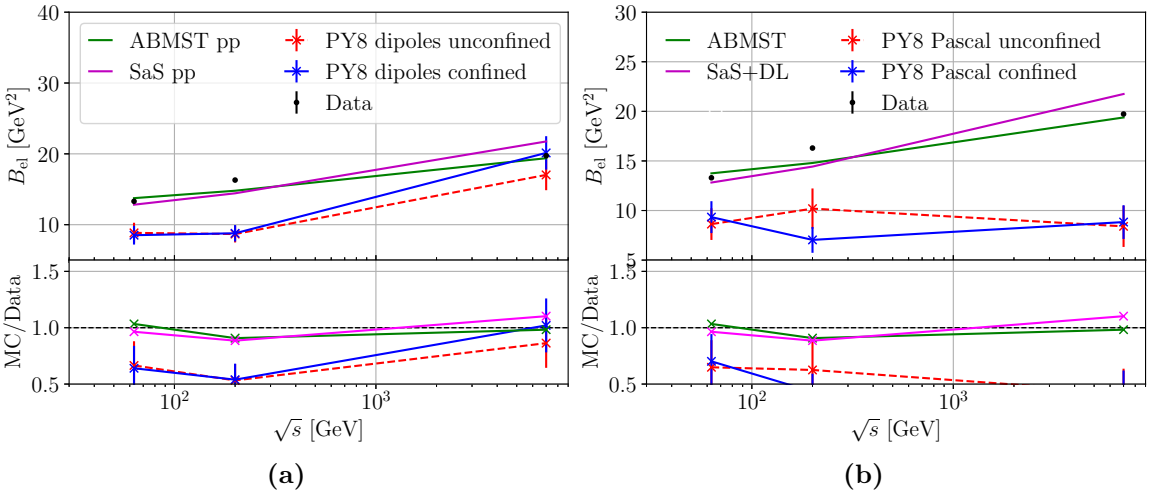


Figure 10. The elastic slope for pp collisions as a function of $\sqrt{s}$ for the dipole (a) and Pascal (b) models. Both show the confined and unconfined versions in solid blue and dashed red lines, respectively. Both figures show the ABMST (solid green lines) and SaS+DL (solid magenta lines) for comparison.

while the effect at higher $\sqrt{s}$ could be tuned away by a slightly smaller value of r_{max} or α_s. Other studies of proton substructure [66] has indicated that the number of hot-spots required for a satisfactory qualitative description of LHC data is larger than three, but at the same time, previous studies of the Mueller dipole formalism in the DIPSY event generator [56] indicated that a triangle configuration of initial dipoles is the most suitable choice for a good description of the cross sections. A study of this sort will therefore likely have to rely on attemps of simultaneous description of both sub-structure and cross sections, and will be referred to a future study.

Table 4 shows the parameters obtained when tuning to all three observables $(\sigma_{\text{tot}}, \sigma_{\text{el}}, B_{\text{el}})$ using Professor2. We also show the χ^2/N_{dof} for three data sets of various sizes. It is striking that the inclusion of the elastic cross section to the χ^2-calculation swaps the behaviour of the full dipole model — without the elastic data set, the confined model has a lower χ^2/N_{dof} than the unconfined model, while the opposite is true with the inclusion of the elastic cross section. This swap is caused by the first two data points in the elastic cross section, where the unconfined version of the full dipole describes data slightly better than the confined version. The parameters of the dipole model obtained with the tunes show the behaviour also observed in γ^*p: adding confinement allows for an increased initial dipole size and slightly larger fluctuations around this size. The initial dipole size seems reasonable for both the confined dipole model and the unconfined Pascal model, giving sizes of the order $r_0 \sim 0.7 - 1.$ fm also confirmed by Dipsy ($r_0 \sim 0.7\,\text{fm}$) and proton charge radii measurements (giving roughly $r_0 \sim 0.9\,\text{fm}$).

As already stated, the inclusion of a running strong coupling is expected to improve results for both the dipole and Pascal model. As we currently are aiming to describe proton substructure at the TeV scale, we can, however, ignore the small deviations from σ_{tot} and B_{el} at lower energies for the moment.

JHEP10(2019)026

Parameter	Full dipole model		Pascal model		ABMST	SaS+DL
	unconfined	confined	unconfined	confined		
r_0 [fm]	0.53	0.70	1.20	1.10		
$r_{\max}$ [fm]	—	3.00	—	2.05		
r_{w} [fm]	0.17	0.27	0.10	0.10		
α_s	0.24	0.22	0.15	0.16		
f_r	—	—	0.25	0.40		
Δy	—	—	2.20	2.45		
χ^2/N_{dof}: σ_{tot}	5.22	3.34	0.75	1.04	0.28	0.41
χ^2/N_{dof}: σ_{tot}, B_{el}	6.89	5.40	2.80	5.34	0.25	0.34
χ^2/N_{Dof}: σ_{tot}, B_{el}, σ_{el}	11.21	13.67	4.63	5.14	0.20	0.46

Table 4. The parameter values obtained when tuning to the $\sigma_{\mathrm{tot}}, \sigma_{\mathrm{el}}, B_{\mathrm{el}}$ data set and the χ^2 obtained for the different models.

6 Results II — eccentricities in small and large systems

In this section we turn our attention to predictions related to the geometry of an event. The parton-level eccentricities of both small and large systems are examined using the matching between the dipole model and the MPI framework described as in section 4.6. Results from the dipole model[5] are shown along with the default models of Pythia 8: in pp collisions, the default scheme of Pythia 8 is a transverse placement according to a 2D Gaussian, while for larger systems two default Pythia 8 methods are available — the usual Glauber approach and the 2D Gaussian pp model extended to larger systems. In order to compare to data, all events are hadronized with Pythia 8 after the parton-level eccentricities are calculated. Results are presented and in a single case compared to data from ALICE [77]. Parton level eccentricities were calculated with a $p_\perp$ weighting, cf. section 3.2, and events accepted if they passed the ALICE high-multiplicity trigger. Eccentricities and normalized symmetric cumulants are presented as a function of average central multiplicity ($|\eta| < 0.8$).

Recall from section 4.6 that Pythia 8 includes a $p_\perp$-dependent Gaussian smearing of parton vertices in the initial- and final-state shower. It is not clear from first principles whether such effects should be included in the calculation of geometric quantities or not. Consider, on one hand, creation of a QGP at early times, right after the collision. Here a parton shower will not be able to influence the geometry of the event, before a hydrodynamic response should be taken into account. On the other hand, one can imagine a system with large gradients (such as a small collision system) which will take time to hydrodynamize, and will therefore be influenced by geometric fluctuations from the final state shower as well. It is important to note, however, that no QGP is assumed in any results presented below as no QGP is assumed in neither Pythia 8 nor Angantyr.

[5] We do not show eccentricities calculated using the Pascal approximation, as it is, at this point, mainly intended as a toy model. If the Pascal approximation should be used for studies of eccentricities, we should point out that the large spread in daughter sizes as seen in figure 4 must be incorporated, in order to provide reasonable estimates for flow fluctuations.

JHEP10(2019)026

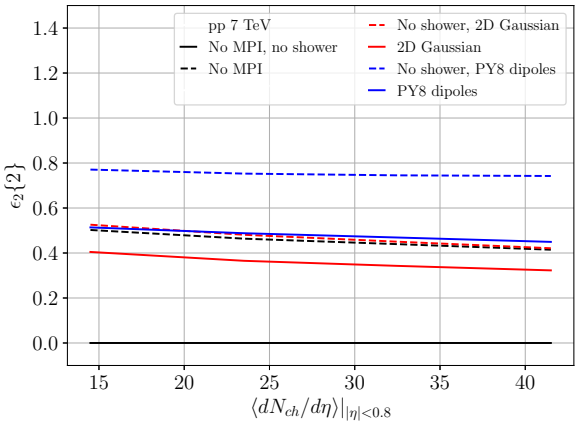


Figure 11. The eccentricities in pp collisions obtained with the several different options: No MPI vertex assignment and no shower smearing (solid black lines), no MPI vertex assignment, with shower smearing (dashed black lines), the 2D Gaussian MPI vertex with/without shower smearing (solid/dashed red lines, respectively) and the dipole model MPI vertex assignment with/without shower smearing (solid/dashed blue lines).

Opening up for a discussion, we show results in figure 11 with and without shower. It is evident that the eccentricities are vastly affected by the models. First consider the simplest case, i.e. placing all MPIs in the proton center and not introducing any shower smearing. This gives no eccentricity as expected, cf. solid black line in figure 11. Symmetric distributions, such as the 2D Gaussian shower smearing and the MPI vertex assignment, should in principle give no eccentricity. But, as we are sampling only a finite number of MPIs from such symmetric distributions, an eccentricity does appear for these models, cf. dashed red and dashed black lines of figure 11. The two methods overlap, thus the exact same effect can be introduced with either (a) no MPI vertex assignment with a Gaussian smearing from the shower or (b) a Gaussian MPI vertex assignment and no shower smearing. That the two overlap is not so surprising as both are Gaussian smearings, and applying such a smearing during the shower or assigning it to the MPIs should make no difference: both methods give rise to a more lenticular overlap region of the two colliding protons.

Applying Gaussian smearing twice, i.e. both in the MPI vertex assignment and during the shower smears the lenticular shape from the MPI assignment slightly, thus causing the eccentricity to drop, cf. the solid red line in figure 11. The largest effect on the eccentricity is seen when purely considering MPI vertex assignment with the dipole model, cf. the dashed blue line in figure 11. The eccentricity with the dipole model is approximately twice as large as with the Gaussian model, thus indicating that event-by-event asymmetries in the initial state gives rise to larger fluctuations and thus larger eccentricities. Adding the Gaussian shower smearing on top of the dipole model, solid blue line of figure 11, washes out some of these features — i.e. makes the almond shape of the overlap region rounder.

Figure 12 shows the eccentricities $\epsilon_{2,3}\{2\}$ in three different systems. Beginning with ϵ_2 we observe in pp that the dipole evolution gives rise to a larger eccentricity than the 2D-Gaussian. In the dipole evolution, the asymmetry is built in at the cascade level, where in the 2D-Gaussian, where MPIs are sampled from a symmetric distribution, asymmetry only

JHEP10(2019)026

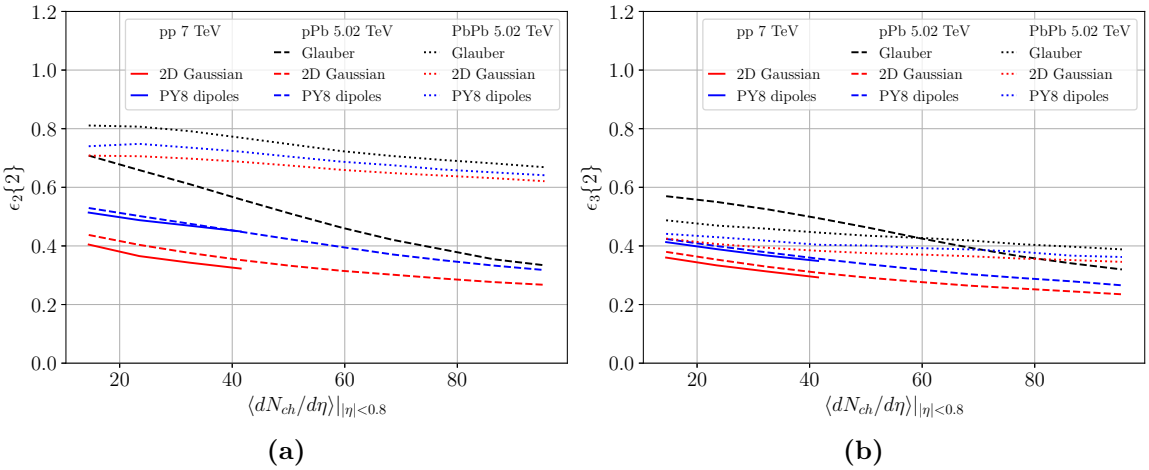


Figure 12. The second (b) and third (c) order eccentricities using two-particle cumulants for pp, pPb, PbPb collisions (solid, dashed, dotted, respectively) using the Glauber (black), Gaussian (red) and dipole (blue) models.

arises due to the sample size. Proceeding to larger systems, pPb and PbPb, it is evident that the same trend is seen: the dipole model gives rise to larger ϵ_2 than the symmetric model. The Glauber model, however, is consistently larger than the other two models at low multiplicity, while all three models appear to approach the same eccentricity at higher multiplicities. Thus it becomes evident that the proton initial state does have an effect on eccentricities, and that it is especially evident in low-multiplicity events, e.g. peripheral PbPb collisions.

Unfortunately, the low-multiplicity events are often marred by large non-flow effects. Measuring the eccentricities with higher-order cumulants can remove some of the contributions from non-flow, thus making it easier to compare data to models. We present results for ϵ_2 with higher-order cumulants in appendix C, as the results are similar in shape as figure 12, but differ in normalisation. Figure 12 (b) show $\epsilon_3\{2\}$ for all three systems. Here, it becomes more difficult to distinguish between the models in symmetric systems, while a large discrepancy between the Glauber approach and the other two is seen in pPb.

Another feature seen in figure 12 is that the dipole model gives roughly the same results for $\epsilon_{2,3}$ in both pp and pPb collisions. If one assumes that the response functions are the same for the two systems (however one may have obtained these response functions, QGP or by string-string interactions), the ratio of pPb to pp eccentricities should thus be comparable to the ratio of flow coefficients measured with the ALICE detector. This ratio is shown in figure 13 for the second-order eccentricity. Both the Gaussian and dipole models are compatible with the ALICE data, however, so we cannot presently discriminate between the two. Additional measurements of the flow coefficients in low-multiplicity events are thus required in order to discriminate between models in this observable.

Figure 13 (b) shows the normalized symmetric cumulant, $NSC(3,2)$. This has been constructed to study the correlations between the eccentricities and normalized to the uncorrelated eccentricities in order to remove the effects of the response function. ALICE

JHEP10(2019)026

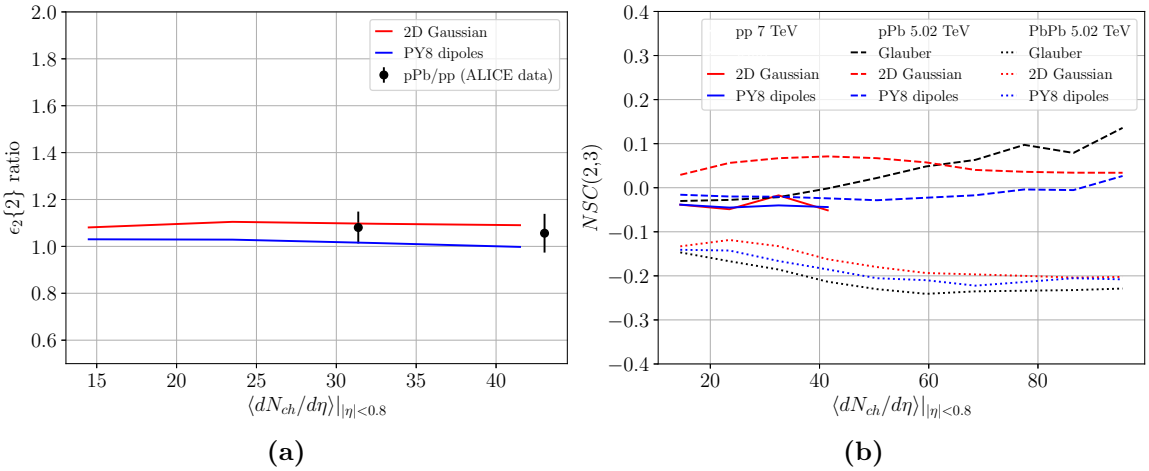


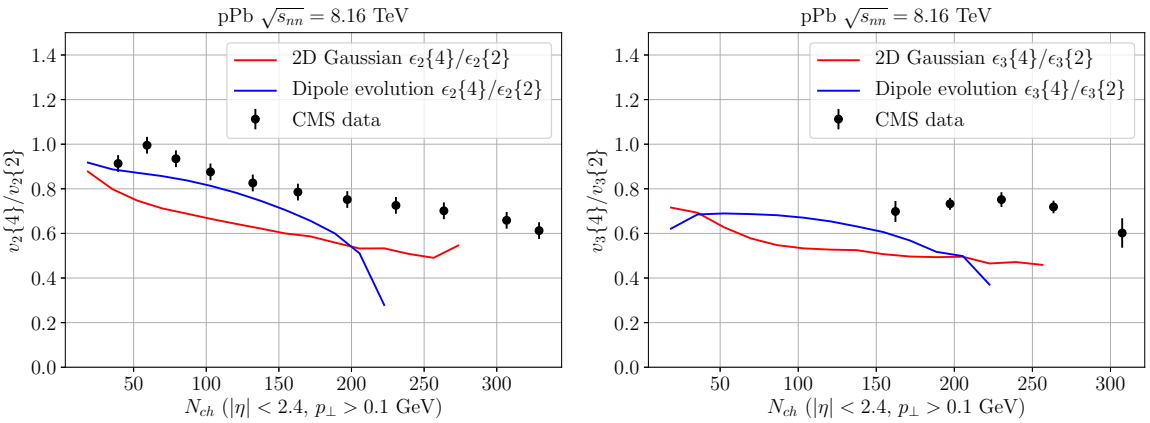
Figure 13. (a) The ratio of second order eccentricities obtained in pPb w.r.t. the baseline pp sample using the Gaussian (red) and dipole (blue) models. Data points calculated from ALICE figures [77]. (b) The normalized symmetric cumulants $NSC(3,2)$ for pp, pPb, PbPb collisions (solid, dashed, dotted, respectively) using the Glauber (black), Gaussian (red) and dipole (blue) models.

reports that all three systems have the same $NSC(3,2)$ at the same average multiplicity, indicating that the correlation between the flow coefficients are the same in different collision systems. We observe no such effect. Focusing on the dipole model, the correlations appear equal in magnitude for pp and pPb, but PbPb results are consistently below the smaller systems. Results for the Gaussian model shows no similarities at all between systems, as the pPb $NSC(3,2)$ is positive, while pp and PbPb are negative. Thus the normalized symmetric cumulants for pPb systems would be an ideal place to discriminate between the symmetric and asymmetric initial state. PbPb results for all three models are in agreement with IP-Glasma predictions presented in the ALICE paper. The main difference between the dipole model and the IP-Glasma approach is the inclusion of saturation in the cascade of the latter. As the two approaches give similar results, we do not find that saturation plays a large role in this observable.

6.1 Flow fluctuations in pPb collisions

Recently, CMS presented results on multi-particle correlations using higher-order particle cumulants in pPb collisions [78]. Ratios of the flow-coefficients based on these cumulants were presented, including the first measurements of the ratio of $v_3\{4\}/v_3\{2\}$ in pPb. In figure 14 we show the predictions for the ratios with the confined dipole model and the default Gaussian model as a function of multiplicity. Both models reasonably reproduce the shape seen in the elliptical ratio, figure 14 (a) showing $v_2\{4\}/v_2\{2\}$, while the normalisation of the dipole model is slightly better than with the Gaussian model. For the triangular ratio, figure 14 (b), both models appear to undershoot data at high multiplicities, where data is available. As opposed to model predictions presented in the CMS paper [79, 80], our predictions have *not* been applied a 10% *ad hoc* increase in normalisation. And where the model predictions presented in [79, 80] predicts roughly the same ratio for both ϵ_2 and

JHEP10(2019)026



(a) **(b)**

Figure 14. Ratios of $v_n\{4\}/v_n\{2\}$ with $n = 2$ (a) and $n = 3$ (b) as measured by CMS as function of multiplicity in pPb collisions, compared to eccentricity ratios calculated with the Gaussian and the dipole models.

ϵ_3, neither the dipole nor the Gaussian model predicts the same normalisation for the two ratios, cf. the height of figure 14 (a) and (b) differs.

Figure 15 shows the higher-order cumulant ratios for elliptic flow as a function of the lower-order ratio presented in figure 14 (a). For higher order cumulants, the Gaussian model predicts purely imaginary values for even powers of the cumulants, hence it has been left out of the figures. The dipole model, however, is able to describe data reasonably well. The dipole predictions decrease with decreasing $v_2\{4\}/v_2\{2\}$ ratio in figure 15 (a), while being roughly constant at unity in figure 15 (b). This is in accordance with the model predictions presented by CMS [81], assuming a non-Gaussian model for the initial state. We note that the eccentricities presented with the dipole model here are (a) based on a pQCD model, and (b) related to final state multiplicities calculated in the same acceptance as the experiment.

7 Results III — dynamic colour fluctuations in Glauber calculations

A general feature of several models describing both collisions of protons and of nuclei, is the notion of interacting nucleons and nuclear sub-collisions, calculated in the formalism of Glauber [82, 83]. The basic formalism is mainly concerned with calculating the full AA scattering matrix or amplitude from knowledge of the nucleon-nucleon amplitude and spatial positions. Multiple interactions between nucleons factorize in transverse coordinates, so in the eikonal limit the S-matrix for scattering between two nuclei A and B becomes:

$$\hat{S}^{(AB)}(\vec{b}) = \prod_{i=1}^{A}\prod_{j=1}^{B}\hat{S}^{(N_iN_j)}(\vec{b}_{ij}), \tag{7.1}$$

where i and j denote the individual nucleons, $\vec{b}$ is the nucleus-nucleus impact parameter and $\vec{b}_{ij}$ is the nucleon-nucleon impact parameter. We will here consider the simplifying

JHEP10(2019)026

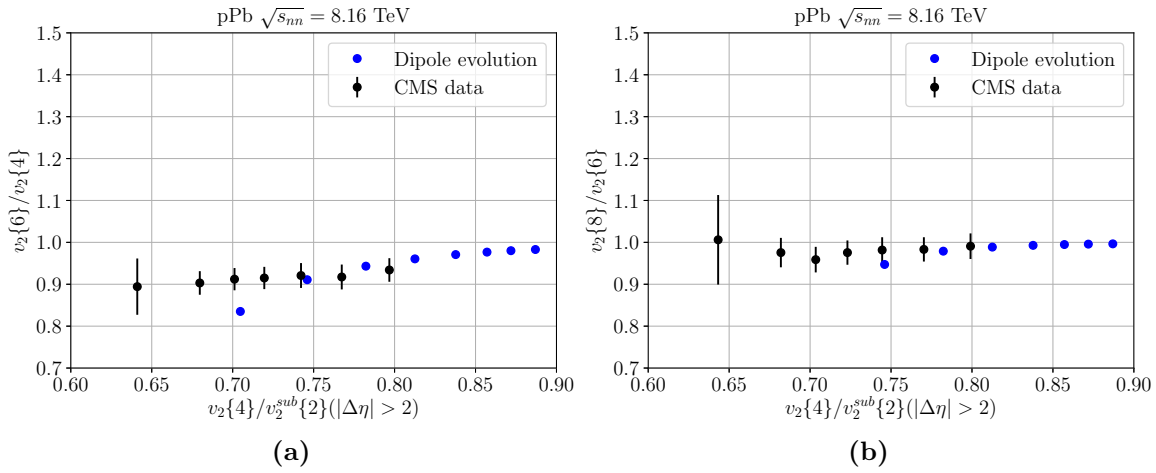


Figure 15. Correlations between higher order flow harmonics as measured by the CMS experiment, compared to correlations between higher order eccentricity ratios calculated in the dipole model.

case where only one projectile (either p or γ^*, called n below) collides with a nucleus (A), which reduces eq. (7.1) considerably:

$$\hat{T}^{(nA)}(\vec{b}) = 1 - \prod_{i=1}^{A} \hat{S}^{(nN_i)}(\vec{b}_{ni}) = 1 - \prod_{i=1}^{A} (1 - \hat{T}^{(nN_i)}(\vec{b}_{ni})). \tag{7.2}$$

If no fluctuations in the interaction are included, the projectile-nucleon elastic amplitude can be inserted, and the total and elastic cross sections can be calculated directly from eqs. (3.4)–(3.5). If fluctuations for projectile and target are included, as calculated for example in the dipole model, the amplitude will depend on the states of target (t_i) and projectile (p) respectively. As shown in section 3.1, the elastic amplitude can be calculated as an average over all states. In ref. [84] it was pointed out that in the evaluation of such an average, the projectile must remain frozen in the same state throughout the passage of the target. Similar to eq. (3.3) the elastic profile function (at fixed Mandelstam s) for a fixed state (k) of the projectile scattered on a single target nucleon (all states) becomes:

$$\begin{aligned}\Gamma_k(\vec{b}) &= \langle \psi_S | \psi_I \rangle = \langle \psi_k, \psi_t | \hat{T}(\vec{b}) | \psi_k, \psi_t \rangle = (c_k)^2 \sum_t |c_t|^2 T_{tk}(\vec{b}) \langle \psi_k, \psi_t | \psi_k, \psi_t \rangle \\ &= (c_k)^2 \sum_t |c_t|^2 T_{tk}(\vec{b}) \equiv \langle T_{tk}(\vec{b}) \rangle_t,\end{aligned} \tag{7.3}$$

where previously suppressed indices k and t on T are spelled out for clarity. For a projectile-nucleus collision, with the projectile frozen in the state k, the relevant projectile-nucleon (nN_i) amplitude becomes:

$$\langle T^{(nN_i)}_{t_i,k}(\vec{b}_{ni}) \rangle_t \equiv T^{(nN_i)}_{k}(\vec{b}_{ni}). \tag{7.4}$$

In the short hand notation on the right hand side, the average over the repeated index t is suppressed. This is the amplitude used to determine which nucleons are "wounded" in a collision. If the purpose is to determine which nucleons participate in the collision either elastically or inelastically, the differential wounded cross section can be calculated with the

JHEP10(2019)026

normal differential pp total cross section as an ansatz, $\mathrm{d}\sigma_{\mathrm{tot}}/\mathrm{d}^2\vec{b} = 2\langle T\rangle_{p,t}$ from eq. (3.4). Since the projectile should be frozen in the state k, the expression for T from eq. (7.4) is inserted to the differential pp total cross section. This just recovers the normal total projectile-nucleon cross section:

$$\frac{\mathrm{d}\sigma_{\mathrm{tot}}}{\mathrm{d}^2\vec{b}} = 2\langle T_k\rangle_p = 2\langle\langle T_{t,k}\rangle_p\rangle_t = 2\langle T\rangle_{p,t}. \tag{7.5}$$

In a Monte Carlo, the number of wounded nucleons can then be generated by assigning each projectile or nucleon a radius of $\sqrt{\sigma_{\mathrm{tot}}/2\pi}$, where the expression in eq. (7.5) has been integrated over $\mathrm{d}^2\vec{b}$ to give σ_{tot}. Normally one is not interested in the number of wounded nucleons including elastic interaction, but rather those that contribute to particle production (i.e. where there is a colour exchange). A usual approach is to just use the inelastic cross section in place of σ_{tot} in the Monte Carlo recipe. This does, however, not account fully for colour fluctuations, as the inelatic cross section is modified when averaging over target states with a frozen projectile. Instead of directly using the inelastic cross section in the Monte Carlo, the modified cross section should be used. This cross section was dubbed the "wounded cross section" in ref. [35], and can be constructed by generalizing the inelastic cross section, using eq. (7.4). The inelastic cross section can from eqs. (3.4)–(3.5) be directly written down as:

$$\frac{\mathrm{d}\sigma_{\mathrm{inel}}}{\mathrm{d}^2\vec{b}} = 2\langle T(\vec{b})\rangle_{p,t} - \langle T(\vec{b})\rangle^2_{p,t}. \tag{7.6}$$

When the frozen projectile is taken into account by inserting T from eq. (7.4), the usual expression is now not recovered, but the average over targets must be made *before* squaring the second term:

$$\frac{\mathrm{d}\sigma_w}{\mathrm{d}^2\vec{b}} = 2\langle T_k(\vec{b})\rangle_p - \langle T_k^2(\vec{b})\rangle_p = 2\langle T(\vec{b})\rangle_{t,p} - \langle\langle T(\vec{b})\rangle_t^2\rangle_p, \tag{7.7}$$

with internal indices again suppressed in the last equality. In a Monte Carlo this can be generated as above, now only by inserting σ_w in place of σ_{tot}.

Generalizing this procedure to γ^*A collisions requires additional considerations. Starting from the elastic profile function for γ^*p, a contribution from the photon fluctuating to a dipole state must be included. Examining only the hadronic (non-VMD) components of the photon state, gives:

$$|\gamma^*\rangle \sim c_1|\mathrm{q\bar{q}}\rangle + c_2|\mathrm{q\bar{q}g}\rangle + \text{higher order Fock states} \tag{7.8}$$

where quark helicities have been neglected. We keep only the first (leading order) term, as the higher order Fock states are included in the dipole evolution. Thus with a photon wave function given in eqs. (4.1)–(4.2), we obtain:

$$|\gamma^*\rangle = \int \mathrm{d}z \int \mathrm{d}^2\vec{r}\left(|\psi_L(z,r)|^2 + |\psi_T(z,r)|^2\right)|\psi_I(r,z)\rangle, \tag{7.9}$$

JHEP10(2019)026

with $|\psi_I\rangle$ a dipole state. The elastic profile is now:

$$\begin{aligned}\Gamma_{\text{el}}(\vec{b}) &= \int \mathrm{d}z \int \mathrm{d}^2\vec{r}\, \langle\psi_S(z,\vec{r})|\hat{T}(\vec{b})|\psi_I(z,\vec{r})\rangle\langle\psi_I(z,\vec{r})|\gamma^*\rangle \\ &= \int \mathrm{d}z \int \mathrm{d}^2\vec{r}\, (|\psi_L(z,\vec{r})|^2 + |\psi_T(z,\vec{r})|^2)\langle T(\vec{b})\rangle_{p,t}. \end{aligned} \tag{7.10}$$

The wounded cross section for γ^*A collisions can now be defined. The first interaction is calculated using the photon wave function in the elastic profile function, leading directly to:

$$\frac{\mathrm{d}\sigma_w}{\mathrm{d}^2\vec{b}} = \int \mathrm{d}z \int \mathrm{d}^2\vec{r}\, (|\psi_L(z,\vec{r})|^2 + |\psi_T(z,\vec{r})|^2)(2\langle T(\vec{b})\rangle_{t,p} - \langle\langle T(\vec{b})\rangle_t^2\rangle_p). \tag{7.11}$$

This first interaction has now turned to photon from a superposition of all dipole states into a single, specific dipole (or vector meson). *This* is the state that the projectile should be frozen to throughout the passage of the nucleus: the first interaction chooses a specific dipole state $|\psi_I\rangle_{z,\vec{r}}$ with given z and $\vec{r}$. This reduces the elastic profile function for the secondary interactions to the well known eq. (3.3), from which a differential wounded cross section has already been calculated (eq. (7.7)).

Thus, in a Monte Carlo, the number of wounded nucleons can be generated with the following method:

- First by selecting, for each event, a dipole with r and z corresponding to the wave function weight, w_γ in eq. (A.35)
- Secondly, testing if any nucleons are hit including the photon wave function normalization proportional to α_{em} (i.e. according to eq. (7.10))
- If any nucleons are hit, then subsequently testing all (other) nucleons, w.r.t. the dipole-target weight (i.e. eq. (3.3))

In the following section, colour fluctuations from the introduced dipole model (where $T(\vec{b})$ can be evaluated directly from eq. (2.7)) are compared to a parameterized approach for fluctuations in pp collisions and γ^*p collisions, and finally for γ^*A.

7.1 Colour fluctuations in pp, γ^*p and γ^*A collisions

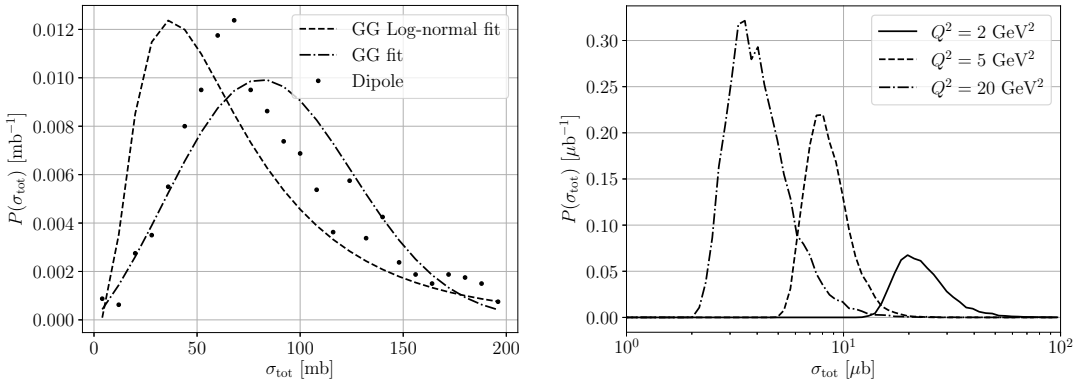
Fluctuations in the pp cross sections, to estimate the influence of fluctuations in pA collisions, are often parametrized using [85–87]:

$$P(\sigma) = \rho \frac{\sigma}{\sigma + \sigma_0} \exp\left(-\frac{(\sigma/\sigma_0 - 1)^2}{\Omega^2}\right), \tag{7.12}$$

where σ_0 and Ω are parameters, and ρ is a normalization constant. In ref. [35] is was found that a log-normal distribution (see eq. (B.2) in appendix B) describes fluctuations generated by a dipole approach better. In figure 16 (a) both parametrizations are compared to the fluctuating total cross section in pp at $\sqrt{s} = 5\,\text{TeV}$, integrated over $\mathrm{d}^2\vec{b}$.

While the log-normal distribution does better in capturing the skewness of the distribution, none of the two parametrizations fully describes the distribution. The problem

JHEP10(2019)026



(a) **(b)**

Figure 16. (a) Fluctuating cross section in pp at $\sqrt{s} = 5\,\text{TeV}$, compared to a GG fit (eq. (7.12)) and a log-normal fit (eq. (B.2)). (b) Fluctuating cross section in γ^*p at $W^2 = 5000\,\text{GeV}^2$ and various Q^2, calculated with the dipole model (the double peak structure for $Q^2 = 20\,\text{GeV}^2$ is a statistical fluctuation). The cross section is shown on a logarithmic horizontal axis, to assess the log-normal approximation (cf. eq. (B.2)).

increases in γ^*p for several reasons. First of all, any parametrization must include the correct dependence on DIS kinematics, which changes the average cross section, cf. figure 16 (b). Here is shown the cross section distributions for three values of Q^2 all with $W^2 = 5000\,\text{GeV}^2$ with a logarithmic first axis. This allows for a by-eye assessment of the validity of a log-normal fit, as a log-normal distribution is Gaussian with such choice of axes. It is seen directly that fluctuations in the high-σ tails are too large to be described by such a parametrization.

Instead of the parametrization approach, the wounded cross section can be calculated directly from the dipole evolution. This also allows for simultaneous calculation of both the part including electromagnetic contributions, and the pure dipole part (given z and r), as introduced in the previous section. This allows directly for a calculation of the distribution of wounded nucleons in a γ^*A collision, as shown in figure 17. In the figure, the nucleus is taken to be Au-197, colliding with a virtual photon with $W^2 = 5000\,\text{GeV}^2$ for a range of Q^2 values, compatible with projected EIC design [37]. Two methods of calculating wounded nucleons are presented: the full treatment using a frozen wave function, where the photon wave function has collapsed to a dipole state when probed by the first collision, and the naive black disk approach, where the photon re-forms after the full collision and no fluctuations are included. In such a treatment, the cross section for additional collisions has an additional factor $\alpha_{\rm em}^2$ compared to the frozen treatment, from the normalization of the wave function. It is directly visible that a full treatment is necessary in order to provide reasonable phenomenological projections for a new collider.

8 Conclusion and outlook

One of the main challenges for the understanding of collective effects, is to grasp how the well-known understanding of flow results from heavy ion collisions can be transferred to

JHEP10(2019)026

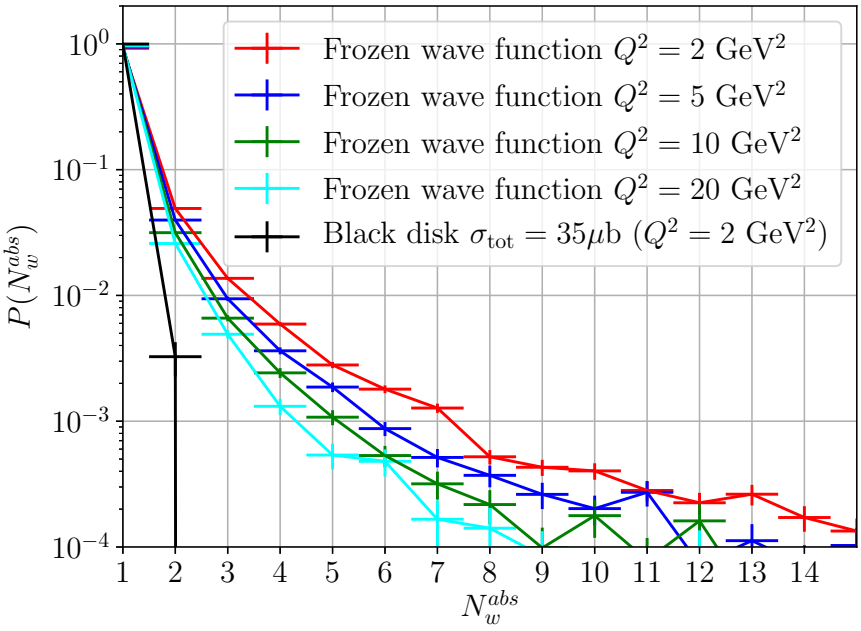


Figure 17. Number of wounded nucleons in a γ^*Au collision with $W^2 = 5000\,\mathrm{GeV}^2$ and a range of Q^2, comparing a treatment with the projectile wave function (denoted wf. in legend) frozen throughout the passage of the nucleus, to a naive black disk approach.

collision of protons with protons and heavy nuclei. In this paper we have presented a Monte Carlo implementation of Mueller's dipole model with several sub-leading corrections, and with all parameters of the model fixed to total and semi-inclusive cross sections calculated within the Good-Walker formalism. This model thus allows for the calculation of proton substructure without tuning any model parameters to observables sensitive to said substructure.

The current implementation of the model includes:

- BFKL evolution of projectile and target states, be it protons or photons, in rapidity and impact parameter space.
- Sub-leading corrections in the evolution:
  1. Energy-momentum conservation.
  2. Non-eikonal corrections in terms of dipole recoils.
  3. Confinement effects by the introduction of a fictitious gluon mass.
- Projectile-target interactions using the unitarized amplitude, which in a Regge field theory language corresponds to multi-Pomeron exchange and Pomeron loops.
- Matching to the Pythia 8 MPI model, in order to assign spatial vertices to produced partons in pp collisions.
- Generalization to heavy ion collisions through the Angantyr framework.

JHEP10(2019)026

Besides the implementation of the dipole model, a simpler version has been provided, based on the geometric properties of the dipole evolution. This model, denoted the Pascal approximation, allows for easy insertion of toy-models of sub-leading effects, thus giving a handle on the importance of such effects.

We have shown that given simple, but reasonable, assumptions of a final-state response (from e.g. hydrodynamics or interacting strings), the eccentricities produced with the implementation provides a reasonable description of flow data from the ALICE and CMS experiments. This includes non-trivial observations such as ratios between pA and pp flow coefficients at fixed event multiplicity, normalized symmetric cumulants in different systems, and ratios between different order flow harmonics in pA collisions. All are signatures which cannot be described in a simpler model, where the spatial structure of MPIs are assumed to be distributed according to a rotationally symmetric distribution. We want to stress that even though we have here chosen flow-type observables to illustrate the effect of the space-time structure of the initial state on observations of collective effects, effects linked to enhancement of strangeness and baryon production [88–90] and even modifications of jets in high multiplicity pp collisions [91, 92] are expected to be influenced as well.

Lastly, we have provided the initial steps towards the generation of fully exclusive final states in electron-ion collisions, by determining the importance of colour-fluctuations in the collisions with virtual photons. We have shown that previous parametrizations from pp collisions do not fully capture the colour-fluctuations predicted by the dipole formulation of BFKL evolution, and thus argue that it is better to calculate the cross sections directly from the dipole model — which has not been possible in the Angantyr model before this work. Secondly, we stress that the collapse of the photon wave function at first interaction provides a larger number of wounded nucleons as compared to the black disk approximation. Each of the wounded nucleons are expected to give rise to final state activity, thus more complicated final states are expected with the proper treatment as opposed to the naive expectations.

The implemented dipole model can be improved in several ways, including:

- Running α_s in the dipole evolution and in the scattering, which will capture some of the NLO corrections in α_s.
- On longer term, an inclusion of full NLO-BK should be the goal, though further theoretical development is needed first.
- Gluon saturation effects in the cascade such as those included in the CGC formalism. To maintain the current treatment of the effect of gluon branchings in the cascade (as opposed to CGC), this could be included by the introduction of a simple swing mechanism, e.g. a mock $2 \to 1$ dipole recombination.
- Several improvements are expected w.r.t. the initial dipole configuration in protons and photons, as well as in the wave functions of these particles. This includes adding the VMD contribution to the photon wave function, to be able to study lower Q^2 and vector meson production in various processes.

JHEP10(2019)026

- New ways of treating MPIs in pp collisions by fully merging the dipole approach with more traditional approaches are foreseen. It is our hope that this could provide new tools to improve understanding of particle production mechanisms across collision systems.

Detailed understanding of the interplay between the proton geometry and the response of final state interactions in hadronic and heavy ion collisions, is crucial for the understanding of collectivity and particle production mechanisms. Since detailed understanding requires tools which are both accessible and transparent, it is our hope that the detailed treatment presented here, and the accompanying open Monte Carlo implementation, can help facilitate this process.

Acknowledgments

We thank Gösta Gustafson, Leif Lönnblad and Torbjörn Sjöstrand for several discussions during the making of this project and Gösta, Torbjörn and Helen Brooks for reading through and commenting on the manuscript. We also thank the ALICE group at the Niels Bohr Institute, University of Copenhagen for hospitality during the writing of this work.

This project has received funding in part from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 668679), and in part by the Swedish Research Council, contract numbers 2016-05996 and 2017-0034 as well as the Marie Skłodowska-Curie Innovative Training Network MCnetITN3 (grant agreement 722104).

A The dipole model

Below we go through the details of the dipole model not included in sections 2.2 and 4.

We here work with light cone momenta,

$$p_\pm = E \pm p_z, \tag{A.1}$$

and can thus define the rapidity as

$$y = \frac{1}{2}\log\left(\frac{p_+}{p_-}\right) = \log\left(\frac{p_+}{p_\perp}\right), \tag{A.2}$$

with the latter equality valid for massless particles. Hence we can express the lightcone momenta in terms of dipole $p_\perp$ and rapidity,

$$p_\pm = p_\perp \exp(\pm y). \tag{A.3}$$

The $p_\perp$ of a dipole can be related to its size through $p_\perp \sim \hbar/r$.

The dipole-dipole scatterings are defined to occur at rapidity zero. Thus the evolution of the beams begin at rapidity $y = \pm y_{\max}$ and evolve to zero, i.e. with negative rapidity steps. For technical reasons, the actual evolution is easier to implement with positive steps

JHEP10(2019)026

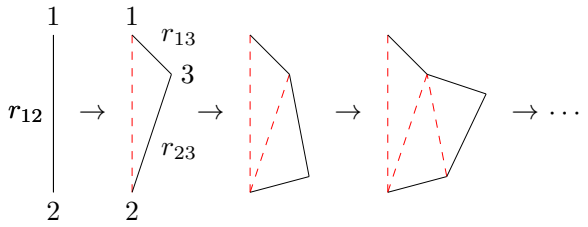


Figure 18. Schematic view of a dipole splitting. The initial dipole is spanned by partons 1 and 2, that emits a new parton (3), thus creating two new dipoles: the dipole spanned by partons 1 and 3 and the dipole spanned by partons 2 and 3. This can be succeeded by additional splittings as indicated by the additional figures following the arrows.

in rapidity. Thus the internal rapidity used in the code (and in the next section) is negated w.r.t. the rapidity defined in eq. (A.2):

$$y_{\mathrm{MC}} = -y = \log\left(\frac{p_\perp}{p_+}\right) \Rightarrow \tag{A.4}$$

$$p_\pm = p_\perp \exp(\mp y_{\mathrm{MC}}), \tag{A.5}$$

where in the forthcoming sections we will skip the subscript MC.

A.1 Mueller's dipole branching

We begin by examining the dipole splitting function,

$$\frac{\mathrm{d}P}{\mathrm{d}y\ \mathrm{d}^2\vec{r}} = \frac{N_c\alpha_s}{2\pi^2}\frac{r_{12}^2}{r_{13}^2 r_{23}^2}, \tag{A.6}$$

where $\vec{r}$ is the transverse location of the emitted parton 3 from the original dipole spanned by partons $1, 2$ and r_{ij} the length of the dipoles, also shown in figure 18. In order to turn this into a dipole evolution, a Sudakov factor, $\Delta(y_{\min}, y)$, restricting emission between $y_{\min}$ and y, has to be introduced. The full dipole splitting kernel then reads,

$$\frac{\mathrm{d}P}{\mathrm{d}y\ \mathrm{d}^2\vec{r}} = \frac{N_c\alpha_s}{2\pi^2}\frac{r_{12}^2}{r_{13}^2 r_{23}^2}\Delta(y_{\min}, y). \tag{A.7}$$

For event generation to proceed, we need to find an overestimate for the above splitting probability. The Sudakov factor is included via the veto algorithm, and is thus neglected in the expressions below.

First we sample a transverse location of the emitting dipole. Assuming partons 1 and 2 located in the (x, y)-plane at $\vec{r}_1 = (0, 0)$ and $\vec{r}_2 = (1, 0)$ with length $r_{12} = 1\,\mathrm{fm}$, while the emitted parton is located at $\vec{r}_3 = (r_x, r_y)$, we can write the splitting probability as,

$$\begin{aligned}\frac{\mathrm{d}P}{\mathrm{d}y\ \mathrm{d}r_x\ \mathrm{d}r_y} &= \frac{N_c\alpha_s}{2\pi^2}\frac{r_{12}^2}{(r_x^2 + r_y^2)((r_{2,x} - r_x)^2 + r_y^2)} \\ &= \frac{N_c\alpha_s}{2\pi^2}\frac{1}{(r_x^2 + r_y^2)((1 - r_x)^2 + r_y^2)}\,,\end{aligned} \tag{A.8}$$

JHEP10(2019)026

where in the second step we have inserted the values for $r_{2,x} = 1\,\text{fm}$ and $r_{12}^2 = 1\,\text{fm}^2$, but suppressed dimensions. These dimensions are suppressed throughout the section. This distribution is symmetric around $r_x = 1/2\,\text{fm}$ and $r_y = 0\,\text{fm}$, so the limits of integration can be changed from $r_{x/y} \in]-\infty, \infty[$ to $r_x \in [-\infty, 1/2]$ and $r_y \in [-\infty, 0]$.

The above splitting probability can be overestimated by the function,

$$\frac{\mathrm{d}P_1}{\mathrm{d}y\ \mathrm{d}r_x\ \mathrm{d}r_y} = \frac{N_c\alpha_s}{2\pi^2}\frac{2}{(r_x^2+r_y^2)(r_x^2+r_y^2+\frac{1}{4})}. \tag{A.9}$$

Changing to cylindrical coordinates we obtain,

$$\frac{\mathrm{d}P_1}{\mathrm{d}y\ r\ \mathrm{d}r\ \mathrm{d}\phi} = \frac{N_c\alpha_s}{2\pi^2}\frac{2}{r^2(r^2+\frac{1}{4})}, \tag{A.10}$$

which can be integrated from a minimal dipole size, ρ. Thus we obtain,

$$\frac{\mathrm{d}P_1}{\mathrm{d}y} = \frac{4N_c\alpha_s}{\pi}\log\left(1+\frac{1}{4\rho^2}\right). \tag{A.11}$$

Without energy ordering, the minimal dipole size ρ has to be fixed to a number larger than zero to avoid the distribution from blowing up. Here, energy the ordering is introduced by ordering of positive lightcone momenta [39]. Again relating the transverse momentum of the dipole to its size, gives an expression for ρ related to the kinematics of the parent dipole (p),

$$p_+^3 \leq p_+^p \Rightarrow \quad p_\perp^3 e^{-y} = \frac{1}{\rho}e^{-y} \leq p_+^p \Rightarrow \quad \rho \geq e^{-y}/p_+^p. \tag{A.12}$$

This expression is then used in eq. (A.11),

$$\frac{\mathrm{d}P_1}{\mathrm{d}y} = \frac{4N_c\alpha_s}{\pi}\log\left(1+\frac{(p_\perp^p)^2}{4}e^{2y}\right). \tag{A.13}$$

This overestimate cannot trivially be integrated, so we find yet another,

$$\frac{\mathrm{d}P_2}{\mathrm{d}y} = \frac{4N_c\alpha_s}{\pi}\log\left[e^{2y}\left(1+\frac{(p_\perp^p)^2}{4}\right)\right] \equiv \frac{4N_c\alpha_s}{\pi}\log\left[e^{2y}A\right] \tag{A.14}$$

which is both integrable and invertible. We take into account the Sudakov factor by using the Veto algorithm, and thus the rapidity can be sampled from this distribution by

$$y_i = 1/2\sqrt{\left[\log(A)+2y_{i-1}\right]^2 - \pi\log(R_1)/(N_c\alpha_s)} - 1/2\log(A), \tag{A.15}$$

where R_1 is a uniformly distributed random number.

From eq. (A.11) we can sample both r and ϕ,

$$\phi = 2\pi R_2, \tag{A.16}$$

$$r = \sqrt{\frac{1}{4}\frac{\rho^{2R_3}}{(\rho^2+1/4)^{R_3}-\rho^{2R_3}}}, \tag{A.17}$$

with R_2, R_3 two new random numbers. Here we should note that we've changed the integration limits, such that any $r_x = r\cos(\phi) > 1/2$ must be rejected in the event generation. Half of the remaining events should be mirrored to $r_x \to 1 - r_x$, and this should be taken into account in the overestimate $\mathrm{d}P_1/\mathrm{d}y\ \mathrm{d}r_x\mathrm{d}r_y$ as well, such that

$$\frac{\mathrm{d}P_1}{\mathrm{d}y\ \mathrm{d}r_x\mathrm{d}r_y} = \frac{N_c\alpha_s}{2\pi^2}\frac{2}{(r_x^2+r_y^2)(r_x^2+r_y^2+\frac{1}{4})} \to \frac{N_c\alpha_s}{2\pi^2}\left[\frac{1}{(r_x^2+r_y^2)(r_x^2+r_y^2+\frac{1}{4})}+\frac{1}{((1-r_x)^2+r_y^2)((1-r_x)^2+r_y^2+\frac{1}{4})}\right]. \tag{A.18}$$

The events are weighted to the correct distributions with,

$$w_r = \frac{(r_x^2+r_y^2+1/4)((1-r_x)^2+r_y^2+1/4)}{((1-r_x)^2+r_y^2)((1-r_x)^2+r_y^2+1/4)+(r_x^2+r_y^2)(r_x^2+r_y^2+1/4)}, \tag{A.19}$$

$$w_y = \frac{\log(1+\frac{(p_\perp^p)^2}{4}e^{2y})}{2y+\log(A)}, \tag{A.20}$$

such that if $w_r w_y < R_4$ the event is rejected and the process is reiterated.

The evolution of an initial dipole thus goes as follows. Firstly, a trial emission from the initial dipole is performed according to eq. (A.6). If the rapidity y_0 of this emission is below the maximally allowed rapidity, then the trial branching is accepted, thus two new dipoles are created. Trial emissions are then allowed from each of these dipoles using $y_{\min} = y_0$ in eq. (A.6). This creates two new emissions with rapidities $y_{1,2}$. But here *only* the dipole with the smallest rapidity is accepted. Thus after the second iteration we have three dipoles, from each of which trial emissions are created and only the emission with the smallest rapidity is accepted, thus creating an additional dipole. The process is reiterated until no trial emissions are produced below the maximally allowed rapidity. The process is visualized in figure 18.

The choice of $p_\perp$ of the emitted parton is not obvious. Here we assign the parton the largest $p_\perp$ of the system,

$$p_\perp^3 = \frac{1}{\min(r_{13}, r_{23})}. \tag{A.21}$$

A.1.1 Ordering of lightcone momenta

We here rely on approximate energy conservation through ordering of p_+. This has already been discussed in the above, where we found the cutoff for small dipoles in the event generation of r, eq. (A.12). Thus we have implemented energy conservation as

$$p_+^3 \leq p_+^p, \tag{A.22}$$

which implies a rapidity-dependent cutoff for smaller dipoles.

Momentum conservation is introduced through the ordering of p_-,

$$p_-^3 \geq \max(p_-^1, p_-^2), \tag{A.23}$$

JHEP10(2019)026

where it should be noted that this requirement is applied *after* the recoils have been taken into account. This choice also sets an upper bound for the dipole size through

$$p_-^3 \geq p_-^p \Rightarrow \quad p_\perp e^{y_3} = \frac{1}{r} e^{y_3} \geq p_-^p \Rightarrow \quad r \leq \frac{e^{y_3}}{p_-} \tag{A.24}$$

A.1.2 Recoil effects

The recoil of the emitted parton is shared equally between the partons spanning the emitting dipole. Energy conservation requires that the energy of the emitter after the emission of a new dipole equals the energy of the emitter before the collision minus the recoil,

$$p_+^{\text{after}} = p_+^{\text{before}} - p_+^{\text{recoil}}. \tag{A.25}$$

The recoil cannot be determined from first principles thus have to make an ansatz. The choice here is also from [39],

$$\begin{aligned} p_+^{1,\text{recoil}} &= \frac{r_{23}}{r_{13} + r_{23}} p_+^3 \\ p_+^{2,\text{recoil}} &= \frac{r_{13}}{r_{13} + r_{23}} p_+^3 \end{aligned} \tag{A.26}$$

thus the recoil on parton 1 depends on the length of the dipole spanned by partons 2, 3 and vice versa. Energy conservation is satisfied in the event generation by always requiring that $p_+^{i,\text{recoil}} \leq p_+^{i,\text{before}}$.

The recoil will also affect the $p_\perp$ of the emitter. Here the choice is

$$p_\perp^{i,\text{after}} = \max\left(p_\perp^i, \frac{1}{r_{i3}}\right), \tag{A.27}$$

where $i = 1, 2$ are the initial partons and 3 is the emitted parton.

Changing both the $p_\perp$ and p_+ of the emitter thus also requires us to change the rapidity of the emitter for consistency,

$$y^{i,\text{after}} = \log\left(\frac{p_\perp^{i,\text{after}}}{p_+^{i,\text{after}}}\right). \tag{A.28}$$

Note here that the rapidity of the parent after the recoil will always be larger than the rapidity of the parent dipole *before* the recoil. This is because p_+ after the recoil is always smaller than p_+ before the recoil, while the $p_\perp$ is after the recoil is always larger than or equal to the $p_\perp$ before the recoil. Because of this, we must require that the rapidity of the emitters after the recoil is smaller than the rapidity of the emitted gluon, $y^{1,\text{after}}, y^{2,\text{after}} \leq y^3$.

A.1.3 Effects of confinement

Here it should be noted that the modified Bessel functions behave as $K_1(x) \sim 1/x$ for small arguments, while falling off exponentially at large arguments, $K_1(x) \sim \sqrt{\pi/x}\exp(-x)$. Thus the confined distribution is overestimated by the unconfined distribution, and the introduction of confinement only adds an additional weight $f(\text{confined})/f(\text{unconfined})$ that vetoes events with large dipole sizes.

JHEP10(2019)026

A.2 Initial states

The initial dipole configuration depends on the particle. Here we present two types: a proton sampler and a photon sampler. The difference here lies both in the number of initial dipoles (three for protons, one for photons) and in the wave function of the particle itself.

A.2.1 Protons

The initial state proton is not known from QCD, but instead has to be described by some phenomenological model. At rest, it consists primarily of three valence quarks, which we can view as endpoints of the initial dipoles. The configuration of these dipoles, however, is not known, thus we here work with a single scenario: An equilateral triangle.

We allow the dipole length to be distributed according to a Gaussian of mean r_0 and width σ_r, such that the length of the initial dipoles is given as:

$$r = r_0 + r_{\mathrm{w}} R_g \tag{A.29}$$

with R_g a Gaussian random number. The center of the triangle is fixed at origo.

A.2.2 Photons

The wave function used in this work is presented in eqs. (4.1)–(4.2). The full cross section for $\gamma^*\mathrm{p}$ is then given as

$$\sigma^{\gamma^*\mathrm{p}}(s) = \int_0^1 \mathrm{d}z \int_0^{r_{\max}} r\mathrm{d}r \int_0^{2\pi} \mathrm{d}\phi \left(|\psi_L(z,r)|^2 + |\psi_T(z,r)|^2 \right) \sigma(z,\vec{r}), \tag{A.30}$$

with $\sigma(z,\vec{r})$ the dipole-dipole scattering cross sections given in equations (3.4)–(3.6). The dipole-dipole scattering cross section goes roughly as the square of the size of the largest dipole, $\sigma(z,\vec{r}) \sim r^2$, thus an overestimate of the $\gamma^*\mathrm{p}$ cross section can be found by sampling the parameters from the following distributions, we obtain

$$z = R_1, \tag{A.31}$$

$$\phi = 2\pi R_2 \tag{A.32}$$

$$r = r_{\max} R_3, \tag{A.33}$$

$$\sigma^O_{\gamma^*\mathrm{p}} = \frac{2\pi r_{\max}}{N} \sum_{i=1}^{N} r_i^3 \left(|\psi_L(z_i,r_i)|^2 + |\psi_T(z_i,r_i)|^2 \right) \tag{A.34}$$

The maximal value obtained in the sum is kept to accept or reject the integrand in the algorithm, where first and z_i, r_i are chosen and then accepted w.r.t.

$$w_\gamma = \frac{r_i^3 \left(|\psi_L(z_i,r_)|^2 + |\psi_T(z_i,r_i)|^2 \right)}{(\text{max.} \quad \text{value})}. \tag{A.35}$$

If this weight is less than a new random number, $w_\gamma < R_4$, the event is rejected. If kept, the event is given a weight $w = \sigma^O_{\gamma^*\mathrm{p}}/r_i^2$ to take into account the overestimation of the dipole-dipole scattering cross section.

JHEP10(2019)026

B The Angantyr model for heavy ion collisions

The Angantyr model for heavy ion collisions is based on the following four components:

- Firstly, the position of the nucleons inside the nuclei has to be determined.
- Secondly, the number of interacting nucleons and binary NN collisions has to be calculated within the Glauber-Gribov (GG) formalism.
- Thirdly, the contribution to the final state of each interacting nucleon has to be determined. Here Angantyr uses the wounded nucleon model by Białas, Bleszyński and Czyż [93].
- Lastly, any hard partonic subcollision has to be modeled, thus introducing the concepts of primary and secondary absorptive interactions.

Each of the four components will here be shortly reviewed. For the full explanation, see [35, 36].

The nucleon distribution is generated using a Woods-Saxon potential:

$$\rho(r) = \frac{\rho_0\left(1 + wr^2/R^2\right)}{1 + \exp\left((r-R)/a\right)} \tag{B.1}$$

with $\rho(r)$ the radial density of the nucleons, R the radius of the nucleus, a the skin width and w the Fermi parameter, introducing a varying density but set to zero in Angantyr. The A nucleons are thus generated randomly according to $P(\vec{r}_i) = \rho(\vec{r}_i)\mathrm{d}^3\vec{r}_i$, assuming isospin invariance, such that p = n. Angantyr uses the hard core assumption, such that a new position for a nucleus is tried if the distance to its neighbours falls below twice the hard-core radius R_h. Once the nuclear distributions are set up, the impact parameter of the collision is sampled using a Gaussian distribution. This information is then passed the GG framework, which determines the fluctuations of the target and projectiles.

The fluctuations arise because of fluctuations in the proton wave function. Because the wave function enters in the cross section calculations and because it is assumed that the projectile state is frozen during its interaction with the target, these fluctuation are then translated into fluctuating cross sections. In the dipole model, the probabilistic nature of the dipole evolution gives rise to different dipole configurations before the collisions, thus giving rise to different dipole-dipole interactions and hence integrated cross sections. Angantyr uses a probability distribution for the cross section in pA extracted from Dipsy:

$$P_{\mathrm{tot}}\left(\log\sigma\right) = \frac{1}{\Omega\sqrt{2\pi}}\exp\left[-\frac{\log^2\left(\sigma/\sigma_0\right)}{2\Omega^2}\right], \tag{B.2}$$

$$\langle T(\vec{b},\sigma)\rangle = T_0\Theta\left(\sqrt{\frac{\sigma}{2\pi T_0}} - b\right), \tag{B.3}$$

with $\sigma = \int \mathrm{d}^2\vec{b}\langle 2T(\vec{b})\rangle$ and eq. (B.3) describing a slightly modified version of the elastic scattering amplitude. The parameters σ_0, Ω, T_0 are tuned to data. For AA the fluctuations

JHEP10(2019)026

are instead determined by a Gamma function,

$$P(r) = \frac{r^{k-1}e^{-r/r_0}}{\Gamma(k)r_0^k}, \tag{B.4}$$

$$T(\vec{b}, r_p, r_t) = T_0(r_p + r_t)\Theta\left(\sqrt{\frac{(r_p + r_t)^2}{2T_0}} - b\right), \tag{B.5}$$

$$T_0(r_p + r_t) = \left(1 - \exp\left[-\frac{\pi(r_p + r_t)^2}{\sigma_t}\right]\right)^{\alpha}, \tag{B.6}$$

where $P(r)$ determines fluctuations in the nucleon radius r_p and r_k. $T(\vec{b}, r_p, r_t)$ again describes a slightly modified elastic amplitude with an opacity T_0 depending on the radii of both the target and projectile. Here, the parameters σ_t, α, k, r_0 are tuned to data. The number of wounded target nucleons in pA collisions is then determined by

$$\frac{\mathrm{d}\sigma_{\mathrm{Wt}}}{\mathrm{d}^2\vec{b}} = 1 - \langle\langle S_{pt}\rangle_t^2\rangle_p, \tag{B.7}$$

with S_{pt} the S-matrix for a given target (t) and projectile (p) state. Subscript on the brackets determines averages over one side only. In AA collisions Angantyr distinguishes between absorptively and diffractively wounded nucleons, with the former dominating given by,

$$\frac{\mathrm{d}\sigma_{\mathrm{abs}}}{\mathrm{d}^2\vec{b}} = 1 - \langle S_{pt}^2\rangle_{pt}, \tag{B.8}$$

and the latter determined by generating the auxiliary states p', t' for both target and projectile, and from these determining the number of wounded target states with either t or t' from eq. (B.7), i.e. using either $S_{pt'}$ or S_{pt} in the derivation. Non-negative probabilities are ensured by shuffling when to use t, t'.

Once the number of wounded target and projectile states has been determined, the wounded nucleon model is used to create final-state partons,

$$\frac{\mathrm{d}N_{\mathrm{ch}}}{\mathrm{d}\eta} = w_p F(\eta) + w_t F(-\eta), \tag{B.9}$$

with the functions $F(\pm\eta)$ determined from the MPI framework of Pythia 8. Each nucleon in the target (and projectile) is allowed to interact several times, similar to an ordinary pp collision containing several MPIs. Thus the pairs of projectile-target nucleons are ordered w.r.t. their impact-parameter $b_{\mu\nu}$. The list is iterated over several times in order to determine which pairs give rise to a primary absorptive scattering, and which are secondary. Once a pair has been selected, the MPI framework of Pythia 8 is used to generate an event, and the pair is marked as having interacted in a primary interaction. If the pair is again chosen to interact, it will be marked as a secondary interaction. After the determination of the absorptive interactions, the diffractive ones are chosen by iterating the list several times, thus creating primary and secondary diffractive interactions. An already wounded nucleon cannot be further excited, but an unwounded nucleon can participate in several diffractive interactions, until itself becomes wounded.

JHEP10(2019)026

After the determination of the absorptive and diffractively primary and secondary interactions, each of the events are passed to Pythia 8 and the parton-level events are stacked on top of each other. The Pythia 8 description of single-diffractive events are modified to look like non-diffractive ones, to describe the secondary absorptive events, while diffractive primary and secondary events remain unmodified. We are thus left with a large set of parton-level events that can be passed to the hadronisation framework of Pythia 8 and further analysed.

C Additional eccentricity figures

In this section, we show additional eccentricity figures not presented in the main body of the text. Figure 19 (a-c) shows ϵ_2 using higher-order cumulants in the evaluation. It is evident that the eccentricities are the same regardless of the number of particles used in the calculation, except for the effects from lack of statistics in the high-multiplicity tail for both the pp and pPb figures. Figure 19 (d) shows the normalized symmetric cumulant $NSC(4,2)$. This cumulant is positive in the entire multiplicity range, consistent with measurements in ALICE. Here, it is evident that discrimination between models would be possible in both pp and pPb collisions, as opposed to $NSC(3,2)$ where discriminatory power was not evident in pp collisions.

For completeness, we also show the eccentricities $\epsilon_{1,3}$ obtained in pp collisions with and without shower smearing in figure 20. Both are shown to give an estimate of the effects on the size of the additional terms in the Fourier expansion of the flow coefficients.

JHEP10(2019)026

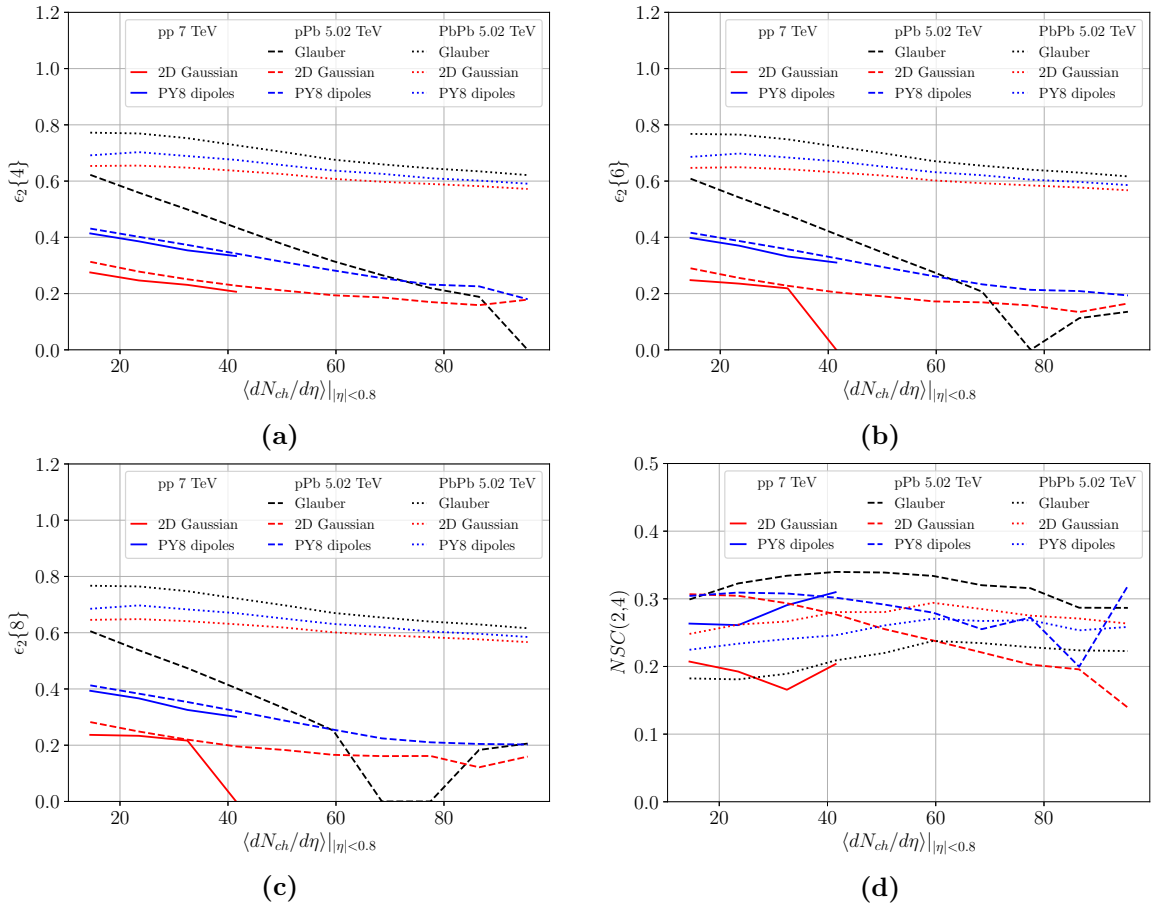


Figure 19. (a-c) The eccentricity ϵ_2 with higher order cumulants $\{4, 6, 8\}$. (d) The normalised symmetric cumulant $NSC(4, 2)$ as a function of average multiplicity for pp, pA, AA systems.

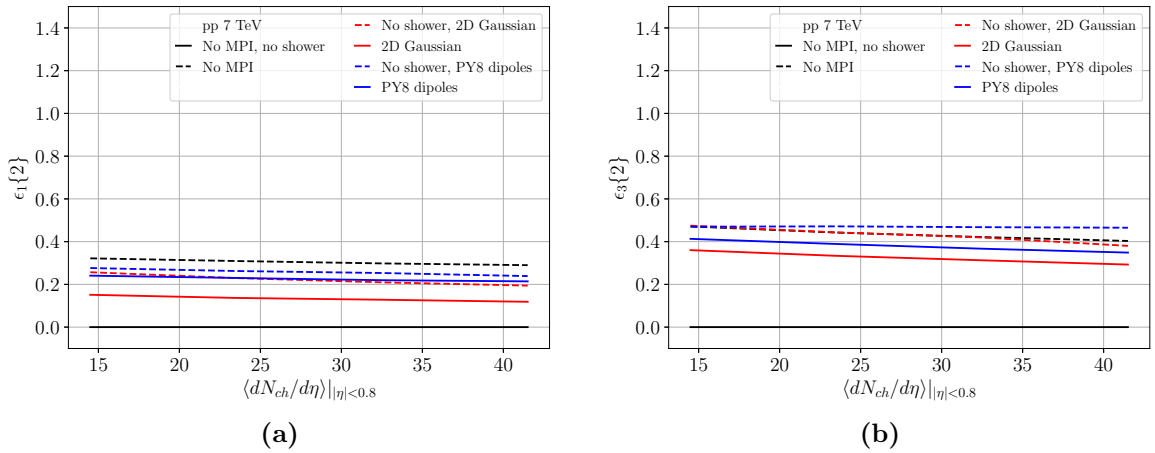


Figure 20. $\epsilon_{1,3}\{2\}$ shown for different MPI vertex assignments with and without the shower smearing.

JHEP10(2019)026

References

[1] STAR collaboration, *Elliptic flow in Au + Au collisions at* ($\sqrt{s_{nn}} = 130$ *GeV*, *Phys. Rev. Lett.* **86** (2001) 402 [nucl-ex/0009011] [INSPIRE].

[2] ALICE collaboration, *Elliptic flow of charged particles in Pb-Pb collisions at 2.76 TeV*, *Phys. Rev. Lett.* **105** (2010) 252302 [arXiv:1011.3914] [INSPIRE].

[3] CMS collaboration, *Long-range and short-range dihadron angular correlations in central PbPb collisions at a nucleon-nucleon center of mass energy of 2.76 TeV*, *JHEP* **07** (2011) 076 [arXiv:1105.2438] [INSPIRE].

[4] ATLAS collaboration, *Measurement of the azimuthal anisotropy for charged particle production in* $\sqrt{s_{NN}} = 2.76$ *TeV lead-lead collisions with the ATLAS detector*, *Phys. Rev.* **C 86** (2012) 014907 [arXiv:1203.3087] [INSPIRE].

[5] J.-Y. Ollitrault, *Anisotropy as a signature of transverse collective flow*, *Phys. Rev.* **D 46** (1992) 229 [INSPIRE].

[6] V.A. Abramovsky, E.V. Gedalin, E.G. Gurvich and O.V. Kancheli, *Long Range Azimuthal Correlations in Multiple Production Processes at High-energies*, *JETP Lett.* **47** (1988) 337 [INSPIRE].

[7] CMS collaboration, *Observation of Long-Range Near-Side Angular Correlations in Proton-Lead Collisions at the LHC*, *Phys. Lett.* **B 718** (2013) 795 [arXiv:1210.5482] [INSPIRE].

[8] PHENIX collaboration, *Measurements of Multiparticle Correlations in* $d+$Au *Collisions at 200, 62.4, 39 and 19.6 GeV and* $p+$Au *Collisions at 200 GeV and Implications for Collective Behavior*, *Phys. Rev. Lett.* **120** (2018) 062302 [arXiv:1707.06108] [INSPIRE].

[9] CMS collaboration, *Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC*, *JHEP* **09** (2010) 091 [arXiv:1009.4122] [INSPIRE].

[10] A. Badea et al., *Measurements of two-particle correlations in* e^+e^- *collisions at 91 GeV with ALEPH archived data*, arXiv:1906.00489 [INSPIRE].

[11] J.L. Nagle and W.A. Zajc, *Small System Collectivity in Relativistic Hadronic and Nuclear Collisions*, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 211 [arXiv:1801.03477] [INSPIRE].

[12] Z. Citron et al., *Future physics opportunities for high-density QCD at the LHC with heavy-ion and proton beams*, in *HL/HE-LHC Workshop: Workshop on the Physics of HL-LHC and Perspectives at HE-LHC Geneva, Switzerland, June 18–20, 2018*, 2018, arXiv:1812.06772 [INSPIRE].

[13] B. Blok, C.D. Jäkel, M. Strikman and U.A. Wiedemann, *Collectivity from interference*, *JHEP* **12** (2017) 074 [arXiv:1708.08241] [INSPIRE].

[14] K. Dusling and R. Venugopalan, *Explanation of systematics of CMS p+Pb high multiplicity di-hadron data at* $\sqrt{s}_{\rm NN} = 5.02$ *TeV*, *Phys. Rev.* **D 87** (2013) 054014 [arXiv:1211.3701] [INSPIRE].

JHEP10(2019)026

[15] R.D. Weller and P. Romatschke, *One fluid to rule them all: viscous hydrodynamic description of event-by-event central p+p, p+Pb and Pb+Pb collisions at* $\sqrt{s} = 5.02$ *TeV*, *Phys. Lett.* **B 774** (2017) 351 [arXiv:1701.07145] [INSPIRE].

[16] K. Werner, *Core-corona separation in ultra-relativistic heavy ion collisions*, *Phys. Rev. Lett.* **98** (2007) 152301 [arXiv:0704.1270] [INSPIRE].

[17] T. Pierog, I. Karpenko, J.M. Katzy, E. Yatsenko and K. Werner, *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, *Phys. Rev.* **C 92** (2015) 034906 [arXiv:1306.0121] [INSPIRE].

[18] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].

[19] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J.* **C 76** (2016) 196 [arXiv:1512.01178] [INSPIRE].

[20] D. d'Enterria et al., *Estimates of hadron azimuthal anisotropy from multiparton interactions in proton-proton collisions at* $\sqrt{s} = 14$ *TeV*, *Eur. Phys. J.* **C 66** (2010) 173 [arXiv:0910.3029] [INSPIRE].

[21] J.L. Albacete and A. Soto-Ontoso, *Hot spots and the hollowness of proton-proton interactions at high energies*, *Phys. Lett.* **B 770** (2017) 149 [arXiv:1605.09176] [INSPIRE].

[22] C. Bierlich, G. Gustafson and L. Lönnblad, *Collectivity without plasma in hadronic collisions*, *Phys. Lett.* **B 779** (2018) 58 [arXiv:1710.09725] [INSPIRE].

[23] A.H. Mueller, *Soft gluons in the infinite momentum wave function and the BFKL pomeron*, *Nucl. Phys.* **B 415** (1994) 373 [INSPIRE].

[24] A.H. Mueller and B. Patel, *Single and double BFKL pomeron exchange and a dipole picture of high-energy hard processes*, *Nucl. Phys.* **B 425** (1994) 471 [hep-ph/9403256] [INSPIRE].

[25] E. Avsar, C. Flensburg, Y. Hatta, J.-Y. Ollitrault and T. Ueda, *Eccentricity and elliptic flow in proton-proton collisions from parton evolution*, *Phys. Lett.* **B 702** (2011) 394 [arXiv:1009.5643] [INSPIRE].

[26] I. Balitsky, *Operator expansion for high-energy scattering*, *Nucl. Phys.* **B 463** (1996) 99 [hep-ph/9509348] [INSPIRE].

[27] I. Balitsky, *Factorization for high-energy scattering*, *Phys. Rev. Lett.* **81** (1998) 2024 [hep-ph/9807434] [INSPIRE].

[28] I. Balitsky, *Effective field theory for the small x evolution*, *Phys. Lett.* **B 518** (2001) 235 [hep-ph/0105334] [INSPIRE].

[29] J. Jalilian-Marian, A. Kovner, A. Leonidov and H. Weigert, *The BFKL equation from the Wilson renormalization group*, *Nucl. Phys.* **B 504** (1997) 415 [hep-ph/9701284] [INSPIRE].

[30] J. Jalilian-Marian, A. Kovner, A. Leonidov and H. Weigert, *The Wilson renormalization group for low x physics: Towards the high density regime*, *Phys. Rev.* **D 59** (1998) 014014 [hep-ph/9706377] [INSPIRE].

[31] E. Iancu, A. Leonidov and L.D. McLerran, *The renormalization group equation for the color glass condensate*, *Phys. Lett.* **B 510** (2001) 133 [hep-ph/0102009] [INSPIRE].

[32] E. Iancu, A. Leonidov and L.D. McLerran, *Nonlinear gluon evolution in the color glass condensate. 1*, *Nucl. Phys.* **A 692** (2001) 583 [hep-ph/0011241] [INSPIRE].

JHEP10(2019)026

[33] H. Weigert, *Unitarity at small Bjorken x*, *Nucl. Phys.* **A 703** (2002) 823 [hep-ph/0004044] [INSPIRE].

[34] M. Deak, K. Kutak, W. Li and A.M. Staśto, *On the different forms of the kinematical constraint in BFKL*, *Eur. Phys. J.* **C 79** (2019) 647 [arXiv:1906.09062] [INSPIRE].

[35] C. Bierlich, G. Gustafson and L. Lönnblad, *Diffractive and non-diffractive wounded nucleons and final states in pA collisions*, *JHEP* **10** (2016) 139 [arXiv:1607.04434] [INSPIRE].

[36] C. Bierlich, G. Gustafson, L. Lönnblad and H. Shah, *The Angantyr model for Heavy-Ion Collisions in PYTHIA8*, *JHEP* **10** (2018) 134 [arXiv:1806.10820] [INSPIRE].

[37] A. Accardi et al., *Electron Ion Collider: The Next QCD Frontier*, *Eur. Phys. J.* **A 52** (2016) 268 [arXiv:1212.1701] [INSPIRE].

[38] G.P. Salam, *OEDIPUS: Onium evolution, dipole interaction and perturbative unitarization simulation*, *Comput. Phys. Commun.* **105** (1997) 62 [hep-ph/9601220] [INSPIRE].

[39] E. Avsar, G. Gustafson and L. Lönnblad, *Energy conservation and saturation in small-x evolution*, *JHEP* **07** (2005) 062 [hep-ph/0503181] [INSPIRE].

[40] C. Flensburg, G. Gustafson and L. Lönnblad, *Inclusive and Exclusive Observables from Dipoles in High Energy Collisions*, *JHEP* **08** (2011) 103 [arXiv:1103.4321] [INSPIRE].

[41] V. Kovalenko and V. Vechernin, *Model of pp and AA collisions for the description of long-range correlations*, PoS(Baldin-ISHEPP-XXI)077 (2012) [arXiv:1212.2590] [INSPIRE].

[42] E. Avsar, G. Gustafson and L. Lönnblad, *Diifractive excitation in DIS and pp collisions*, *JHEP* **12** (2007) 012 [arXiv:0709.1368] [INSPIRE].

[43] E. Avsar, *On the dipole swing and the search for frame independence in the dipole model*, *JHEP* **11** (2007) 027 [arXiv:0709.1371] [INSPIRE].

[44] B. Schenke, P. Tribedy and R. Venugopalan, *Fluctuating Glasma initial conditions and flow in heavy ion collisions*, *Phys. Rev. Lett.* **108** (2012) 252301 [arXiv:1202.6646] [INSPIRE].

[45] A.H. Mueller, *A simple derivation of the JIMWLK equation*, *Phys. Lett.* **B 523** (2001) 243 [hep-ph/0110169] [INSPIRE].

[46] E.A. Kuraev, L.N. Lipatov and V.S. Fadin, *The Pomeranchuk Singularity in Nonabelian Gauge Theories*, *Sov. Phys. JETP* **45** (1977) 199 [INSPIRE].

[47] I.I. Balitsky and L.N. Lipatov, *The Pomeranchuk Singularity in Quantum Chromodynamics*, *Sov. J. Nucl. Phys.* **28** (1978) 822 [INSPIRE].

[48] Y.V. Kovchegov, *Small x F(2) structure function of a nucleus including multiple Pomeron exchanges*, *Phys. Rev.* **D 60** (1999) 034008 [hep-ph/9901281] [INSPIRE].

[49] Y.V. Kovchegov and E. Levin, *Quantum chromodynamics at high energy*, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **33** (2012) 1 [INSPIRE].

[50] I. Balitsky and G.A. Chirilli, *Next-to-leading order evolution of color dipoles*, *Phys. Rev.* **D 77** (2008) 014019 [arXiv:0710.4330] [INSPIRE].

[51] A. Kovner, M. Lublinsky and Y. Mulian, *Jalilian-Marian, Iancu, McLerran, Weigert, Leonidov, Kovner evolution at next to leading order*, *Phys. Rev.* **D 89** (2014) 061704 [arXiv:1310.0378] [INSPIRE].

[52] A. Kovner, M. Lublinsky and Y. Mulian, *NLO JIMWLK evolution unabridged*, *JHEP* **08** (2014) 114 [arXiv:1405.0418] [INSPIRE].

JHEP10(2019)026

[53] M. Lublinsky and Y. Mulian, *High Energy QCD at NLO: from light-cone wave function to JIMWLK evolution*, *JHEP* **05** (2017) 097 [arXiv:1610.03453] [INSPIRE].

[54] T. Lappi and H. Mäntysaari, *Direct numerical solution of the coordinate space Balitsky-Kovchegov equation at next to leading order*, *Phys. Rev.* **D 91** (2015) 074016 [arXiv:1502.02400] [INSPIRE].

[55] B. Ducloué, E. Iancu, A.H. Mueller, G. Soyez and D.N. Triantafyllopoulos, *Non-linear evolution in QCD at high-energy beyond leading order*, *JHEP* **04** (2019) 081 [arXiv:1902.06637] [INSPIRE].

[56] E. Avsar, G. Gustafson and L. Lönnblad, *Small-x dipole evolution beyond the large-N_c imit*, *JHEP* **01** (2007) 012 [hep-ph/0610157] [INSPIRE].

[57] G. Beuf, *Improving the kinematics for low-x QCD evolution equations in coordinate space*, *Phys. Rev.* **D 89** (2014) 074039 [arXiv:1401.0313] [INSPIRE].

[58] V.S. Fadin and L.N. Lipatov, *BFKL Pomeron in the next-to-leading approximation*, *Phys. Lett.* **B 429** (1998) 127 [hep-ph/9802290] [INSPIRE].

[59] M. Ciafaloni and G. Camici, *Energy scale(s) and next-to-leading BFKL equation*, *Phys. Lett.* **B 430** (1998) 349 [hep-ph/9803389] [INSPIRE].

[60] G.P. Salam, *An introduction to leading and next-to-leading BFKL*, *Acta Phys. Polon.* **B 30** (1999) 3679 [hep-ph/9910492] [INSPIRE].

[61] T. Altinoluk, N. Armesto, G. Beuf, M. Martínez and C.A. Salgado, *Next-to-eikonal corrections in the CGC: gluon production and spin asymmetries in pA collisions*, *JHEP* **07** (2014) 068 [arXiv:1404.2219] [INSPIRE].

[62] H. Niemi, G.S. Denicol, H. Holopainen and P. Huovinen, *Event-by-event distributions of azimuthal asymmetries in ultrarelativistic heavy-ion collisions*, *Phys. Rev.* **C 87** (2013) 054901 [arXiv:1212.1008] [INSPIRE].

[63] C. Bierlich, G. Gustafson and L. Lönnblad, *A shoving model for collectivity in hadronic collisions*, arXiv:1612.05132 [INSPIRE].

[64] Z. Qiu and U.W. Heinz, *Event-by-event shape and flow fluctuations of relativistic heavy-ion collision fireballs*, *Phys. Rev.* **C 84** (2011) 024911 [arXiv:1104.0650] [INSPIRE].

[65] Y. Zhou, *Anisotropic Flow and Flow Fluctuations at the Large Hadron Collider*, Ph.D. thesis, Utrecht University, The Netherlands (2016).

[66] J.L. Albacete, H. Petersen and A. Soto-Ontoso, *Symmetric cumulants as a probe of the proton substructure at LHC energies*, *Phys. Lett.* **B 778** (2018) 128 [arXiv:1707.05592] [INSPIRE].

[67] E. Avsar and G. Gustafson, *Geometric scaling and QCD dynamics in DIS*, *JHEP* **04** (2007) 067 [hep-ph/0702087] [INSPIRE].

[68] A.H. Mueller and G.P. Salam, *Large multiplicity fluctuations and saturation effects in onium collisions*, *Nucl. Phys.* **B 475** (1996) 293 [hep-ph/9605302] [INSPIRE].

[69] T. Sjöstrand and M. van Zijl, *A Multiple Interaction Model for the Event Structure in Hadron Collisions*, *Phys. Rev.* **D 36** (1987) 2019 [INSPIRE].

[70] T. Sjöstrand and P.Z. Skands, *Multiple interactions and the structure of beam remnants*, *JHEP* **03** (2004) 053 [hep-ph/0402078] [INSPIRE].

JHEP10(2019)026

[71] A. Donnachie and P.V. Landshoff, *Total cross-sections*, *Phys. Lett.* **B 296** (1992) 227 [hep-ph/9209205] [INSPIRE].

[72] G.A. Schuler and T. Sjöstrand, *Hadronic diffractive cross-sections and the rise of the total cross-section*, *Phys. Rev.* **D 49** (1994) 2257 [INSPIRE].

[73] R.B. Appleby, R.J. Barlow, J.G. Molson, M. Serluca and A. Toader, *The Practical Pomeron for High Energy Proton Collimation*, *Eur. Phys. J.* **C 76** (2016) 520 [arXiv:1604.07327] [INSPIRE].

[74] C.O. Rasmussen and T. Sjöstrand, *Models for total, elastic and diffractive cross sections*, *Eur. Phys. J.* **C 78** (2018) 461 [arXiv:1804.10373] [INSPIRE].

[75] H1 collaboration, *Deep inelastic inclusive e p scattering at low x and a determination of* α_s, *Eur. Phys. J.* **C 21** (2001) 33 [hep-ex/0012053] [INSPIRE].

[76] A. Buckley, H. Hoeth, H. Lacker, H. Schulz and J.E. von Seggern, *Systematic event generator tuning for the LHC*, *Eur. Phys. J.* **C 65** (2010) 331 [arXiv:0907.2973] [INSPIRE].

[77] ALICE collaboration, *Investigations of anisotropic flow using multi-particle azimuthal correlations in pp, p-Pb, Xe-Xe and Pb-Pb collisions at the LHC*, arXiv:1903.01790 [INSPIRE].

[78] CMS collaboration, *Multiparticle correlation studies in pPb collisions at* $\sqrt{s_{\rm NN}} = 8.16$ *TeV*, arXiv:1904.11519 [INSPIRE].

[79] G. Giacalone, J. Noronha-Hostler and J.-Y. Ollitrault, *Relative flow fluctuations as a probe of initial state fluctuations*, *Phys. Rev.* **C 95** (2017) 054910 [arXiv:1702.01730] [INSPIRE].

[80] J.E. Bernhard, J.S. Moreland, S.A. Bass, J. Liu and U. Heinz, *Applying Bayesian parameter estimation to relativistic heavy-ion collisions: simultaneous characterization of the initial state and quark-gluon plasma medium*, *Phys. Rev.* **C 94** (2016) 024907 [arXiv:1605.03954] [INSPIRE].

[81] L. Yan and J.-Y. Ollitrault, *Universal fluctuation-driven eccentricities in proton-proton, proton-nucleus and nucleus-nucleus collisions*, *Phys. Rev. Lett.* **112** (2014) 082301 [arXiv:1312.6555] [INSPIRE].

[82] R.J. Glauber, *Cross-sections in deuterium at high-energies*, *Phys. Rev.* **100** (1955) 242 [INSPIRE].

[83] M.L. Miller, K. Reygers, S.J. Sanders and P. Steinberg, *Glauber modeling in high energy nuclear collisions*, *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 205 [nucl-ex/0701025] [INSPIRE].

[84] M. Alvioli and M. Strikman, *Color fluctuation effects in proton-nucleus collisions*, *Phys. Lett.* **B 722** (2013) 347 [arXiv:1301.0728] [INSPIRE].

[85] H. Heiselberg, G. Baym, B. Blaettel, L.L. Frankfurt and M. Strikman, *Color transparency, color opacity and fluctuations in nuclear collisions*, *Phys. Rev. Lett.* **67** (1991) 2946 [INSPIRE].

[86] B. Blaettel, G. Baym, L.L. Frankfurt, H. Heiselberg and M. Strikman, *Hadronic cross-section fluctuations*, *Phys. Rev.* **D 47** (1993) 2761 [INSPIRE].

[87] V.P. Gonçalves, R.P. da Silva and P.V. R.G. Silva, *Diffractive excitation in pp and pA collisions at high energies*, *Phys. Rev.* **D 100** (2019) 014019 [arXiv:1905.00806] [INSPIRE].

[88] T.S. Biro, H.B. Nielsen and J. Knoll, *Color Rope Model for Extreme Relativistic Heavy Ion Collisions*, *Nucl. Phys.* **B 245** (1984) 449 [INSPIRE].

JHEP10(2019)026

[89] C. Bierlich, G. Gustafson, L. Lönnblad and A. Tarasov, *Effects of Overlapping Strings in pp Collisions*, *JHEP* **03** (2015) 148 [arXiv:1412.6259] [INSPIRE].

[90] C. Bierlich and J.R. Christiansen, *Effects of color reconnection on hadron flavor observables*, *Phys. Rev.* **D 92** (2015) 094010 [arXiv:1507.02091] [INSPIRE].

[91] M.L. Mangano and B. Nachman, *Observables for possible QGP signatures in central pp collisions*, *Eur. Phys. J.* **C 78** (2018) 343 [arXiv:1708.08369] [INSPIRE].

[92] C. Bierlich, *Soft modifications to jet fragmentation in high energy proton-proton collisions*, *Phys. Lett.* **B 795** (2019) 194 [arXiv:1901.07447] [INSPIRE].

[93] A. Białas, M. Bleszyński and W. Czyż, *Multiplicity Distributions in Nucleus-Nucleus Collisions at High-Energies*, *Nucl. Phys.* **B 111** (1976) 461 [INSPIRE].

JHEP10(2019)026