

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Helenius, Ilkka; Utheim, Marius

**Title:** Hadron-ion collisions in Pythia and the vector-meson dominance model for photoproduction

**Year:** 2024

**Version:** Published version

**Copyright:** © The Author(s) 2024

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Helenius, I., & Utheim, M. (2024). Hadron-ion collisions in Pythia and the vector-meson dominance model for photoproduction. *European Physical Journal C*, 84, Article 1155.  
<https://doi.org/10.1140/epjc/s10052-024-13543-6>



Regular Article

Theoretical Physics

# Hadron-ion collisions in Pythia and the vector-meson dominance model for photoproduction

Ilkka Helenius<sup>1,2,a</sup>, Marius Utheim<sup>1,2,3</sup> 

<sup>1</sup> Department of Physics, University of Jyväskylä, P.O. Box 35, FI-40014 Jyväskylä, Finland

<sup>2</sup> Helsinki Institute of Physics, University of Helsinki, P.O. Box 64, FI-00014 Helsinki, Finland

<sup>3</sup> Vikaveien 214, 9303 Silsand, Norway

Received: 22 June 2024 / Accepted: 15 October 2024  
© The Author(s) 2024

**Abstract** We present an extension to the PYTHIA Monte Carlo event generator that enables simulations of collisions between a generic hadron beam on a nuclear target with energy variation in event-by-event basis. This builds upon PYTHIA's module for heavy ions, ANGANTYR, as well as previous work on simulating hadron-proton collisions. As such, the extensions in this work are largely technical, except for a rudimentary model for hadronic fluctuations. With hadron-ion simulations, we implement an explicit vector-meson dominance (VMD) model that can be used to simulate interactions of hadronic component of real photons in photo-nuclear collisions. Such processes can be studied in ultra-peripheral heavy-ion collisions and in the future also with the upcoming Electron-Ion Collider. Our work also has applications to hadronic showers, e.g. air showers initiated by high-energy cosmic rays. We first validate the VMD model by comparing to HERA photoproduction data on proton target. Then we apply this to generate events for ultra-peripheral heavy-ion collisions at the LHC and present the results corresponding to the event-selection criteria matching to a recent ATLAS analysis. We find that single-particle multiplicity and rapidity distributions are well in line with the measured ones. We also construct the Fourier coefficients from two-particle correlations for the simulated events and study whether the resulting azimuthal anisotropies are consistent with the ATLAS results.

## 1 Introduction

PYTHIA is a general-purpose event generator that can simulate high-energy collisions events [1]. The underlying physics modelling is built upon collinear factorization [2] where

hard-process cross sections can be factorized from long-distance physics describing the structure of the colliding particles. The latter can be obtained by applying DGLAP evolution for a non-perturbative input that is fitted to experimental data in a global analysis of parton distribution functions (PDFs) [3]. In event generators the DGLAP equations can be applied also to generate parton showers where highly-virtual partons formed in a hard process radiate more partons [4], which later on turn into hadrons at non-perturbative scales that construct the complete final state. Within the previous decade the main focus in PYTHIA has been in proton-proton collisions performed at the Large Hadron Collider (LHC) at CERN. More recently there has, however, been several developments to extend the modelling also to other beam configurations including photoproduction in lepton-proton collisions [5], collisions involving heavy ions [6], and collisions with varying hadron beam and a proton target [7]. In this work we combine the two latter features to enable event generation for collisions of a generic hadron and a heavy ion. Such modelling has applications to photon-induced processes such as ultra-peripheral heavy-ion collisions at the LHC and electron-ion collisions in the future Electron-Ion Collider (EIC) [8,9], as well as hadronic cascades such as air showers initiated by high-energy cosmic rays.

In ultra-peripheral collisions (UPCs) [10,11] the beam particles encounter with a large impact parameter such that no short-distance strong interaction, described with quantum chromodynamics (QCD), can take place but long-distance electromagnetic interactions can still occur. These interactions can be described by exchanging a quasi-real photon and thus allow us to study photon-photon and photon-hadron collisions. Unlike in case of charged point-like leptons, the photons emitted by charged hadrons typically have a limited virtuality which essentially follows from the requirement

<sup>a</sup> e-mail: [ilkka.m.helenius@jyu.fi](mailto:ilkka.m.helenius@jyu.fi) (corresponding author)

of hadron staying intact when emitting a photon. Therefore these processes are similar to photoproduction [12], studied extensively e.g. in electron-proton collisions at HERA [13]. With heavy-ion collisions at the LHC one can also study interactions between a real photon and a nuclear target at collider energies. Several recent experimental studies have been performed, including exclusive particle production [14–17], inclusive dijets [18] and two-particle correlations [19]. Interestingly, the latter study have found signs of long-range correlations typically associated with hydrodynamical flow of strongly interacting medium [20]. Similar effects have been observed earlier in case high-multiplicity proton-nucleus [21, 22] and proton-proton [23, 24] collisions which have raised a question whether there are other mechanisms that could result as similar signatures for collective behaviour. The high-multiplicity photo-nuclear interactions offer thus a complementary “small” collisions system to study onset of hydrodynamic behaviour and the origin of the observed collective effects.

When studying collisions of a real photon there are several contributions that should be accounted for. The photons can of course interact as point-like particle (direct photon) or it can fluctuate into a hadronic state (resolved photon) with equal quantum numbers. The latter may happen non-perturbatively, or the photon can split into quark-antiquark pair perturbatively. The partonic structure of these hadronic states can be encoded in PDFs which now contain a perturbatively calculable point-like component and a hadron-like component. As the latter contribution is not known, a non-perturbative input that can be fixed in a global QCD analysis is needed. Another approach for this is to apply the so called vector-meson dominance (VMD) model where a hadron-like photon is treated as a linear combination of vector-mesons states [25]. Independently of the chosen approach, one ends up on modelling collisions between a (generalized) hadron and a hadron or nuclear target.

The modelling of photon-induced processes with general-purpose Monte Carlo event generators has been recently gaining some attention. In PYTHIA, a new implementation of such processes has been available for a few years already and recently a similar framework has been set up in Sherpa [26] and to some extent also in Herwig [27]. In all cases a reasonable agreement with various HERA and LEP data has been found [28, 29]. However, so far no attempt has been made to model interactions between a real photon and a nuclear target. Such a setup would, however, be very useful for experimental analyses and would provide a realistic baseline for correlation studies and underlying event for jet production. In case of HERA photoproduction data it has been observed that multiparton interactions (MPIs) are a necessary ingredient to describe the measured multiplicity distributions [30], which can be naturally explained with additional partonic interactions between the remnant of the resolved photon and the

target hadron. In the case of nuclear target, one needs to take an additional step forward and account also for multi-nucleon interactions, where the remnant of the resolved-photon state interacts with the other nucleons in the nuclei. Comparison between the measured multiplicity distributions in photon-proton [31] and photon-lead [19] collisions shows that such effects are important and should be included to obtain realistic estimates for particle production.

Another place where hadron-ion collisions occur in nature is in hadronic cascades initiated by high-energy cosmic rays striking our atmosphere [32]. One open question in cosmic ray physics concerns the production mechanism for ultra-high-energy cosmic rays. Another is the observation that the number of muons measured in these showers does not match the predictions of air shower simulations such as CORSIKA 8 [33] – the so-called “muon puzzle”. Air shower simulations can be divided into one component for simulating interactions between hadrons and gas atoms in the air, and one component for the propagation of the produced hadrons to their next interaction point. For the interactions, CORSIKA uses external models, the most prominent ones being Sibyll 2.3d [34], EPOS-LHC [35], and QGSJet-II.04 [36]. With the development of generic hadron-hadron interactions in PYTHIA also included a toy model for including nuclear effects, but recent work has shown that it gives predictions that are not fully consistent with the other reigning models [37].

In PYTHIA, proton-nucleus and nucleus-nucleus collisions are simulated with the ANGANTYR model [6]. Originally developed as a plugin, this model is now integrated as a part of core program. In case of heavy-ion collisions, the model relies on partonic interactions instead of explicitly including a strongly interacting thermalized medium, i.e. the quark gluon plasma. It has been shown that by including string interactions in hadronization and final-state rescattering, some of the observed collective effects can be reproduced in lead-lead collisions at the LHC [38]. In case of proton-lead collisions, the ANGANTYR model provide a good description of the measured multiplicities [6].

In this work we introduce an extension of ANGANTYR model in PYTHIA that accepts any hadron (except multi-heavy flavoured baryons, whose PDFs are not implemented in PYTHIA) as a beam particle colliding with a target nucleus. Here our main motivation is to apply this to setup up an explicit VMD model that can be applied to simulate the minimum bias photon-nucleus collisions, and only briefly discuss the application to cosmic rays. For realistic simulations, the shape of photon energy-spectrum needs to be considered which we have here accounted by allowing varying beam energies that can be matched to the calculated photon flux. We also include a feature to switch between different vector mesons on an event-by-event basis. We first introduce our simulation framework in Sect. 3, validate it against HERA

data in Sect. 4.1 and present comparisons to UPC data in Sect. 4.2. We pay special attention to two-particle correlations and perform the template fitting procedure applied by ATLAS to obtain Fourier decomposition with and without non-flow subtraction in Sect. 4.3. The work is summarized in Sect. 5 where also possible future directions are outlined.

## 2 Theoretical background

### 2.1 Generic hadronic interactions

Interactions between generic hadrons scattering against a nucleon were implemented in PYTHIA 8.307 [7]. In this framework, an interaction type is determined randomly according to the partial cross sections of the hadrons involved, whose relative compositions vary from species to species. Once a process type has been selected, a hadron can either scatter elastically, or break up in a diffractive or non-diffractive scattering. In the latter case, the hadron is decomposed into its constituent quarks as determined by its PDFs. After this decomposition, the rest of the process interacts only with the partons, and is agnostic to the original hadron species. Hence, the differences between proton-proton and hadron-proton interactions in PYTHIA can be captured by changing the cross sections and the PDFs.

The total cross section can be written as the sum of the partial ones,

$$\sigma_{\text{tot}} = \sigma_{\text{el}} + \sigma_{\text{ND}} + \sigma_{XB} + \sigma_{AY} + \sigma_{XY} + \dots \quad (1)$$

In our approach, we parameterize the total, elastic, and diffractive cross sections, and define the non-diffractive cross section as the remainder,

$$\sigma_{\text{ND}} = \sigma_{\text{tot}} - \sigma_{\text{el}} - \sigma_{\text{ND}} - \sigma_{XB} + \sigma_{AY} - \sigma_{XY} - \dots \quad (2)$$

In Ref. [7], the total cross section is based on the Donnachie–Landshoff model [39],

$$\sigma_{\text{tot}}^{AB} = X^{AB} s^{\epsilon} + Y^{AB} s^{-\eta}, \quad (3)$$

where  $s$  is the squared CM energy in units of  $\text{GeV}^2$ , the exponents  $\epsilon = 0.0808$  and  $\eta = 0.4525$  are parameters that do not depend on the hadron in question, and  $X^{AB}$  and  $Y^{AB}$  are parameters that do depend on the hadron species. The  $s^{\epsilon}$  and  $s^{-\eta}$  terms correspond to pomeron and reggeon exchanges, respectively. Going beyond the most common hadrons, such as protons and pions, very little data exists for these cross sections, and the parameters are based on very simple considerations such as taking linear combinations of other hadrons with the same valence content, or applying the additive quark model (AQM) [40,41], as described in Ref. [7]. Diffractive cross sections are calculated using the Schuler and Sjöstrand ansatz [42,43], with expressions such as

$$d\sigma_{\text{SD}} = \frac{g_{3\mathbb{P}}\beta_{A\mathbb{P}}\beta_{B\mathbb{P}}^2}{16\pi} \frac{dM_X^2}{M_X^2} (e^{B_{XB}t} dt) F_{\text{SD}}(M_X^2, s) \quad (4)$$

for single diffraction  $AB \rightarrow XB$ . Here, the hadron species-dependent parameters are the Pomeron couplings  $\beta$ , which with the appropriate ansatz and normalization can be expressed in terms of  $X^{A\mathbb{P}}$ , and the exponential slope  $B_{XB}$ , which depends on whether hadron  $B$  is a baryon or meson. Similar expressions are used for the elastic and double diffractive components,

$$d\sigma_{\text{el}} = (1 + \rho^2) \frac{\sigma_{\text{tot}}^2}{16\pi} e^{B_{\text{el}}t} dt, \quad (5)$$

$$d\sigma_{XY} = \frac{g_{3\mathbb{P}}\beta_{A\mathbb{P}}\beta_{B\mathbb{P}}}{16\pi} \frac{dM_X^2 dM_Y^2}{M_X^2 M_Y^2} (e^{B_{XY}t} dt) F_{\text{DD}}(M_X^2, M_Y^2, s). \quad (6)$$

The non-diffractive cross section is given in the end as the difference between the total and the other partial cross sections.

There is very little data for PDFs beyond nucleons, and as such, the models used in PYTHIA are very simple. The basic ansatz is that of Glück, Reya et al. [44,45], which takes the valence, sea and gluon distributions at some initial scale  $Q_0^2$  to be on the form

$$f(x, Q_0^2) = N x^a (1-x)^b (1 + A\sqrt{x} + BX), \quad (7)$$

which can then be evolved to higher scales through DGLAP evolution. In PYTHIA, this is simplified further for the valence content by assuming  $A = B = 0$ . The normalization factor  $N$  is given by the valence sum relations, but there is no solid theory for determining  $a$  and  $b$ . One guiding principle is that all valence quarks must have roughly the same velocity in order for the hadron to stay intact, so that heavier quarks must have a smaller  $b$  and a larger momentum fraction  $\langle x \rangle$ . Even so, there is no simple way to determine these quantities, so their exact values are chosen for each hadron based on heuristic guesses (see Ref. [7] for further discussion). The gluon and u/d sea distributions at the initial scale are assumed to be the same as for the pion, except multiplied by a factor  $x^d$  in order to soften them in light of the increased  $\langle x \rangle$ . Finally, the gluon and sea distributions are scaled by a normalization factor in order to satisfy the momentum sum relation.

### 2.2 ANGANTYR

ANGANTYR is the heavy ion model of PYTHIA [6]. The core idea of this model is that an interaction between two nuclei can be described as a number of nucleon-nucleon interactions stacked on top of each other, with a special treatment of wounded nucleons that have already interacted, where the individual interactions can be simulated by PYTHIA's existing minimum-bias machinery.

Event generation in ANGANTYR begins with a Glauber modelling of subcollisions [46]. First, the spatial distribution of nucleons in each beam is determined, by default using the Glissando model [47], and the beam–beam impact parameter  $b_{AA}$  picked. By default,  $b_{AA}$  is sampled according to a Gaussian, which gives a bias towards lower values that is compensated by reweighting. This is in order to improve the statistics for central collisions at the expense of precision at peripheral collisions.

Knowing the impact parameter of the collision,  $b_{AA}$ , and the nucleon configurations of each beam, one can derive the nucleon–nucleon impact parameter  $b_{NN}$  for each nucleon pair. For each such pair, a potential subcollision type is chosen. The relative probabilities are functions of  $b_{NN}$  and the hadronic partial cross sections from Sect. 2.1. The exact dependence on the impact parameter is not well known, and ANGANTYR offers a few options. The most simplistic model is a black disk with a sharp division between the region for each type, starting with absorptive being the most central (here, “absorptive” basically means non-diffractive, except it may lead to diffractive topologies if either nucleon is wounded). That is, the collision is labelled as absorptive if  $\pi b^2 \leq \sigma_{\text{abs}}$ , double diffractive if  $\sigma_{\text{abs}} < \pi b^2 \leq \sigma_{\text{DD}}$ , and so on, continuing with single diffractive, and elastic scatterings being the most peripheral. If  $\sigma_{\text{tot}} < \pi b^2$ , the nucleon pair is labelled as not interacting.

A more sophisticated collision model, which is the default in ANGANTYR, is based on an approach by Strikman et al. [48–52], wherein the radius  $r$  of each nucleon is allowed to fluctuate. In proton–proton collisions, the sampling of the overlap is based on a parametrization fitted to data and thus already accounts for fluctuations. In heavy ion collisions, however, fluctuations have larger impact: when the projectile proton fluctuates to a large radius, it leads to more subcollisions, which in turn leads to a significantly longer tail in the multiplicity distribution that is difficult to model without such fluctuations. Especially in  $pA$  collisions, fluctuations in the projectile are more significant than fluctuations of nucleons in the target nucleus. Similarly to the black disk approach, more central collisions are more likely to be absorptive, but unlike the black disk, the division is continuous.

Specifically, the radius  $r$  of each nucleon follows a gamma distribution with mean radius  $r_0$  and shape parameter  $k_0$ . As the “size” of the nucleons is what determines the cross section, these parameters can be related to the total cross section according to

$$\sigma_{\text{tot}}/\pi = 4(k_0 + k_0^2)r_0^2, \quad (8)$$

and as the total cross section is given by PYTHIA’s hadronic model, we treat  $r_0$  as a dependent parameter. After sampling the projectile- and target radii  $r_p$  and  $r_t$ , the *opacity* is calculated as

$$T_0 = \left[ 1 - \exp\left(-\frac{\pi(r_p + r_t)^2}{\sigma_d}\right) \right]^\alpha, \quad (9)$$

introducing two new parameters,  $\sigma_d$  and  $\alpha$ , for a total of three free parameters overall. A given set of these parameters fixes the total and partial cross sections, which are already known from the hadronic collision model, constraining the possible values of the parameters. In ANGANTYR, the parameters are determined by a genetic algorithm. It begins by choosing random parameter sets, and selects the set that gives the best fit to the cross sections. Next, it produces random variations of this set, and again selects the best fit. This process is iterated for several generations (20 by default), eventually converging to a parameter set that accurately reproduces the cross sections. In addition, the elastic  $b$ -slope (appearing in  $d\sigma_{\text{el}}/dt \sim e^{bt}$ ) is used as a fitting target in this process.

After all subcollision candidates have been found they are sorted according to type, starting with absorptive collisions, followed by double diffractive, single diffractive, and finally elastic collisions. Within each process type, the collisions are sorted according to  $b_{NN}$ , starting with the most central. The collisions are then simulated in order, using PYTHIA to generate interactions to parton level. If a particle participates in several subcollisions, the subsequent ones are handled using the wounded nucleon model. When a wounded projectile particle collides with a target, the projectile emits a pomeron-like particle, which imparts some momentum on the beam remnants from the original collision. The secondary collision is then treated as a non-diffractive collision between the pomeron-like particle and the target, leading to a diffractive topology between the projectile and target.

The impact parameter  $b_{NN}$  determines the probability of each type of process, but  $b_{NN}$  also plays a different role in event generation, namely in determining MPI activity: in non-diffractive events, a smaller  $b_{NN}$  is correlated with a larger number of parton interactions. However, the different subcollision models may give different  $b_{NN}$  distributions within each type of collisions. For example, for a black disk where  $b_{NN}$  is assumed to be distributed uniformly on a disk with radius  $\sqrt{\sigma_{\text{abs}}/\pi}$ , the mean radius is  $\langle b_{\text{ND}} \rangle = \frac{2}{3}\sqrt{\sigma_{\text{abs}}/\pi}$ , whereas a fluctuating model could lead to an increased number of collisions with large impact parameter resulting as an increased  $\langle b_{\text{ND}} \rangle$  compared to black-disk one.

The MPI framework in PYTHIA [53] does not use the physical impact parameter values but the overlap  $\mathcal{O}(b)$  is parametrized in terms of scaled impact parameter  $b_{\text{MPI}}$  for which  $\langle b_{\text{ND}}^{\text{MPI}} \rangle = 1$ . Thus it is necessary to scale the  $b_{NN}$  values with  $\langle b_{\text{ND}} \rangle$  when passing this information to PYTHIA for MPI generation for a given nucleon–nucleon collision. The scaling is calculated as  $b_{\text{MPI}} = f \cdot b_{\text{NN}}/\langle b_{\text{ND}} \rangle$  where parameter  $f$  (0.85 by default) can be used to compensate for potentially different  $b$  distributions between ANGANTYR and the MPI framework in PYTHIA.



Once all subcollisions have been resolved this way, the resulting partons from each event are all combined into one overall event. Partons from different subevents may become correlated through colour reconnections [53], and optional features such as rope formation [54] and string shoving [55, 56] may be simulated at this point. Finally, the partons are hadronized and post-hadronization effects such as decays and rescattering are simulated.

### 2.3 Hadronic structure of quasi-real photons

One of the main motivations of developing the generic hadron-ion collision framework is that it allows to simulate an important aspect of the photon-nucleus collisions. As in the considered processes the photons have low virtuality, they may fluctuate into a hadronic state with equal quantum numbers. Thus to have a complete picture of these photon-induced processes, in addition to interactions with direct photons, also the component where photon has a resolved partonic structure has to be accounted for. The resolved photons consists of a component of perturbative origin, which can be directly calculated, and from a non-perturbative transition to a hadronic state. The latter component can be modelled with the VMD where the hadronic state is given by a linear combination of different vector mesons. While the direct photons have a significant contribution for high- $p_T$  observables, such as jets, it turns out that the bulk of the cross section is actually arising from collisions with resolved photons. Furthermore, as these collisions are essentially just like any other hadronic collisions, there can be several partonic interactions at the same collisions. A full-fledged framework for photoproduction with proton target is been included in PYTHIA 8.3 and described in detail in Ref. [1] and some applications considered in Refs. [5, 57].

In this work we focus on collisions where the interactions of quasi-real photons and protons or heavy nuclei generate a large number of final state particles. The production of these high-multiplicity events is dominated by the hadron-like part for the reasons explained above. Therefore, instead of including all components of real photons on one go, we implement an explicit VMD model and model real photons as a combination of hadronic states. Following Ref. [58], the total cross section for this contribution in photon-proton collision at a collision energy  $s$  is given by

$$\sigma_{\text{VMD}}^{\gamma p}(s) = \sum_V \frac{4\pi\alpha_{\text{em}}}{f_V^2} \sigma^{Vp}(s), \quad (10)$$

where the coupling factor  $4\pi\alpha_{\text{em}}/f_V^2$  is the probability for a photon to fluctuate into a vector-meson state  $V$  and the  $\sigma^{Vp}(s)$  the cross section for vector-meson-proton interaction. We use  $f_V$  values from Ref. [59] where  $f_\rho = 2.2$ ,  $f_\omega = 23.6$ ,  $f_\phi = 18.4$ ,  $f_{J/\psi} = 11.5$  and cross section parametrizations

follow the form in Eq. (3) with the state-dependent parameter values from Ref. [60]. The vector mesons state participating to the interaction is sampled according to the relative contributions of different vector-meson state to the total cross section at a given collision energy as given in Eq. (10).

### 2.4 Photon fluxes from different beams

All charged particles accelerated to high energies may emit photons but the energy spectrum and possible virtualities depend on the structure and size of the particle. Within the equivalent photon approximation (EPA) [61], derived in the limit  $Q^2 \rightarrow 0$ , the photon flux can be factorized from the hard scattering. In case of point-like charged leptons the photon flux,  $f_\gamma$ , can be calculated from the well known Weizsäcker-Williams formula [62, 63]

$$f_\gamma^l(x, Q^2) = \frac{\alpha_{\text{em}}}{2\pi} \frac{1 + (1-x)^2}{x} \frac{1}{Q^2}, \quad (11)$$

where  $x$  is the momentum fraction carried by the photon wrt. the beam particle emitting the photon and  $Q^2$  the (space-like) virtuality of the photon. Photoproduction is usually separated from deep-inelastic scattering events by requiring an upper limit for the virtuality  $Q_{\text{max}}^2 = \mathcal{O}(\text{GeV})$  to ensure that the EPA is applicable. The lower limit for the virtuality follows from kinematics and is proportional to the lepton mass, see details for the PYTHIA implementation in Ref. [1].

In case of charged hadrons and nuclei, one needs to take into account the finite size of a particle emitting the photon by including a form factor into the flux calculation. The form factor ensures that the emitting particle stays intact and suppresses the virtuality of the photons. In case of protons the photon flux rapidly vanishes beyond  $Q^2 > 1 \text{ GeV}^2$  whereas with heavy ions the virtualities are restricted to some tens of MeVs [10]. In case of heavy ions, a good approximation of the photon flux in the impact parameter space  $b$  can be calculated by assuming a point-like charge and rejecting interactions with small enough impact parameter where also strong interactions would be possible. Usually the cut in impact parameter space  $b_{\text{min}}$  is taken to be around  $2R_A$ , where  $R_A$  is the radius of the colliding nuclei  $A$ . Integrating the flux over the allowed impact-parameter space give the flux as

$$f_\gamma^A(x) = \frac{\alpha_{\text{em}} Z^2}{\pi x} \left[ 2\xi K_1(\xi) K_0(\xi) - \xi^2 (K_1^2(\xi) - K_0^2(\xi)) \right], \quad (12)$$

where  $Z$  is the charge of the nucleus,  $K_i$  modified Bessel functions of the second kind and  $\xi = b_{\text{min}} x m_N$ ,  $m_N$  being the per-nucleon mass. In the presented results for UPCs in the Pb-Pb collisions at the LHC we use the flux in Eq. (12) with  $b_{\text{min}} = 13.272 \text{ fm}$ ,  $m_N = 0.9314 \text{ GeV}$  and  $Z = 82$ .

### 3 Modelling hadron-ion interactions

#### 3.1 Subcollisions with generic hadrons

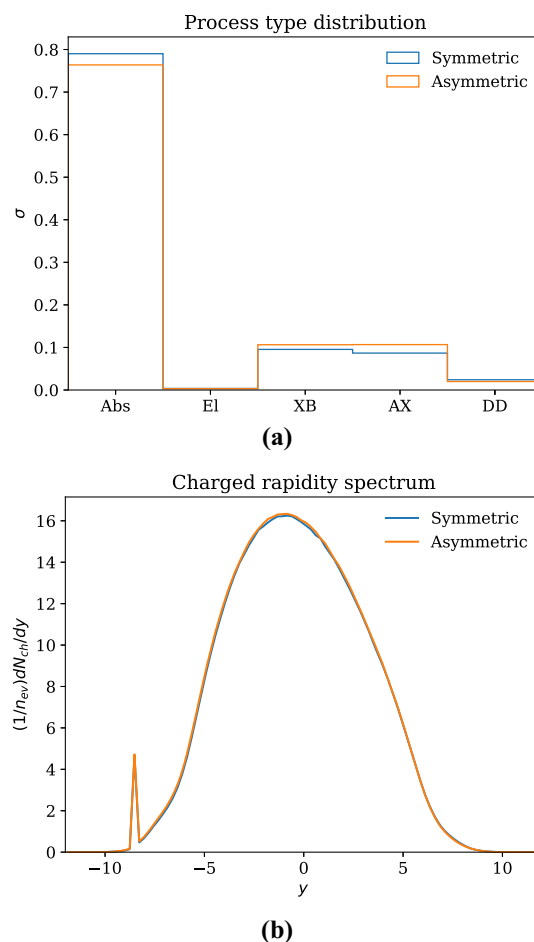
Our goal is now to extend ANGANTYR and implement  $hA$  collisions for generic projectile hadrons  $h$ . To this end, many of the components we need already exist in PYTHIA/ ANGANTYR. In fact, once the subcollisions are all determined, the hadronic interactions are fully implemented in PYTHIA. Furthermore, in the wounded nucleon model, the momentum imparted by the pomeron-like particle on the beam remnants of the wounded hadron follows the same distribution regardless of the species of the original beam hadron [6]. After subcollisions have been generated, we are left with a partonic state that is also agnostic to the original hadrons, and the existing framework is completely applicable after this point. The only unresolved question is then how to determine subcollisions for different projectile hadrons.

As all partial cross sections are given by PYTHIA, in the black disk model we can simply apply the sharp thresholds in the impact parameter to determine each collision type. Complications arise in the fluctuating model. While it is possible to assume only three free parameters as before, the projectile and target hadrons may now be of different species, and thus can have different parameters for fluctuations. Assuming the projectile/target each follow a gamma distribution with mean radius  $r_{p/t,0}$  and shape parameter  $k_{p/t,0}$ , the relation in (8) instead becomes

$$\frac{\sigma_{\text{tot}}}{\pi} = (k_{p,0} + k_{p,0}^2)r_{p,0}^2 + (k_{t,0} + k_{t,0}^2)r_{t,0}^2 + 2k_{p,0}k_{t,0}r_{p,0}r_{t,0}. \quad (13)$$

Although it is possible in principle to solve this for one of the parameters, say  $r_{t,0}$ , it is a second degree equation, which means that real solutions only exist in some regions of parameter space. For the sake of simplicity, we ignore this relation and simply work with a rectilinear parameter space with six independent parameters,  $k_p$ ,  $r_p$ ,  $k_t$ ,  $r_t$ ,  $\sigma_d$ , and  $\alpha$ . The drawback of this is that with more parameters, the number of iterations required for a converged fit in the evolutionary algorithm increases significantly.

In Fig. 1, we evaluate the effect of the six-parameter model compared to one where the two sides are assumed to have the same parameters, in the case of  $\pi^0\text{Pb}$  collisions at 5.02 TeV. After training, the goodness of fit was  $\chi^2/\text{dof} = 2.6$  for the symmetric case and 6.4 for the asymmetric case, despite the fact that for the symmetric case, the model was trained for 30 generations, while the asymmetric case needed more than 50,000 generations to achieve this level of precision. In the asymmetric case, we observed that the precision plateaued at several times during the training, indicating that the model found and stayed at a local minimum in parameter space until random variations found a better point. Considering Fig. 1a,



**Fig. 1** Comparison between model with symmetric and asymmetric parameters for  $\pi^0 p$  collisions at 5.02 TeV. **(a)** Relative composition of event types (absorptive, elastic, projectile diffraction (XB), target diffraction (AX) and double diffraction). **(b)** Charged rapidity spectra

one advantage of the second model is that we see an asymmetry between the rate of diffraction in the projectile and target. This difference cannot be captured by the symmetric model because it gives equal cross sections for diffraction on each side, even when the target cross sections are different. However, the difference here is small enough compared to the overall uncertainty in the cross sections that it has little practical value. Indeed, as we see in Fig. 1b, the actual observables are virtually indistinguishable.

In order to properly handle the effects of independent fluctuations, one would have to develop the underlying model further and choose fitting targets that are sensitive to the individual fluctuations, e.g. the expected mean radius of each hadron. In the current version, given the similarity of the results between the symmetric and asymmetric models and the large computational overhead in the latter, we conclude that there is no clear reason to favour the asymmetric model over the symmetric one.

### 3.2 Model tests

As a basic model test, the charged multiplicity and rapidity distributions for different  $h$ Pb combinations are shown in Fig. 2. In the charged multiplicities with default parameters (Fig. 2a), we see a bimodal distribution, particularly prominent in the  $J/\psi$  case. Furthermore, we see a convoluted progression where the  $\phi$  peak is below both  $\rho^0$  and  $J/\psi$ , even though one might expect a hierarchical relationship between  $\rho^0$ ,  $\phi$  and  $J/\psi$  as they contain progressively heavier  $q\bar{q}$  pairs. The reason for these behaviours is that with the fluctuating model, hadrons like  $J/\psi$  would sometimes fluctuate to very large radii, resulting in an unnaturally large  $\langle b_{\text{ND}} \rangle$ . In turn, this gives a very small  $b_{\text{MPI}}$ , basically treating all collisions as head-on for the purposes of calculating MPI activity, and all absorptive collisions are pushed towards higher multiplicities, giving a prominent second peak.

In order to circumvent this issue, we offer an alternative where the average impact parameter is calculated under the black disk assumption,

$$\langle b_{\text{ND}} \rangle = \frac{2}{3} \sqrt{\sigma_{\text{abs}}/\pi}, \quad (14)$$

but still allowing hadron size to fluctuate and maintain some of the multiplicity tail. The resulting multiplicities, shown in Fig. 2b, have a smoother behaviour and a clear hierarchy between  $\rho^0$ ,  $\phi$  and  $J/\psi$ . Applying the  $\langle b_{\text{ND}} \rangle$  from the black-disk approximation could potentially underestimate the result impact-parameter values and the variation between the two options could be to some extent taken as a modelling uncertainty.

In the rapidity spectra, shown in Fig. 2c and d, we see more clearly how the two different models affect the average multiplicity. For  $J/\psi$ , the effect is massive, with the two differing by a factor of roughly 4. For  $\rho^0$  and  $\phi$  the difference is around a factor 2, which is still large, but in line with other model uncertainties such as PDFs. The rapidity spectra also show a shift due to additional subcollisions, especially clear in the proton case where one would normally expect  $pp$  rapidity spectra to be symmetric. For the heavier  $\phi$  and  $J/\psi$ , there are fewer subcollisions and the projectile itself is heavier, and therefore lead to a lesser shift, as expected.

### 3.3 Cross sections

Cross sections for various hadron-lead combinations at 5.02 TeV are shown in Table 1, together with the corresponding hadron-proton cross sections from the DL/SaS models. As a first naïve estimate, one might expect that the  $hA$  cross section should equal the  $hN$  cross section multiplied by the nucleon number  $A$  of the ion, e.g.  $A = 208$  for lead. In practice, however, the target nucleons are bunched up, causing a shadowing effect. That is, in a central collision where

the projectile has several chances to interact with nucleons in the target, a collision is almost guaranteed. Hence, we see that even though the  $J/\psi$  cross section is almost an order of magnitude smaller than the  $pp$  one, the corresponding nuclear cross sections differ by less than a factor of 2.

Nuclear cross sections with an Oxygen target are shown in Fig. 3a for  $\pi^+O$  and  $pO$  as a function of  $p_{\text{Lab}}$ , comparing to the values calculated by Sibyll 2.3d [34], EPOS-LHC [35], and QGSJet-II.04 [36]. Whereas a comparison with the earlier toy model in PYTHIA gave results that were inconsistent with the other models [37], we now see a much closer match. Note that in PYTHIA 8.312 there is a numerical issue in ANGANTYR where event generation fails at the highest lab momenta in Fig. 3a, which we circumvented by calculating the cross section in the corresponding CM frame, where the numerical values are smaller and boosts more numerically stable.

At the lowest energies, ANGANTYR gives a higher  $\pi O$  cross section than the other models, but the discrepancy is overall quite small. This cross section then grows noticeably more slowly than in the other models, in accordance with how the  $\pi^+p$  cross section in PYTHIA grows more slowly than the other models [37]. A similar trend is seen for  $pO$ , but here the difference between the hadronic cross sections is smaller, suggesting there may be other model effects at play. However, since the difference is very small, we have not investigated this further.

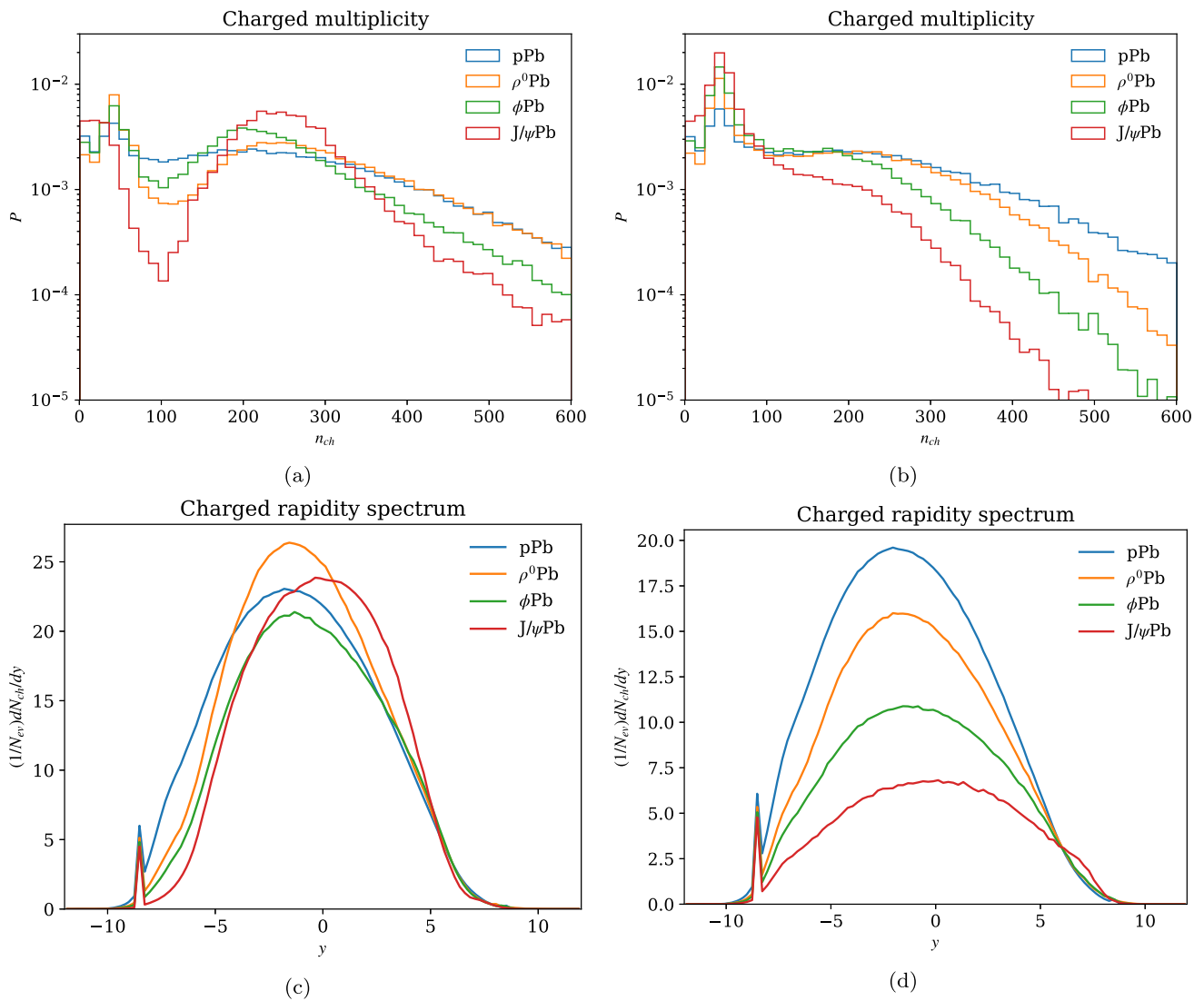
Once the nuclear configuration and impact parameter have been determined and it has been decided that at least one  $hN$  collision occurs, the number of additional nucleons within range scales roughly with the total hadronic cross section (in the case of a uniform target, the scaling would be exact). Therefore, we expect that the mean number of subcollisions roughly follows the shape  $\langle n_{\text{coll}} \rangle = 1 + k\sigma_{\text{tot}}$ , where  $k$  represents the density of nucleons in the projection of the nucleus onto the  $xy$ -plane. This is shown in Fig. 3b where a line with a fitted value for  $k$  capture the trend of the calculated points well.

### 3.4 Variable energies

In photon-induced VMD processes, the available energy in the hard interaction changes on an event-by-event basis, depending on the photon flux. The energy also varies in hadronic cascades, with less energy further down in the cascade. Thus, there is a technical need for efficiently changing the energy in each event. Moreover, there is similarly a need to switch between different beam particles. In VMD processes, one needs to switch between different vector mesons, while in hadronic cascades, one must include all hadrons that are long-lived enough to rescatter in the medium.

This feature is implemented in PYTHIA hadron-hadron. It requires a number of technical steps in the code, most signif-





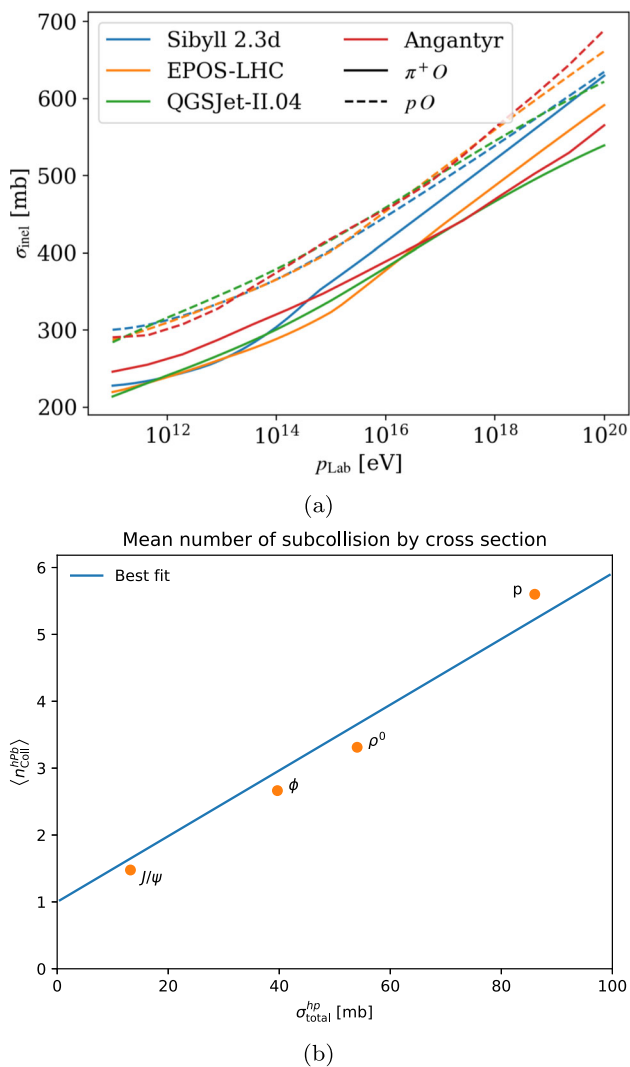
**Fig. 2** Comparison of observables for various hadron-proton and hadron-lead beam combinations at 5.02 TeV. The left-hand figures correspond to an  $\langle b_{\text{ND}} \rangle$  calculated by the fluctuating model, while the fig-

ures on the right use the black disk ( $b_{\text{ND}}$ ). The spikes around  $y = -8$  in the rapidity spectra correspond to the nuclear remnant, constructed as a combination of spectator nucleons

**Table 1** Hadron–proton and hadron–lead cross sections for different hadron species at 5.02 TeV, in units of mb. The first column corresponds to  $h\text{Pb}$ , and subsequent columns are  $hp$ . Note that  $hA$  cross

sections cannot easily be calculated a priori in ANGANTYR, and were calculated using Monte-Carlo integration

$h$	$\sigma_{\text{tot}}^{h\text{Pb}}$	$\sigma_{\text{tot}}^{hp}$	$\sigma_{\text{ND}}^{hp}$	$\sigma_{\text{diff}}^{hp}$	$\sigma_{\text{el}}^{hp}$
$p$	$2.05 \times 10^3$	86.0	47.7	19.9	18.4
$\rho^0$	$1.95 \times 10^3$	54.0	33.5	12.5	8.0
$\phi$	$1.81 \times 10^3$	39.7	26.9	8.5	4.3
$J/\psi$	$1.26 \times 10^3$	13.2	10.5	2.2	0.5



**Fig. 3** (a) A comparison of ANGANTYR inelastic cross sections to corresponding cross sections from other models [37], ranging from  $p_{Lab} = 10^{11}$  eV ( $E_{CM} \sim 13$  GeV) to  $10^{20}$  eV ( $E_{CM} \sim 250$  TeV). (b) Average number of subcollisions in  $hPb$  collisions at 5.02 TeV. The best fit is assumed to be on the form  $1 + k\sigma_{tot}$

icantly handling energy-dependent MPI parameters. These parameters are calculated at the beginning of the run and constitute the slowest part of initialization, which makes it unfeasible to perform this calculation every time the beam energies change. PYTHIA solves this by calculating the parameters at several energy values during initialization, then using interpolation to determine values at a given energy during event generation.

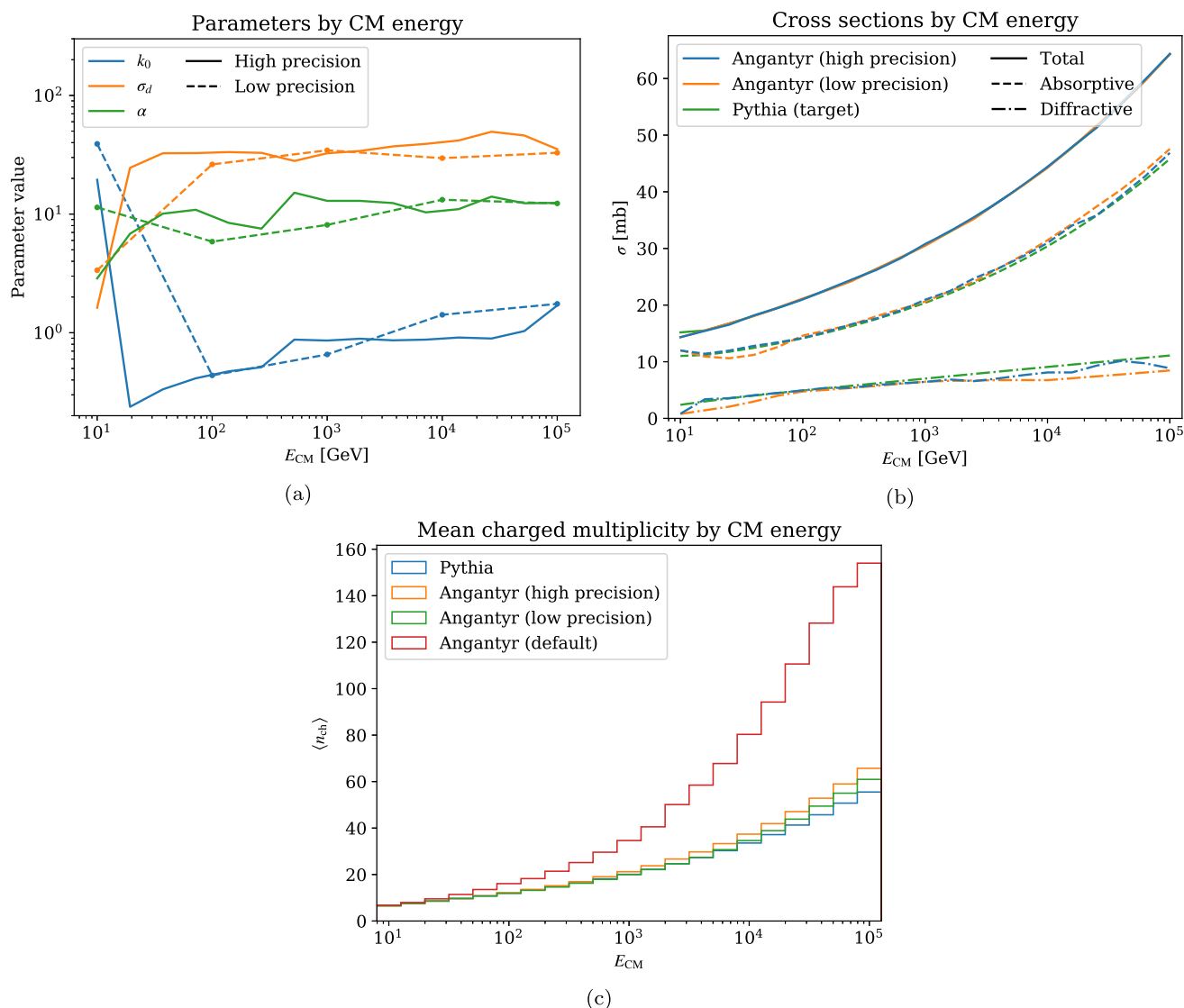
Implementing variable energies in ANGANTYR is the same story. Here, the bottleneck is the evolutionary algorithm used to determine nuclear fluctuations. We again solve it by calculating the subcollision parameter values during initialization and interpolating during event generation. As an additional optimization, we assume the parameters vary relatively

slowly with energy, so that by using the parameters at each interpolation point as the initial values when running the algorithm at the next point, we can achieve a slightly faster convergence. To also include beam switching, the framework automatically handles initializing for different hadron types, saving the initialization data to one file for easy reuse.

The question is then how many generations and interpolation points are necessary. We investigate this in Fig. 4, where we consider  $\phi p$  collisions at energies ranging from  $E_{CM} = 10^5$  GeV down to 10 GeV, comparing a case with 5 interpolation points to one with 21 points (recall that the parameters are only a function of the hadron species, i.e. the parameters will be identical for all  $\phi A$ ). In Fig. 4a, we do see a noticeable discrepancy between the parameter values. However, the fitting procedure might find different local minima in parameter space (as indicated by point-by-point fluctuations in the high-precision fitting), and this discrepancy in itself is not cause for concern. What matters is the resulting cross sections and observables. The cross sections in Fig. 4b show that the two models give similar results, with the high-precision run giving a marginally better fit to the target.

Figure 4c shows the mean charged multiplicities as a function of the CM energy. Here, the high- and low-precision runs use the black disk  $\langle b_{ND} \rangle$  as described in Eq. (14), and we compare to the case where  $\langle b_{ND} \rangle$  is calculated from the fluctuations (labelled “ANGANTYR (default)”). First, we see that the high- and low-precision runs both are close to the PYTHIA target, and in fact the low-precision version gives a closer match. We have also validated this for  $\rho^0$  and  $J/\psi$ , and conclude that one interpolation point per decade gives a precision that is well within other model uncertainties. With the fluctuating  $\langle b_{ND} \rangle$ , the discrepancy grows large, especially at higher energies relevant for cosmic ray applications. We therefore recommend the black disk  $\langle b_{ND} \rangle$  for such use cases until a more generic model for hadronic fluctuations capable to handle these very high energies is developed.

In the preceding studies, we have gone all the way down to 10 GeV where we expect to see non-perturbative effects such as resonance formation. While default PYTHIA can model hadronic interactions at low energies, it is not immediately clear how to extend this to heavy ions. In particular, all low-energy processes implemented in PYTHIA have purely hadronic final states with no beam remnants, so it is not straightforward to model secondary subcollisions with the wounded nucleon model. Furthermore, a sudden variance in the parameters at low energies in Fig. 4a indicates that the genetic algorithm might not be as accurate in this region. As a brute force approximation, one might simply assume that hadrons wounded in non-perturbative interactions cannot participate in further collisions. While this would not necessarily be an incorrect physical picture, we note that such a hadron-ion collision would be indistinguishable from



**Fig. 4** Comparison of a low-precision cross section fit to one with much higher precision for  $p p$  collisions. **(a)** Parameters output from the genetic algorithm; **(b)** resulting cross sections, compared to target cross sections; **(c)** mean charged multiplicity in  $p p$  collisions as a function

of energy, comparing different model choices. Here “default” refers to using  $\langle b_{ND} \rangle$  calculated from fluctuations, while the other ANGANTYR plots use the black disk  $\langle b_{ND} \rangle$

a hadron-nucleon collision, except for the addition of nuclear remnants.

We have not implemented a way to do this automatically in ANGANTYR, but a user who wishes to do this can work around it by initializing another PYTHIA object that is configured to generate only low energy events, and using that object to generate a fraction of the events. Specifically, in PYTHIA this fraction is given by  $P_{\text{nonpert}} = \exp\left(-\frac{E_{CM}-E_{\text{min}}}{E_{\text{width}}}\right)$ , where the parameters are by default  $E_{\text{min}} = E_{\text{width}} = 10$  GeV. This gives a fraction of around 2% at 50 GeV, above which we consider low-energy processes negligible.

## 4 Applications

### 4.1 ZEUS data for high-multiplicity events in photoproduction

A recent ZEUS analysis studied azimuthal correlations in high-multiplicity events in DIS and photoproduction [30] in electron-proton collisions at HERA. In addition to two- and four-particle correlations, the analysis presented results for charged-particle multiplicity and rapidity distributions in photoproduction with a proton target that can be used to validate the vector-meson-dominance model considered in this work. In the analysis, the experimental data were com-

pared to PYTHIA simulations for photoproduction by varying the value of  $p_{T,0}^{\text{ref}}$  that determines the probability for MPIs in case of resolved photons and found that values around  $3 < p_{T,0}^{\text{ref}} < 4$  GeV were preferred. This maps to slightly lower MPI probability wrt. what has been found optimal in proton-proton collisions [64] but also the considered energy range is quite different. In any case, the data clearly disfavored results where MPIs were disabled and is in line what has been obtained from a comparison to charged-particle  $p_T$  spectra in  $\gamma\gamma$  collisions at LEP [5].

In Fig. 5 we compare our PYTHIA simulations to the ZEUS measurement for charged-particle multiplicity,  $N_{\text{ch}}$ , and pseudorapidity,  $\eta$ , distribution and the  $p_T$  spectrum. The simulations include results with full photoproduction and with only the resolved part modelled with the VMD. In both cases we have varied  $p_{T,0}^{\text{ref}}$  to account for the underlying uncertainty in the modelling of MPIs. In case of  $N_{\text{ch}}$  we notice that both full photoproduction and VMD calculation enclose the data within the  $p_{T,0}^{\text{ref}}$  variation and that impact of this variation is somewhat smaller in case of VMD setup with a slight preference on  $p_{T,0}^{\text{ref}} = 3$  GeV. In case of rapidity distribution the variations with different contributions are found small and well in line with the data. Only in case of the  $p_T$  spectrum do we find significant differences between full photoproduction and VMD only, where the former is again enclosing the data within  $p_{T,0}^{\text{ref}}$  variations whereas the latter seem to fall below the data at high values of  $p_T$ . This likely follows from differences in PDFs for vector mesons and resolved photons, as the latter tend to have more large- $x$  partons. Additionally we show also the break down of events from the different contributions for the multiplicity distributions with the default photoproduction setup in Fig. 5. We find that when event selection does not require any high- $p_T$  particles, the direct contribution is at most around 10% and negligible at high multiplicities. We conclude that the VMD-only setup is sufficient to describe the event structure for high-multiplicity events we focus on this study and that it provides a reasonable baseline to extend this to collisions with a heavy-ion target.

#### 4.2 Ultra-peripheral photo-nuclear collisions at the LHC

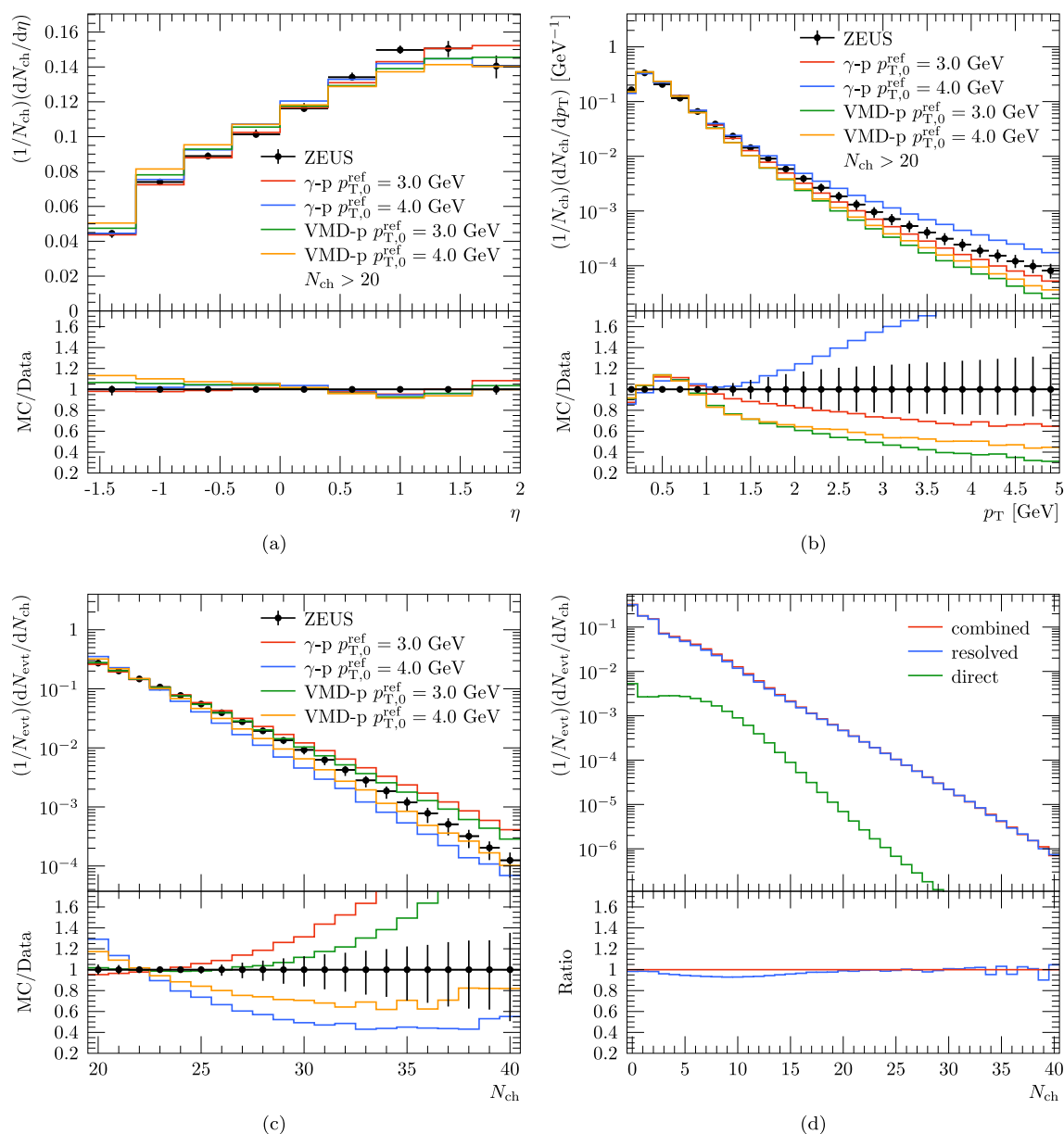
The main application of our hadron-ion modelling is to enable simulations of photon-ion collisions applying the VMD model within PYTHIA. Here we focus on high multiplicity UPC events recently analysed by the ATLAS collaboration [19]. The caveat of the data provided by the experimental analysis is that the shown distributions are not corrected for the finite acceptance and only the Fourier coefficients fitted to two-particle correlations are publicly available. This prevents direct comparisons, but in order to draw at least qualitative conclusions based on the data, we have adjusted the calculated MC multiplicity,  $N_{\text{ch}}$ , to the experimentally

reconstructed charged-particle multiplicity  $N_{\text{ch}}^{\text{rec}}$  by scaling the former with a factor 0.8. This acceptance correction was estimated based on the PYTHIA simulations for  $\gamma$ -p presented in ATLAS publication including the limited acceptance and matching this with simulations including all particles with similar event selection and PYTHIA configuration. Similar efficiencies have been reported earlier by ATLAS in other collisions systems [22] though instead of a constant factor some dependence on  $\eta$  was observed.

In addition to the efficiency correction, we perform a similar event selection criteria as quoted by the experimental study including a cut on sum-of-rapidity-gap measure  $\Sigma_{\gamma} \Delta\eta$  of 2.5. This is calculated by summing over all rapidity gaps larger than 0.5 in the range  $0 \leq \eta \leq 4.9$  in the photon-going direction. The resulting  $\Sigma_{\gamma} \Delta\eta$  distributions with the default VMD-Pb and photoproduction setup with proton target are shown in Fig. 6 together with the invariant mass distribution of the  $\gamma$ -nucleon system,  $W_{\gamma p}$ , when a multiplicity cut of  $20 < N_{\text{ch}}^{\text{rec}} < 60$  is applied. In case of the rapidity-gap measure we notice that the results with nuclear and nucleon targets are close to each other which can be expected since the experimental data was well in line with the simulations with proton target in the  $\Sigma_{\gamma} \Delta\eta > 2.5$  region where photon-originated processes dominate. Similarly there are no big differences in the distribution of the photon-nucleon centre-of-mass energy  $W_{\gamma p}$  when applying cuts for the event multiplicity. In both cases the minimum centre-of-mass energy to produce events with  $N_{\text{ch}}^{\text{rec}} > 20$  is found to be around 200 GeV. The average  $W_{\gamma p}$  distribution is slightly lower than with the VMD-Pb setup compared to simulation with proton target only which follows from the increased multiplicity for events with nuclear target. There is also a correlation between the event multiplicity and centre-of-mass energy as the average  $W_{\gamma p}$  is around 470 GeV for the reconstructed multiplicities at 15 – 20 and at 570 GeV for  $20 < N_{\text{ch}}^{\text{rec}} < 60$ .

In Fig. 7 we include results with different subcollisions models included within the ANGANTYR setup. Two of these include the cross section fluctuations with an average impact-parameter for non-diffractive events taken from the black-disk approximation ( $b_{\text{ND}}^{\text{BD}}$ ) or from the fluctuating default model ( $b_{\text{ND}}^{\text{fluc}}$ ). In the black-disk model the nucleon radius and cross section fractions are fixed. To have a baseline to which the effects from a nuclear target can be compared, we also include results using the default photoproduction setup for  $\gamma$ -p that corresponds to the PYTHIA results in the ATLAS publication [19]. We present results for per-event yields as a function of  $N_{\text{ch}}$ , and  $\eta$ . With all available nuclear collision models we notice that the multiplicities are increased compared to the simulation with just a proton target which is explained by subsequent interactions between the projectile and target nucleus.

For the two models including cross section fluctuations we find similar shape in the multiplicity distribution, but some-



**Fig. 5** (a) Pseudorapidity, (b)  $p_T$  spectra and (c) multiplicity distribution from full photoproduction and VMD model with different  $p_{T,0}^{\text{ref}}$  compared to ZEUS data taken from electron-proton collisions at

$\sqrt{s} = 318 \text{ GeV}$  [30], and (d) multiplicity distribution with contributions from direct- and resolved-photon originated processes

what lower tail of high-multiplicity events in case  $\langle b_{\text{ND}} \rangle$  calculated from the black-disk approximation. Keeping in mind that the multiplicities shown by ATLAS were not corrected for the limited efficiency, the simulated multiplicity distributions including fluctuations seem to enclose the measured distribution. However, the fixed-radii black-disk model high-multiplicity tail seems to fall short compared to data.

The observations are very similar for the pseudorapidity distributions where the fluctuating subcollision models predict a larger increase of particle production at negative rapidi-

ties as the fixed-radii model and the results with only a proton target. Again the data would seem to prefer models including cross-section fluctuations but, as above, the observations are only indicative due to missing unfolded data.

In Fig. 8 we compare the simulations with varying  $p_{T,0}^{\text{ref}}$  using similar values for the parameter as in case of HERA data in Fig. 5 using the fluctuating subcollision model with  $\langle b_{\text{ND}} \rangle$  from black-disk approximation and comparing this with the results from applying photoproduction in  $\gamma$ -p. For the multiplicity distribution we find some dependence on this



MPI parameter, but interestingly the effect is significantly reduced compared to the effect seen with a proton target. The reason for this is that now the multiplicity of an event is not only driven by the number of MPIs in a single collision between the projectile and a nucleon, but also by the number of collisions between the projectile and target nucleons. The latter seems to be the dominant mechanism at high multiplicities but some effects from the  $p_{T,0}^{\text{ref}}$  variation is seen around  $10 < N_{\text{ch}}^{\text{rec}} < 50$ . Similarly, only small effects from  $p_{T,0}^{\text{ref}}$  variation are found for the  $\eta$  distribution considered in Fig. 8. Thus we conclude that while the probability of MPIs in case of real photons has some uncertainty, the results with a heavy-ion target are fairly robust against this.

### 4.3 Collectivity in $\gamma$ -Pb

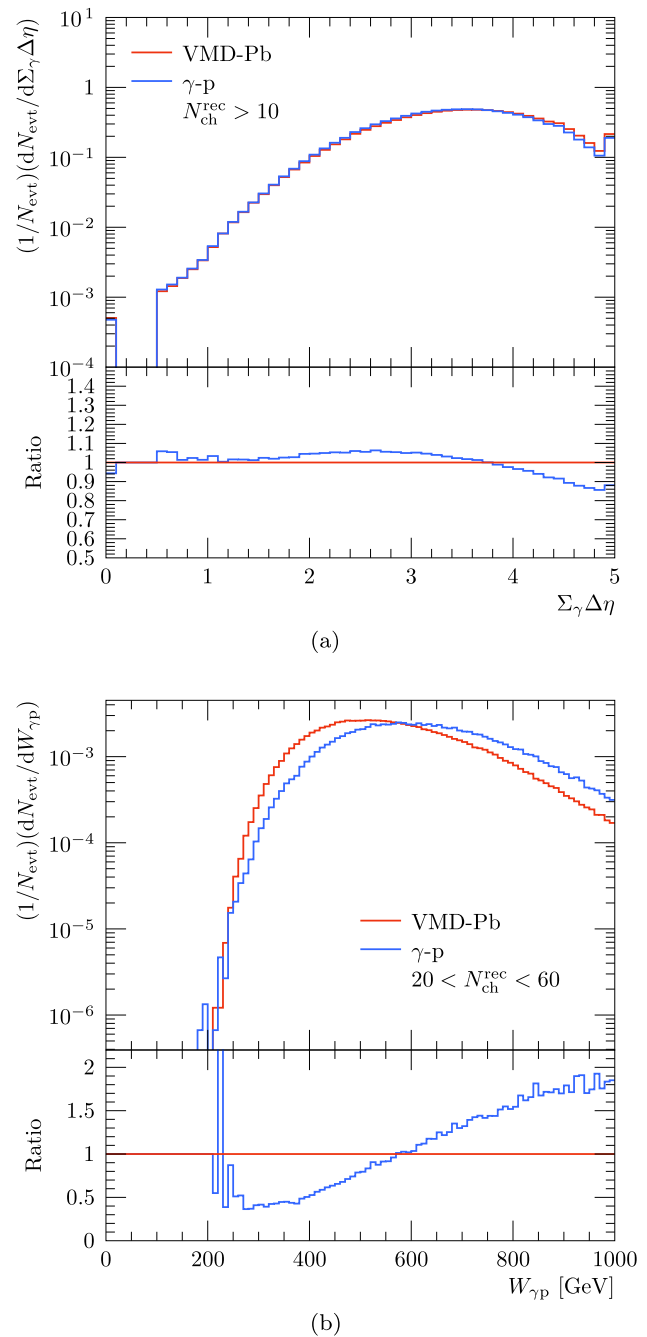
Having a framework to simulate complete UPC events we can apply this also to study many-particle correlations. A particularly interesting observation in the ATLAS analysis [19] was a finite value for the Fourier coefficients in high-multiplicity UPC events from the template fitting where the correlations in low-multiplicity events are subtracted from the high-multiplicity ones. In the UPCs, there is a clear correlation between the collision energy  $W$  and the multiplicity, so it is not trivial whether the template fit would result as zero  $v_n$  even if there are no collectivity from explicit final-state interactions. To verify this in detail, we have performed a similar template fit with the simulated events, using again the fluctuating subcollision model with  $\langle b_{\text{ND}} \rangle = b_{\text{ND}}^{\text{BD}}$ . The calculated two-particle yield as a function of the azimuth angle  $\phi$  and pseudorapidity separation between the two particles  $|\Delta\eta|$ ,  $Y(\Delta\phi, |\Delta\eta|)$ , is integrated over  $2.0 < |\Delta\eta| < 5.0$  to obtain the one-dimensional two-particle yield  $Y(\Delta\phi)$ . For the low-multiplicity (LM) events, a direct Fourier fit is performed by accounting terms up to order 4 in the Fourier series of the form

$$Y^{\text{LM}}(\Delta\phi) = c_0 + 2 \sum_{n=1}^4 c_n \cos(n \cdot \Delta\phi). \quad (15)$$

For high-multiplicity (HM) events a parametrization is used,

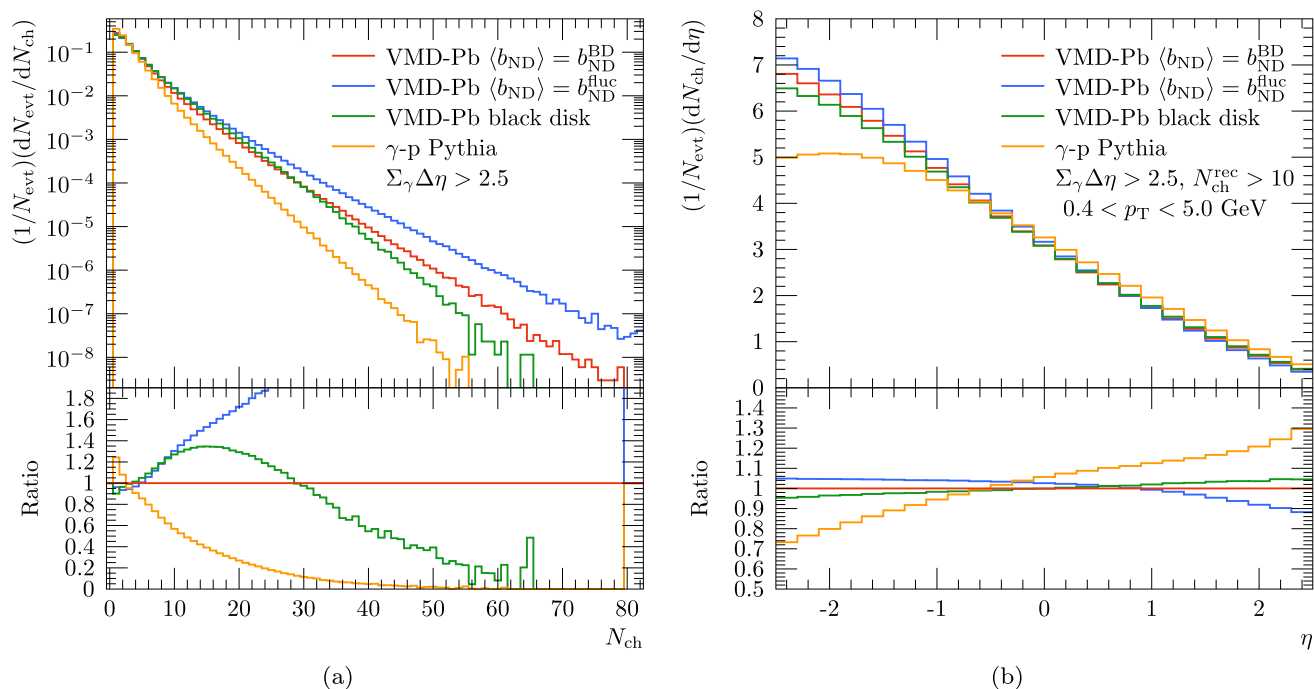
$$Y^{\text{HM}}(\Delta\phi) = F \cdot Y^{\text{LM}}(\Delta\phi) + G \left[ 1 + 2 \sum_{n=2}^4 v_{n,n} \cos(n \cdot \Delta\phi) \right], \quad (16)$$

where now  $F$  and  $G$  are free parameters, but are connected by the requirement of having the same value for the integral of both sides in Eq. (16). In total there are then 9 free parameters that are fitted simultaneously. In addition, we perform a direct Fourier fit to the HM sample using the truncated series in Eq. (15) to extract the Fourier coefficients without the template fitting and a separate 5-parameter Fourier fit to LM sample to check whether the simultaneous template fit

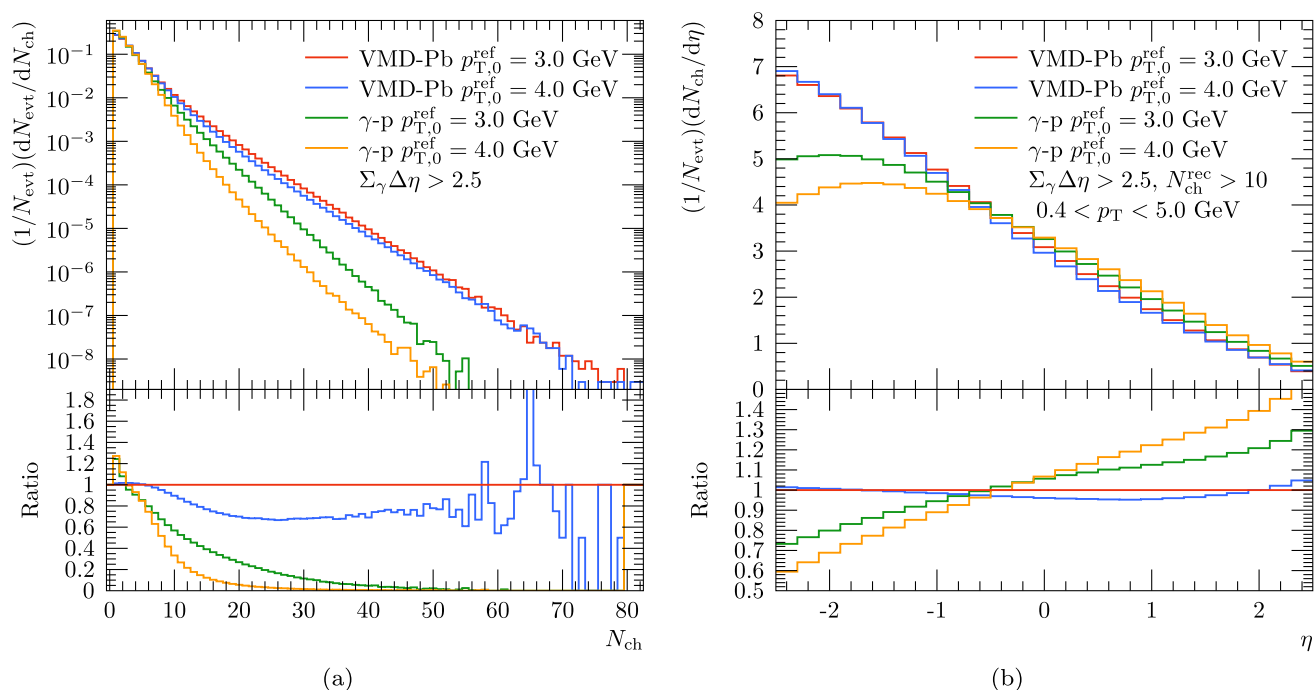


**Fig. 6** (a) Rapidity gap measure applied by ATLAS [19] and (b) invariant mass of the vector meson-nucleon system from VMD-Pb configuration and photoproduction with proton target from PYTHIA

modifies the values of the Fourier coefficients. An example of this procedure is shown in Fig. 9 where the LM sample has  $15 < N_{\text{ch}}^{\text{rec}} < 20$ , the HM sample has  $20 < N_{\text{ch}}^{\text{rec}} < 30$  and the  $p_T$  of the both particles is between 0.4 GeV and 2.0 GeV. In case of a direct Fourier fit we identify  $c_n = v_{n,n}$ . For the HM fit we plot also part  $G + F \cdot Y^{\text{LM}}(\Delta\phi)$  which provides the scaled LM reference to which the additional HM modulation can be compared to. The simulated results have been



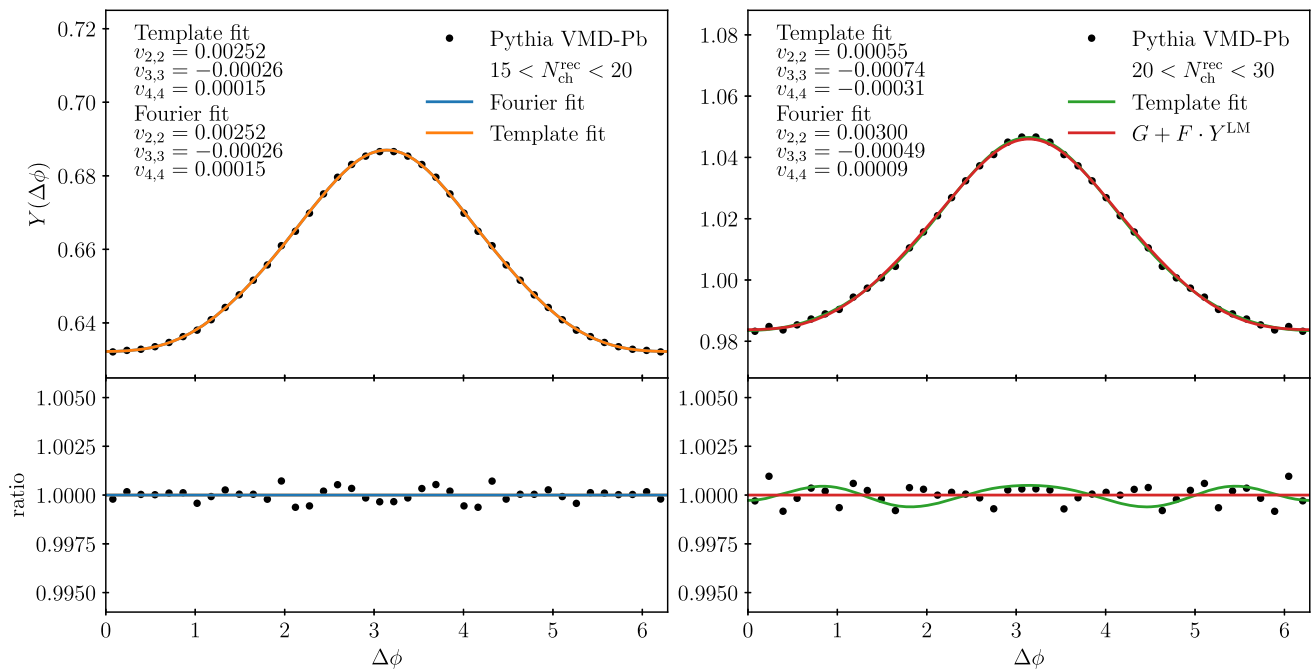
**Fig. 7** PYTHIA results for **(a)** per-event multiplicity and **(b)** rapidity distribution. For the VMD-Pb configuration results with different assumptions for ANGANTYR subcollision model are compared to photoproduction with proton target



**Fig. 8** PYTHIA results for **(a)** per-event multiplicity and **(b)** rapidity distribution. PYTHIA results include simulations with two different values for  $p_{T,0}^{\text{ref}}$  in case of VMD-Pb with the ANGANTYR and photoproduction with proton target

obtained with  $p_{T,0}^{\text{ref}}$  using the default subcollision model for the VMD-Pb simulations. We notice that for the LM sample both the template fit and the direct Fourier fit yields the same values for the parameters confirming that the LM events are

equally well described by both fits. In case of HM events we find that the scaled modulation from LM events is sufficient to describe the simulated two-particle yield as the values of Fourier coefficients  $v_{n,n}$  obtained from the Template fit are



**Fig. 9** Template fit to two-particle yields for low- (left) and high-multiplicity events (right) following the fit procedure in Ref. [19]. The multiplicity cuts have been adjusted to account for limited detector acceptance, see text for details

consistent with zero when accounting for the fit uncertainties related to finite statistics in the simulated events. This is also reflected by the ratio of the simulated events and the  $G + F \cdot Y^{\text{LM}}(\Delta\phi)$  where no further significant modulation remains. We find also that the Fourier coefficients from a direct Fourier fit are very similar in both event samples further reflecting similarity of the azimuthal modulation in LM and HM samples.

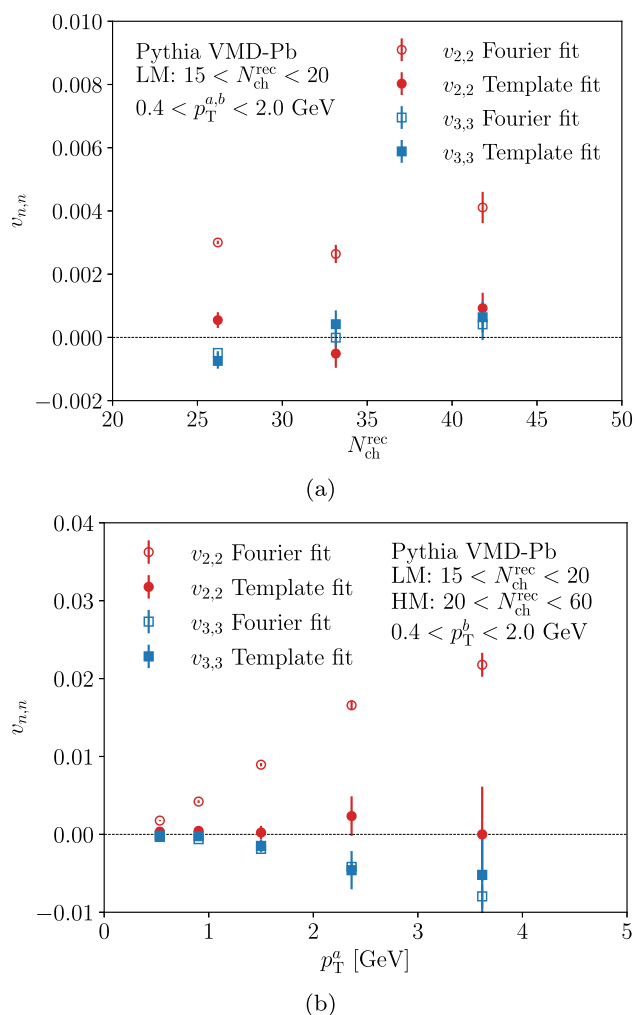
As in the ALICE UPC analysis [19], we have repeated the fitting procedure for different selections of high-multiplicity events and by varying the transverse momentum of the particle  $a$ . The similar event selection criteria as in the ATLAS analysis have been applied accounting also for the limited efficiency for  $N_{\text{rec}}^{\text{ch}}$  discussed above. The multiplicity and  $p_{\text{T}}^a$  dependencies are shown in Fig. 10. For the direct Fourier fits (open markers) we notice that the trends are very similar with the experimental results. The fitted  $v_{2,2}$  are of the same order and roughly constant as a function of  $N_{\text{rec}}^{\text{ch}}$  and increase linearly as a function of  $p_{\text{T}}^a$ . In case of  $v_{3,3}$  the values are consistent with zero apart from the pairs with the highest  $p_{\text{T}}^a$  values where the negative  $v_{3,3}$  are found, similar to the experimental data. In case of the template fit (filled markers) the clear difference to the measurement is that in all cases the resulting  $v_{n,n}$  values are consistent with zero whereas in the data finite and positive values for  $v_{2,2}$  and  $v_{3,3}$  were found from the template fit. This further confirms the observation from the result in Fig. 9 showing no room for additional modulation in HM events wrt. LM sample in our simulations. We conclude that our PYTHIA VMD-Pb simulations without

any explicit final-state interactions is not compatible with the ATLAS data where finite positive values for  $v_{n,n}$  were found after the template fit.

## 5 Summary and outlook

In this article, we have presented a framework for hadron-ion collisions in PYTHIA/ANGANTYR where the allowed projectile hadrons include pions, vector mesons and quarkonia. The full list of available hadrons match the previous study related to hadron-nucleon collisions [7]. The application considered in this study is modelling of photon-ion interactions in ultra-peripheral heavy-ion collisions using the vector-dominance model for the quasi-real photon. This framework can be also applied to study air showers created by high-energy cosmic rays when hitting to the Earth's atmosphere.

The implementation is built upon PYTHIA's existing framework for hadron-nucleon interactions, using partial hadronic cross sections as input for the Glauber subcollision modelling, and invoking PYTHIA to simulate hadron-nucleon interactions. The component that requires most new modelling is the hadronic fluctuations, which are calculated by a genetic algorithm fitting to partial cross sections. In existing ANGANTYR pA/AA, the fluctuation parameters are the same for both beams, and we extended this to include a model with asymmetric fluctuations for the general  $hA$  case. However, we found that using the asymmetric model gives only a small difference, especially compared to existing model uncertain-



**Fig. 10** The  $v_{n,n}$  values from the fit to PYTHIA events at (a) different multiplicities and (b) different  $p_T^a$  bins corresponding to ATLAS study in Ref. [19]. Error bars quantify the statistical uncertainty in the fit

ties such as in the cross sections and PDFs for uncommon hadrons. Considering that it takes significantly longer to run the genetic algorithm with additional parameters, we believe the symmetric model gives sufficient precision for now. In the future, a more accurate modelling of these fluctuations may be relevant e.g. for the tail shape of multiplicity distributions.

With the current fluctuation model, some hadrons (e.g.  $J/\psi$ ) would fluctuate to unreasonably large sizes, contradicting the expectation that it should have a narrower wavefunction. This cannot be regulated in the current model because the fit is only to the inclusive cross sections, and it would require making deeper changes to the approach in order to accurately model different hadronic fluctuations. As a consequence to these large fluctuations, the mean impact parameter  $\langle b_{ND} \rangle$  also becomes very large, which causes the MPI framework to treat all collisions as essentially head-on, leading to too high MPI activity. As a workaround, we presented the option to use the black disk value,  $\langle b_{ND} \rangle = \frac{2}{3} \sqrt{\sigma_{abs}/\pi}$ , but it

is clear that additional modelling is necessary to accurately represent fluctuations of generic hadrons.

This development also includes a technical feature for changing the energy on an event-by-event basis in ANGANTYR. For varying energies, one of the initialization bottlenecks is the genetic algorithm for hadronic fluctuations. We reduce the initialization time by saving the resulting parameters to disk at certain energies, and interpolating between these values whenever the energy is changed. Our study indicates that the parameters change slowly enough that roughly one interpolation point per order of magnitude is sufficient. As we move to energies below  $\sim 50$  GeV, a growing fraction of events will include non-perturbative effects such as resonance formation. These interactions are implemented in PYTHIA but not yet in ANGANTYR, and it is not clear how accurate our framework is below this energy.

In addition to switching energies, we now also allow switching the beam particles. This is relevant for example in VMD models where the species depends on which vector meson the photon fluctuates into, and in cosmic rays where the hadronic cascades includes a wide variety of species. When switching beams, the projectile is allowed to be any hadron or nucleus, while the target can be a proton, neutron or nucleus.

We compared the VMD model first in case of photon-proton collisions with the full photoproduction implementation in PYTHIA and with the experimental data from ZEUS for high-multiplicity photoproduction events. For the applied event selection the two models were in a good agreement with each other and the experimental data. In case of the multiplicity distribution we noticed significant differences when varying the value  $p_{T,0}^{ref}$  that controls the probability for the MPIs. With both models the data was, however, enclosed with  $p_{T,0}^{ref}$  range in line with previous studies. In case of photon-lead UPCs we found that the multiplicity distributions were not very sensitive to the value of this parameter, but the increased multiplicity when applying a nuclear target was rather driven by the number of collisions between the resolved photon and different nucleons in the nucleus. We found some dependence on the applied fluctuating cross-section model in ANGANTYR and noticed that the ATLAS data do seem to disfavour simulations when these fluctuations are completely turned off and a simplified black disk model is applied. We conclude that the presented framework for photon-ion collisions do have some uncertainties related to subcollision modelling but the generic features are in line with the limited data and future measurements with more direct comparisons would provide useful constraints.

Having a model that can produce complete final states in photon-ion collisions, we were able to test whether the observed collective behaviour in two-particle correlations could arise from simulations without including any explicit

final-state interactions. We noticed that while the direct Fourier fits to the obtained two-particle correlations were in line with ATLAS data, the  $v_{n,n}$  values from the template fit were consistent with zero. In case of the ATLAS data, they found finite values for these Fourier coefficients after subtracting correlations from the low-multiplicity events with the template fitting procedure. This suggests that further final-state interactions are required to match the experimental values for the coefficients quantifying collective behaviour. In case of PYTHIA context, such final-state effects could arise from string interactions such as rope hadronization and string shoving that we plan to revisit in a future study.

**Acknowledgements** We thank Leif Lönnblad and Christian Bierlich for useful discussions related to the ANGANTYR model and Sruthy Das for information related to ATLAS particle-reconstruction efficiency. We also thank Maximilian Reininghaus for sharing cross section data from the Sibyll, EPOS-LHC and QGSJet models. We acknowledge the financial support from the Research Council of Finland, Project No. 331545, and through the Centre of Excellence in Quark Matter. The reported work is associated with the European Research Council project ERC-2018-ADG-835105 YoctoLHC. We acknowledge grants of computer capacity from the Finnish Grid and Cloud Infrastructure (persistent identifier urn:nbn:fi:research-infras-2016072533).

**Data Availability Statement** This manuscript has no associated data. [Authors' comment: This manuscript has no associated data but the associated code can be used to generate the presented results.]

**Code Availability Statement** This manuscript has associated code/software in a data repository. [Authors' comment: This manuscript has associated code available at <https://pythia.org> and in a data repository at <https://gitlab.com/Pythia8/releases>.]

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. Funded by SCOAP<sup>3</sup>.

## References

1. C. Bierlich et al., A comprehensive guide to the physics and usage of PYTHIA 8.3. SciPost Phys. Codeb. (2022)
2. J.C. Collins, D.E. Soper, G.F. Sterman, Factorization of hard processes in QCD. Adv. Ser. Direct. High Energy Phys. **5**, 1–91 (1989)
3. K. Kovarik, P.M. Nadolsky, D.E. Soper, Hadronic structure in high-energy collisions. Rev. Mod. Phys. **92**(4), 045003 (2020)
4. S. Höche, Introduction to parton-shower event generators, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, pp. 235–295 (2015)
5. I. Helenius, Photon-photon and photon-hadron processes in Pythia 8. CERN Proc. **1**, 119 (2018)
6. C. Bierlich, G. Gustafson, L. Lönnblad, H. Shah, The Angantyr model for heavy-ion collisions in PYTHIA8. JHEP **10**, 134 (2018)
7. T. Sjöstrand, M. Uthmeim, Hadron interactions for arbitrary energies and species, with applications to cosmic rays. Eur. Phys. J. C **82**(1), 21 (2022)
8. A. Accardi et al., Electron ion collider: the next QCD frontier: understanding the glue that binds us all. Eur. Phys. J. A **52**(9), 268 (2016)
9. R. Abdul Khalek et al. Science requirements and detector concepts for the electron-ion collider: EIC Yellow Report. Nucl. Phys. A **1026**, 122447 (2022)
10. C.A. Bertulani, S.R. Klein, J. Nystrand, Physics of ultra-peripheral nuclear collisions. Ann. Rev. Nucl. Part. Sci. **55**, 271–310 (2005)
11. S. Klein, P. Steinberg, Photonuclear and two-photon interactions at high-energy nuclear colliders. Ann. Rev. Nucl. Part. Sci. **70**, 323–354 (2020)
12. M. Klasen, Theory of hard photoproduction. Rev. Mod. Phys. **74**, 1221–1282 (2002)
13. J.M. Butterworth, M. Wing, High energy photoproduction. Rep. Prog. Phys. **68**, 2773–2828 (2005)
14. E. Abbas et al., Charmonium and  $e^+e^-$  pair photoproduction at mid-rapidity in ultra-peripheral Pb-Pb collisions at  $\sqrt{s_{NN}}=2.76$  TeV. Eur. Phys. J. C **73**(11), 2617 (2013)
15. S. Acharya et al., Coherent  $J/\psi$  and  $\psi'$  photoproduction at midrapidity in ultra-peripheral Pb-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV. Eur. Phys. J. C **81**(8), 712 (2021)
16. V. Khachatryan et al., Coherent  $J/\psi$  photoproduction in ultra-peripheral PbPb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV with the CMS experiment. Phys. Lett. B **772**, 489–511 (2017)
17. R. Aaij et al., Study of coherent  $J/\psi$  production in lead-lead collisions at  $\sqrt{s_{NN}} = 5$  TeV. JHEP **07**, 117 (2022)
18. Photo-nuclear jet production in ultra-peripheral Pb+Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV with the ATLAS detector (2022)
19. G. Aad et al., Two-particle azimuthal correlations in photonuclear ultraperipheral Pb+Pb collisions at 5.02 TeV with ATLAS. Phys. Rev. C **104**(1), 014903 (2021)
20. U. Heinz, R. Snellings, Collective flow and viscosity in relativistic heavy-ion collisions. Ann. Rev. Nucl. Part. Sci. **63**, 123–151 (2013)
21. V. Khachatryan et al., Evidence for collective multiparticle correlations in p-Pb collisions. Phys. Rev. Lett. **115**(1), 012301 (2015)
22. M. Aaboud et al., Measurements of long-range azimuthal anisotropies and associated Fourier coefficients for  $pp$  collisions at  $\sqrt{s} = 5.02$  and 13 TeV and  $p$ +Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV with the ATLAS detector. Phys. Rev. C **96**(2), 024908 (2017)
23. G. Aad et al., Observation of long-range elliptic azimuthal anisotropies in  $\sqrt{s} = 13$  and 2.76 TeV  $pp$  collisions with the ATLAS detector. Phys. Rev. Lett. **116**(17), 172301 (2016)
24. M. Aaboud et al., Measurement of long-range two-particle azimuthal correlations in Z-boson tagged  $pp$  collisions at  $\sqrt{s}=8$  and 13 TeV. Eur. Phys. J. C **80**(1), 64 (2020)
25. M. Gluck, E. Reya, A. Vogt, Photonic parton distributions. Phys. Rev. D **46**, 1973–1979 (1992)
26. E. Bothmann et al., Event Generation with Sherpa 2.2. SciPost Phys. **7**(3), 034 (2019)
27. J. Bellm et al., Herwig 7.0/Herwig++ 3.0 release note. Eur. Phys. J. C **76**(4), 196 (2016)
28. S. Hoeche, F. Krauss, P. Meizinger, Resolved photons in Sherpa. Eur. Phys. J. C **84**(2), 178 (2024)
29. I. Helenius, P. Meizinger, S. Plätzer, P. Richardson, Photoproduction in general-purpose event generators [arXiv:2406.08026](https://arxiv.org/abs/2406.08026) [hep-ph] (2024)
30. I. Abt et al., Azimuthal correlations in photoproduction and deep inelastic  $ep$  scattering at HERA. JHEP **12**, 102 (2021)



31. A. Tumasyan et al., Two-particle azimuthal correlations in  $\gamma p$  interactions using pPb collisions at  $\sqrt{s_{NN}}=8.16$  TeV. *Phys. Lett. B* **844**, 137905 (2023)
32. R.L. Workman et al. Review of particle physics. *PTEP* **2022**, 083C01 (2022)
33. R. Engel, D. Heck, T. Huege, T. Pierog, M. Reininghaus, F. Riehn, R. Ulrich, M. Unger, D. Veberic, Towards a next generation of CORSIKA: a framework for the simulation of particle cascades in astroparticle physics. *Comput. Softw. Big Sci.* **3**(1), 2 (2019)
34. F. Riehn, R. Engel, A. Fedynitch, T.K. Gaisser, T. Stanev, Hadronic interaction model Sibyll 2.3d and extensive air showers. *Phys. Rev. D* **102**(6), 063002 (2020)
35. T. Pierog, I. Karpenko, J.M. Katzy, E. Yatsenko, K. Werner, EPOS LHC: test of collective hadronization with data measured at the CERN Large Hadron Collider. *Phys. Rev. C* **92**(3), 034906 (2015)
36. S. Ostapchenko, Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: I. QGSJET-II model. *Phys. Rev. D* **83**, 014018 (2011)
37. M. Reininghaus, T. Sjöstrand, M. Uthman, Pythia 8 as hadronic interaction model in air shower simulations. *EPJ Web Conf.* **283**, 05010 (2023)
38. C. Bierlich, T. Sjöstrand, M. Uthman, Hadronic rescattering in pA and AA collisions. *Eur. Phys. J. A* **57**(7), 227 (2021)
39. A. Donnachie, P.V. Landshoff, Total cross-sections. *Phys. Lett. B* **296**, 227–232 (1992)
40. H.J. Lipkin, Quarks for pedestrians. *Phys. Rep.* **8**, 173–268 (1973)
41. E.M. Levin, L.L. Frankfurt, The quark hypothesis and relations between cross-sections at high-energies. *JETP Lett.* **2**, 65–70 (1965)
42. G.A. Schuler, T. Sjöstrand, Hadronic diffractive cross-sections and the rise of the total cross-section. *Phys. Rev. D* **49**, 2257–2267 (1994)
43. G.A. Schuler, T. Sjöstrand, A scenario for high-energy gamma gamma interactions. *Z. Phys. C* **73**, 677–688 (1997)
44. M. Glück, E. Reya, A. Vogt, Pionic parton distributions. *Z. Phys. C* **53**, 651–656 (1992)
45. M. Glück, E. Reya, I. Schienbein, Pionic parton distributions revisited. *Eur. Phys. J. C* **10**, 313–317 (1999)
46. R.J. Glauber, Cross-sections in deuterium at high-energies. *Phys. Rev.* **100**, 242–248 (1955)
47. M. Rybczynski, G. Stefanek, W. Broniowski, P. Bozek. GLIS-SANDO 2: GLauber initial-state simulation and more..., ver. 2. *Comput. Phys. Commun.* **185**, 1759–1772 (2014)
48. H. Heiselberg, G. Baym, B. Blaettel, L.L. Frankfurt, M. Strikman, Color transparency, color opacity, and fluctuations in nuclear collisions. *Phys. Rev. Lett.* **67**, 2946–2949 (1991)
49. B. Blaettel, G. Baym, L.L. Frankfurt, H. Heiselberg, M. Strikman, Hadronic cross-section fluctuations. *Phys. Rev. D* **47**, 2761–2772 (1993)
50. M. Alvioli, M. Strikman, Color fluctuation effects in proton-nucleus collisions. *Phys. Lett. B* **722**, 347–354 (2013)
51. M. Alvioli, L. Frankfurt, V. Guzey, M. Strikman, Revealing “flickering” of the interaction strength in pA collisions at the CERN LHC. *Phys. Rev. C* **90**, 034914 (2014)
52. M. Alvioli, B.A. Cole, L. Frankfurt, D.V. Perepelitsa, M. Strikman, Evidence for  $x$ -dependent proton color fluctuations in pA collisions at the CERN Large Hadron Collider. *Phys. Rev. C* **93**(1), 011902 (2016)
53. T. Sjöstrand, P.Z. Skands, Multiple interactions and the structure of beam remnants. *JHEP* **03**, 053 (2004)
54. C. Bierlich, G. Gustafson, L. Lönnblad, A shoving model for collectivity in hadronic collisions [arXiv:1612.05132](https://arxiv.org/abs/1612.05132) [hep-ph] (2016)
55. C. Bierlich, G. Gustafson, L. Lönnblad, Collectivity without plasma in hadronic collisions. *Phys. Lett. B* **779**, 58–63 (2018)
56. C. Bierlich, S. Chakraborty, G. Gustafson, L. Lönnblad, Setting the string shoving picture in a new frame. *JHEP* **03**, 270 (2021)
57. I. Helenius, C.O. Rasmussen, Hard diffraction in photoproduction with Pythia 8. *Eur. Phys. J. C* **79**(5), 413 (2019)
58. G.A. Schuler, T. Sjöstrand, The hadronic properties of the photon in gamma p interactions. *Phys. Lett. B* **300**, 169–174 (1993)
59. T.H. Bauer, R.D. Spital, D.R. Yennie, F.M. Pipkin, The hadronic properties of the photon in high-energy interactions. *Rev. Mod. Phys.* **50**, 261 (1978). [Erratum: *Rev. Mod. Phys.* **51**, 407 (1979)]
60. G.A. Schuler, T. Sjöstrand, Gamma gamma and gamma P events at high-energies, in *Two-Photon Physics from DAPHNE to LEP200 and Beyond* (1994)
61. V.M. Budnev, I.F. Ginzburg, G.V. Meledin, V.G. Serbo, The two photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation. *Phys. Rep.* **15**, 181–281 (1975)
62. C.F. von Weizsacker, Radiation emitted in collisions of very fast electrons. *Z. Phys.* **88**, 612–625 (1934)
63. E.J. Williams, Nature of the high-energy particles of penetrating radiation and status of ionization and radiation formulae. *Phys. Rev.* **45**, 729–730 (1934)
64. P. Skands, S. Carrazza, J. Rojo, Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J. C* **74**(8), 3024 (2014)