Eur. Phys. J. C (2022) 82:21
https://doi.org/10.1140/epjc/s10052-021-09953-5

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

Hadron interactions for arbitrary energies and species, with applications to cosmic rays

Torbjörn Sjöstrand[1], **Marius Utheim**[1,a]

[1] Department of Astronomy and Theoretical Physics, Lund University, Sölvegatan 14A, 223 62 Lund, Sweden

Received: 23 August 2021 / Accepted: 17 December 2021 / Published online: 10 January 2022

Abstract The PYTHIA event generator is used in several contexts to study hadron and lepton interactions, notably pp and p$\bar{\text{p}}$ collisions. In this article we extend the hadronic modelling to encompass the collision of a wide range of hadrons h with either a proton or a neutron, or with a simplified model of nuclear matter. To this end we model hp total and partial cross sections as a function of energy, and introduce new parton distribution functions for a wide range of hadrons, as required for a proper modelling of multiparton interactions. The potential usefulness of the framework is illustrated by a simple study of the evolution of cosmic rays in the atmosphere, and by an even simpler one of shower evolution in a solid detector material. The new code will be made available for future applications.

1 Introduction

Throughout the history of high energy particle physics, one of the most studied processes is proton–proton collisions. Originally the PYTHIA (+ JETSET) event generator [1,2] was designed to simulate e^+e^-/pp/p$\bar{\text{p}}$ collisions. Later it was extended partly to ep and some photon physics, while the coverage of other hadron and lepton collision types has remained limited. For QCD studies, as well as other Standard-Model and Beyond-the-Standard-Model ones, e^+e^-/pp/p$\bar{\text{p}}$/ep has provided the bulk of data, and so there has been little incentive to consider other beam combinations.

In recent years, however, there has been an increasing interest to extend the repertoire of beams. Prompted by the ongoing heavy-ion experiments at RHIC and the LHC, the most significant addition to PYTHIA is the ANGANTYR framework for heavy-ion interactions [3,4], which implements pA and AA collisions building on PYTHIA's existing framework for nucleon–nucleon interactions.

A second new addition is low-energy interactions, which was developed as part of a framework for hadronic rescattering in PYTHIA [5,6]. In this framework, common collisions (ie. mainly those involving nucleons or pions) are modelled in detail, including low-energy versions of standard high-energy processes like diffractive and non-diffractive interactions, as well as low-energy-only non-perturbative processes like resonance formation and baryon number annihilation. Less common collisions (involving eg. excited baryons or charm/ bottom hadrons) use simplified descriptions, the most general being the Additive Quark Model (AQM) [7,8], which gives a cross section that depends only on the quark content of the involved hadrons. This way, the low-energy framework supports interactions for all possible hadron–hadron combinations.

These non-perturbative models are accurate only for low energies, however, up until CM energies around $E_{\text{CM}} = 10$ GeV. This means that, at perturbative energies, still mainly nucleon–nucleon interactions are supported. While other hadron species seldom are used directly as beams in experiments, their collisions still have relevance, in particular for hadronic cascades in a medium. One such example is cosmic rays entering the atmosphere, with collision center-of-mass (CM) energies that stretch to and above LHC energies, and thus give copious particle production. Secondary hadrons can be of rare species, and may interact with the atmosphere at perturbative energies. The objective of this article is to implement general perturbative hadron–nucleon interactions in PYTHIA, using cosmic rays as a test case for the resulting framework.

Two significant extensions are introduced to this end. One is a modelling of total, elastic, diffractive and nondiffractive cross sections for the various beam combinations, as needed to describe collision rates also at energies above $E_{\text{CM}} = 10$ GeV. The other is parton distribution functions (PDFs) for a wide selection of mesons and baryons, as needed to describe the particle production in high-energy collisions. Important

[a] e-mail: marius.m.utheim@jyu.fi (corresponding author)

is also a recent technical improvement, namely the support for selecting beam energies on an event-by-event basis for the main QCD processes, made possible by initializing relevant quantities on an interpolation grid of CM energies. At the time of writing, this is supported for hadron–hadron beams, but not yet for heavy-ion collisions in ANGANTYR, which will prompt us to introduce a simplified handling of nuclear effects in hadron–nucleus collisions. Nucleus–nucleus ones, such as iron hitting the atmosphere, is not yet considered.

Key to the understanding of atmospheric cascades is the model for hadronic interactions. Several different ones are used in the community, such as SIBYLL [9–13], QGSJET [14–18], DPMJET [19–21], VENUS/NEXUS/EPOS [22–26], and HDPM (described in Ref. [27]). It is in this category that PYTHIA could offer an alternative model, constructed completely independently of either of the other ones, and therefore with the possibility to offer interesting cross-checks. In some respects it is likely to be more sophisticated than some of the models above, eg. by being able to handle a large range of beam particles almost from the threshold to the highest possible energies, with semi-perturbative interactions tailored to the incoming hadron type. In other respects it is not yet as developed, like a limited handling of nuclear effects and a lack of tuning to relevant data.

Neither of these programs can describe the important electromagnetic cascades, which instead typically are delegated to EGS [28]. At low energies GHEISHA [29] is often used for nuclear effects, with ISOBAR (described in Ref. [27]), UrQMD [30,31] and FLUKA [32] as alternatives. Generally a typical full simulation requires many components to be combined, under the control of a framework that does the propagation of particles through the atmosphere, taking into account eg. the atmospheric density variation and the bending of charged particles by the earth magnetic field. Two well-known examples of such codes are CORSIKA [27] and AIRES [33,34]. Interestingly for us, the new CORSIKA 8 [35,36] framework is written in C++, like PYTHIA 8 but unlike the other hadronic interaction models, and PYTHIA 8 is already interfaced to handle particle decays, so a further integration is a possibility.

One should also mention that an alternative to Monte Carlo simulation of cascades is to construct a numerical simulation from the cascade evolution equations, one example being MCeq [28]. Other examples include hybrid approaches, where the high-energy part is tracked explicitly but the low-energy part is handled by evolution equations, eg. in SENECA [28] and CONEX [37,38]. Also in these cases the hadronic interaction models provide necessary input. ANGANTYR has in fact already been used to this end, to describe $\mathrm{p}/\pi/\mathrm{K}$ interactions with nuclei [39].

Another application of hadronic cascades is in detector simulations with programs such as FLUKA [32] and GEANT [40–43], which have also been used for cascades in the atmosphere, see eg. [44–46]. GEANT4 depends on external frameworks for simulating collisions, like CORSIKA 8, and has been explicitly designed with an object-oriented architecture that allows users to insert their own physics implementations, one of the current possibilities being PYTHIA 6 [1]. One central difference is that the medium is much denser in a detector, so particles propagate shorter distances before interacting. Hence, some particles that are too short-lived to interact in the atmosphere can do so in detector simulations, eg. D, B, Λ_c^+ and Λ_b^0. For the rest of this article we will focus on the atmospheric case, but we still implement all hadronic interactions relevant for either medium.

To describe hadron–hadron interactions, we need to set up the relevant cross sections and event characteristics. In particular, the latter includes modelling the parton distribution functions for the incoming hadrons. These aspects are developed in Sect. 2. Some simple resulting event properties for hp collisions are shown in Sect. 3. In practice, mediums consist of nuclei such as nitrogen or lead, rather than of free nucleons. Since ANGANTYR does not yet efficiently support nuclear collisions with variable energy, we also introduce and test a simplified handling of nuclear effects in Sect. 3. The main intended application of this framework is to cascades in a medium, so we implement a simple atmospheric model in Sect. 4, and give some examples of resulting distributions. There is also a quick look at passage through a denser medium. Either setup is much simplified relative to CORSIKA or GEANT4, so has no scientific value except to to test and explore features of our new hadronic interactions. The atmospheric toy-model code will be included in a future release of PYTHIA as an example of how to interface a cascade simulation with PYTHIA. Finally we present some conclusions and an outlook in Sect. 5.

2 Cross sections and parton distributions

The first step in modelling the evolution of a cascade in a medium is to have access to the total cross sections for all relevant collisions. Crucially, this relates to how far a particle can travel before it interacts. Once an interaction occurs, the second step is to split the total cross section into partial ones, each with a somewhat different character of the resulting events. Each event class therefore needs to be described separately. At high energies a crucial component in shaping event properties is multiparton interactions (MPIs). PYTHIA models MPIs in terms of parton distribution functions (PDFs), and these have to be made available for all relevant hadrons. Special attention also has to be given to particles produced in the forward direction, that take most of the incoming energy and therefore will produce the most energetic subsequent interactions. These topics will be discussed in the following.

2.1 New total cross sections

The description of cross sections depends on the collision energy. At low energies various kinds of threshold phenomena and resonance contributions play a key role, and these can differ appreciably depending on the incoming hadron species. At high energies a more smooth behaviour is expected, where the dominating mechanism of pomeron exchange should give common traits in all hadronic cross sections [47–50].

In a recent article [5] we implemented low-energy cross sections for most relevant hadron–hadron collisions, both total and partial ones, all the way down to the energy threshold. Input came from a variety of sources. The main ones were mostly based on data or well-established models, while others involved larger measures of uncertainty. Extensions were also introduced to the traditional string fragmentation framework, to better deal with constrained kinematics at low energy.

In cases where no solid input existed, the Additive Quark Model (AQM) [7,8] was applied to rescale other better-known cross sections. In the AQM, the total cross section is assumed to be proportional to the product of the number of valence quarks in the respective hadron, so that eg. a meson–meson cross section is 4/9 that of a baryon–baryon one. The contribution of a heavy quark is scaled down relative to that of a u/d quarks, however, presumably by mass effects giving a narrower spatial wave function. Assuming that quarks contribute inversely proportionally to their constituent masses, we define an effective number of valence quarks in a hadron to be approximately

$$n_{\mathrm{q,AQM}} = n_{\mathrm{u}} + n_{\mathrm{d}} + 0.6\, n_{\mathrm{s}} + 0.2\, n_{\mathrm{c}} + 0.07\, n_{\mathrm{b}}. \tag{1}$$

This expression will also be used as a guide for high-energy cross sections, as we shall see.

The emphasis of the low-energy cross sections lies on the description of collisions below $E_{\mathrm{CM}} = 5$ GeV, say, but the models used should be valid up to 10 GeV. Many processes also have a sensible behaviour above that, others gradually less so.

At the other extreme then lies models intended to describe high-energy cross sections. Here $\mathrm{pp}/\mathrm{p\bar{p}}$ collisions are central, given the access to data over a wide energy range, and the need to interpret this data. A few such models have been implemented in PYTHIA [51], giving the possibility of comparisons. Fewer models are available for diffractive topologies than for the total and elastic cross sections.

For the purposes of this study we will concentrate on the SaS/DL option, not necessarily because it is the best one for $\mathrm{pp}/\mathrm{p\bar{p}}$ but because we have the tools to extend it to the necessary range of collision processes in a reasonably consistent manner. The starting point is the Donnachie–Landshoff modelling of the total cross section [52]. In it, a common ansatz

$$\sigma_{\mathrm{tot}}^{AB} = X^{AB} s^{\epsilon} + Y^{AB} s^{-\eta} \tag{2}$$

is used for the collisions between any pair of hadrons A and B. Here s is the squared CM energy, divided by 1 GeV^2 to make it dimensionless. The terms s^{ϵ} and $s^{-\eta}$ are assumed to arise from pomeron and reggeon exchange, respectively, with tuned universal values $\epsilon = 0.0808$ and $\eta = 0.4525$. The X^{AB} and Y^{AB}, finally, are process-specific. $X^{\bar{A}B} = X^{AB}$ since the pomeron is charge-even, whereas generally $Y^{\bar{A}B} \neq Y^{AB}$, which can be viewed as a consequence of having one charge-even and one charge-odd reggeon. Recent experimental studies [53,54] have shown that the high-energy picture should be complemented by a charge-odd odderon [55] contribution, but as of yet there is no evidence that such effects have a major impact on total cross sections.

In the context of $\gamma\mathrm{p}$ and $\gamma\gamma$ studies, the set of possible beam hadrons was extended by Schuler and Sjöstrand (SaS) to cover vector meson collisions [56,57]. Now we have further extended it to cover a range of additional processes on a p/n target, Table 1. The extensions have been based on simple considerations, notably the AQM, as outlined in the table. They have to be taken as educated guesses, where the seeming accuracy of numbers is not to be taken literally. For simplicity, collisions with protons and with neutrons are assumed to give the same cross sections, which is consistent with data, so only the former are shown. The reggeon term for $\phi^0\mathrm{p}$ is essentially vanishing, consistent with the OZI rule [58–60], and we assume that this suppression of couplings between light u/d quarks and s quarks extends to c and b. Thus, for baryons, the reggeon Y^{AB} values are assumed proportional to the number of light quarks only, while the AQM of Eq. (2) is still used for the pomeron term. Another simplification is that D/B and $\bar{\mathrm{D}}/\bar{\mathrm{B}}$ mesons are assigned the same cross section. Baryons with the same flavour content, or only differing by the relative composition of u and d quarks, are taken to be equivalent, ie. $\Lambda\mathrm{p} = \Sigma^{+}\mathrm{p} = \Sigma^{0}\mathrm{p} = \Sigma^{-}\mathrm{p}$.

The DL parametrizations work well down to 6 GeV, where testable. Thus there is an overlap region where either the low-energy or the high-energy cross sections could make sense to use. Therefore we have chosen to mix the two in this region, to give a smooth transition. More precisely, the transition is linear in the range between

$$\begin{aligned} E_{\mathrm{CM}}^{\mathrm{begin}} =& E_{\mathrm{min}} + \max(0., m_A - m_{\mathrm{p}}) \\ &+ \max(0., m_B - m_{\mathrm{p}}) \end{aligned} \tag{3}$$

and

$$E_{\mathrm{CM}}^{\mathrm{end}} = E_{\mathrm{CM}}^{\mathrm{begin}} + \Delta E, \tag{4}$$

Table 1 Coefficients X^{AB} and Y^{AB}, in units of mb, in Eq. (2) for various beam combinations. First section is from DL [52], second from SaS [56] and the rest are new for this study

AB	X^{AB}	Y^{AB}	$Y^{\bar{A}B}$	comment
pp	21.70	56.08	98.39	
pn	21.70	54.77	92.71	not used, see text
π^+p	13.63	27.56	36.02	
K$^+$p	11.82	8.15	26.36	
π^0p	13.63	31.79	–	$(\pi^+\mathrm{p}+\pi^-\mathrm{p})/2$
ϕ^0p	10.01	−1.51	–	K$^+$p + K$^-$p − π^-p
K^{0}p	11.82	17.26	–	$(\mathrm{K}^+\mathrm{p}+\mathrm{K}^-\mathrm{p})/2$
ηp	12.18	19.68	–	$0.6\,\pi^0\mathrm{p}+0.4\,\phi^0\mathrm{p}$
η'p	11.46	13.62	–	$0.4\,\pi^0\mathrm{p}+0.6\,\phi^0\mathrm{p}$
J/ψp	3.33	−0.50	–	ϕ^0p/3
D$^{0,+}$p	8.48	15.65		$(\pi^0\mathrm{p}+\mathrm{J}/\psi\mathrm{p})/2$
D$_s^+$p	6.67	−1.00		$(\phi^0\mathrm{p}+\mathrm{J}/\psi\mathrm{p})/2$
Υp	1.17	−0.18	–	$0.07\,\phi^0$p/0.6
B$^{0,+}$p	7.40	15.81		$(\pi^0\mathrm{p}+\Upsilon\mathrm{p})/2$
B$_s^0$p	5.59	−0.85		$(\phi^0\mathrm{p}+\Upsilon\mathrm{p})/2$
B$_c^+$p	2.25	−0.34		$(\mathrm{J}/\psi^0\mathrm{p}+\Upsilon\mathrm{p})/2$
Λp	18.81	37.39	65.59	AQM, 2 pp/3
Ξp	15.91	18.69	32.80	AQM, pp/3
Ωp	13.02	0.00	0.00	AQM, 0
Λ_cp	15.91	37.39	65.59	AQM, 2 pp/3
Ξ_cp	13.02	18.69	32.80	AQM, pp/3
Ω_cp	10.13	0.00	0.00	AQM, 0
Λ_bp	14.97	37.39	65.59	AQM, 2 pp/3
Ξ_bp	12.08	18.69	32.80	AQM, pp/3
Ω_bp	9.19	0.00	0.00	AQM, 0

where E_{min} is 6 GeV and ΔE is 8 GeV by default.

2.2 New partial cross sections

The total cross section can be split into different components

$$\begin{aligned}\sigma_{\mathrm{tot}} =& \sigma_{\mathrm{ND}} + \sigma_{\mathrm{el}} + \sigma_{\mathrm{SD}(XB)} + \sigma_{\mathrm{SD}(AX)} + \sigma_{\mathrm{DD}} + \sigma_{\mathrm{CD}} \\ &+ \sigma_{\mathrm{exc}} + \sigma_{\mathrm{ann}} + \sigma_{\mathrm{res}} + \cdots\end{aligned} \tag{5}$$

Here ND is short for nondiffractive, el for elastic, SD(XB) and SD(AX) for single diffraction where either beam is excited, DD for double diffraction, CD for central diffraction, exc for excitation, ann for annihilation and res for resonant. Again slightly different approaches are applied at low and at high energies, where the former often are based on measurements or models for exclusive processes, whereas the latter assume smoother and more inclusive distributions. The last three subprocesses in Eq. (5) are only used at low energies. In the transition region between low and high energies, the two descriptions are mixed the same way as the total cross section.

High-energy elastic cross sections are modelled using the optical theorem. Assuming a simple exponential fall-off $\mathrm{d}\sigma_{\mathrm{el}}/\mathrm{d}t \propto \exp(B_{\mathrm{el}}t)$ and a vanishing real contribution to the forward scattering amplitude ($\rho = 0$)

$$\sigma_{\mathrm{el}} = \frac{\sigma_{\mathrm{tot}}^2}{16\pi\, B_{\mathrm{el}}} \tag{6}$$

(with $c = \hbar = 1$). The slope is given by

$$\begin{aligned}\mathrm{B}_{\mathrm{el}}^{AB} &= 2b_A + 2b_B + 2\alpha' \ln\left(\frac{s}{s_0}\right) \\ &\to 2b_A + 2b_B + 2(2.0\, s^{\epsilon} - 2.1),\end{aligned} \tag{7}$$

where $\alpha' \approx 0.25$ GeV^{-2} is the slope of the pomeron trajectory and $s_0 = 1/\alpha'$. In the final expression the SaS replacement is made to ensure that $\sigma_{\mathrm{el}}/\sigma_{\mathrm{tot}}$ goes to a constant below unity at large energies, while offering a reasonable approximation to the logarithmic expression at low energies. The

hadronic form factors $b_{A,B}$ are taken to be 1.4 for mesons and 2.3 for baryons, except that mesons made only out of c and b quarks are assumed to be more tightly bound and thus have lower values. As a final comment, note that a simple exponential in t is only a reasonable approximation at small $|t|$, but this is where the bulk of the elastic cross section is. For pp and p$\bar{\text{p}}$ more sophisticated larger-$|t|$ descriptions are available [51].

Also diffractive cross sections are calculated using the SaS ansatz [57,61]. The differential formulae are integrated numerically for each relevant collision process and the result suitably parametrized, including a special threshold-region ansatz [5]. Of note is that, if the hadronic form factor from pomeron-driven interactions is written as $\beta_{A\mathbb{P}}(t) = \beta_{A\mathbb{P}}(0)\,\exp(b_A t)$ then, with suitable normalization, $X^{AB} = \beta_{A\mathbb{P}}(0)\,\beta_{B\mathbb{P}}(0)$ in Eq. (2). Thus we can define $\beta_{\text{p}\mathbb{P}}(0) = \sqrt{X^{\text{pp}}}$ and other $\beta_{A\mathbb{P}}(0) = X^{A\text{p}}/\beta_{\text{p}\mathbb{P}}(0)$. These numbers enter in the prefactor of single diffractive cross sections, eg. $\sigma_{AB\to AX} \propto \beta^2_{A\mathbb{P}}(0)\,\beta_{B\mathbb{P}}(0) = X^{AB}\,\beta_{A\mathbb{P}}(0)$. This relation comes about since the A side scatters (semi)elastically while the B side description is an inclusive one, cf. the optical theorem. In double diffraction $AB \to X_1 X_2$ neither side is elastic and the rate is directly proportional to X^{AB}.

In addition to the approximate $\mathrm{d}M_X^2/M_X^2$ mass spectrum of diffractive systems, by default there is also a smooth low-mass enhancement, as a simple smeared representation of exclusive resonance states. In the low-energy description of nucleon–nucleon collisions this is replaced by a set of explicit low-mass resonances (eg. $AB \to AR$) [5]. The low-energy description also includes single-resonance ($AB \to R$) and baryon–antibaryon annihilation contributions that are absent in the high-energy one.

The nondiffractive cross section, which is the largest fraction at high energies, is defined as what remains when the contributions above have been subtracted from the total cross section.

Some examples of total and partial cross sections are shown in Fig. 1.

2.3 Hadronic collisions

At low energies the character of an event is driven entirely by nonperturbative processes. In a nondiffractive topology, this can be represented by the exchange of a single gluon, so soft that the momentum transfer can be neglected. The colour exchange leads to two colour octet hadron remnants, however. Each can be split into a colour triplet and a colour antitriplet part, q–$\bar{\text{q}}$ for a meson and q–qq for a baryon. This leads to two (Lund [62]) strings being pulled out, each between the colour of one hadron and the anticolour of the other. In diffraction either a quark or a gluon is kicked out from the diffracted hadron, giving either a straight string or one with a kink at the gluon. Other processes have their own descriptions [5].

At high energies, on the other hand, perturbative processes play a key role. A suitable framework is that of multiparton interactions, MPIs [63,64]. In it, it is assumed that the composite nature of the hadrons leads to several separate parton–parton interactions, each dressed up with associated parton showers. At first glance the interactions occur independently, but at closer look they are connected by energy–momentum-flavour-colour conservation. Especially the last is nontrivial to model, and requires a special colour reconnection step. There the total string length is reduced relative to a first assignment where the MPIs are largely decoupled from each other.

The probability to offer a perturbative description of a nondiffractive event is assumed to be

$$P_{\text{pert}} = 1 - \exp\left(-\frac{E_{\text{CM}} - E_{\text{min}}}{E_{\text{wid}}}\right), \tag{8}$$

when $E_{\text{CM}} > E_{\text{min}}$, and else vanishing. Here E_{CM} is the collision energy in the rest frame, and

$$\begin{aligned} E_{\text{min}} =& E_{\text{min},0} + 2\,\max(0., m_A - m_p) \\ &+ 2\,\max(0., m_B - m_p), \end{aligned} \tag{9}$$

while $E_{\text{min},0}$ and E_{wid} are two free (within reason) parameters, both 10 GeV by default. The same transition can be used for the handling of diffraction, with E_{CM} replaced by the mass of the diffractive system. Note that it is separate from the transition from low- to high-energy cross section expressions.

In perturbative events the parton–parton collision rate (neglecting quark masses) is given by

$$\begin{aligned} \frac{\mathrm{d}\sigma^{AB}}{\mathrm{d}p_\perp^2} =& \sum_{i,j,k} \iiint f_i^A(x_1, Q^2)\, f_j^B(x_2, Q^2) \\ &\times \frac{\mathrm{d}\hat{\sigma}_{ij}^k}{\mathrm{d}\hat{t}}\, \delta\left(p_\perp^2 - \frac{\hat{t}\hat{u}}{\hat{s}}\right) \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}\hat{t} \end{aligned} \tag{10}$$

differentially in transverse momentum $p_\perp$. Here the PDF $f_i^A(x, Q^2)$ represents the probability to find a parton i in a hadron A with momentum fraction x if the hadron is probed at a scale $Q^2 \approx p_\perp^2$. Different subprocesses are possible, labelled by k, but the dominant one is t-channel gluon exchange. It is convenient to order MPIs in falling order of $p_\perp$, like in a parton shower.

A problem is that the perturbative QCD cross section in Eq. (10) is divergent in the $p_\perp \to 0$ limit. This can be addressed by multiplying it with a factor

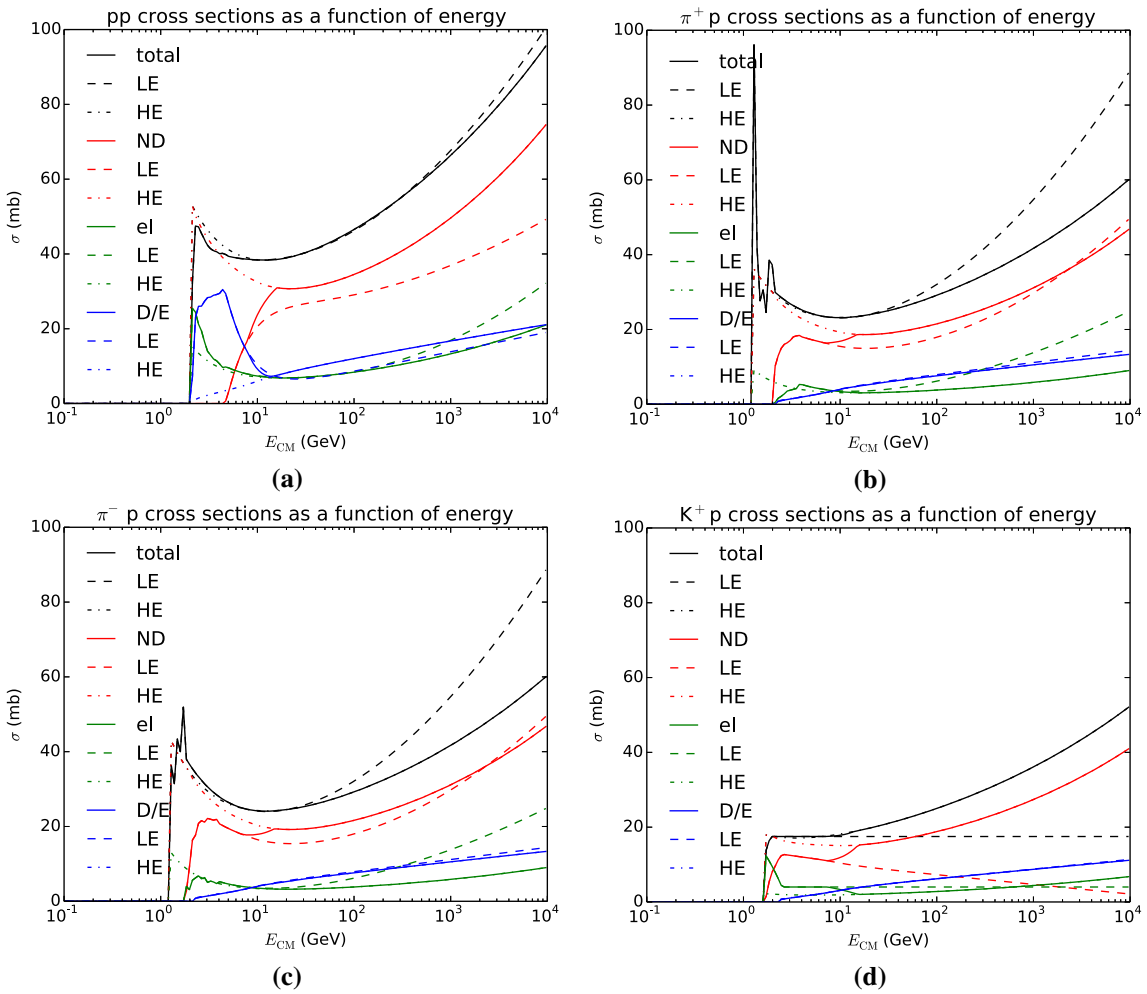


Fig. 1 Total, nondiffractive (ND), elastic (el) and diffractive/excitation (D/E) cross sections for some common collision processes, **a** pp, **b** π^+p, **c** π^-p and **d** K^+p. Full lines show the cross sections actually used, while dashed show the low-energy (LE) and dash-dotted the high-energy (HE) separate inputs. The LE/HE curves are shown also outside of their regions of intended validity, so should be viewed as illustrative only

$$f_{\text{damp}}(p_\perp) = \left(\frac{\alpha_s(p_{\perp 0}^2 + p_\perp^2)}{\alpha_s(p_\perp^2)} \frac{p_\perp^2}{p_{\perp 0}^2 + p_\perp^2} \right)^2 . \tag{11}$$

which is finite in the limit $p_\perp \to 0$. Such a modification can be viewed as a consequence of colour screening: in the $p_\perp \to 0$ limit a hypothetical exchanged gluon would not resolve individual partons but only (attempt to) couple to the vanishing net colour charge of the hadron. The damping could be associated only with the PDFs or only with the $\mathrm{d}\hat{\sigma}/\mathrm{d}\hat{t}$ factor, according to taste, but we remain agnostic on this count. The new $p_{\perp 0}$ parameter is assumed to be varying with the collision energy, with current default

$$p_{\perp 0} = (2.28\ \text{GeV}) \left(\frac{E_{\text{CM}}}{7\ \text{TeV}} \right)^{0.215} , \tag{12}$$

which can be related to the increase of PDFs at low x, leading to an increasing screening with energy.

Most of the MPIs occur in the nondiffractive event class. The average number is given by

$$\langle n_{\text{MPI}} \rangle = \frac{1}{\sigma_{\text{ND}}} \int_0^{E_{\text{CM}}/2} f_{\text{damp}}(p_\perp) \frac{\mathrm{d}\sigma^{AB}}{\mathrm{d}p_\perp} \mathrm{d}p_\perp . \tag{13}$$

MPIs can also occur in high-mass diffraction, and is simulated in PYTHIA [65], but this is a smaller fraction.

The amount of MPIs in a collision directly impacts the event activity, eg. the average charged multiplicity. MPIs have almost exclusively been studied in pp and $\mathrm{p\bar{p}}$ collisions, however, so we have no data to go on when we now want to extend it to all the different collision types listed in Table 1. As a guiding principle we assume that $\langle n_{\text{MPI}} \rangle$ should remain roughly constant, ie. plausibly hadronic collisions at a given

(large) energy have a comparable event activity, irrespective of the hadron types. But we already assumed that total cross sections are lower for mesons than for baryons, and falling for hadrons with an increasing amount of strange, charm or bottom quarks, so naively then Eq. (13) would suggest a correspondingly rising $\langle n_{\mathrm{MPI}} \rangle$. There are (at least) two ways to reconcile this.

One is to increase the $p_{\perp 0}$ scale to make the MPI cross section decrease. It is a not unreasonable point of view that a lower cross section for a hadron is related to a smaller physical size, and that this implies a larger screening. But it is only then interactions at small $p_\perp$ scales that are reduced, while the ones at larger scales remain.

The alternative is to modify the PDFs and to let heavier quarks take a larger fraction of the respective total hadron momentum, such that there are fewer gluons and sea quarks at small x values and therefore a reduced collision rate. (A high-momentum quark will have an enhanced high-$p_\perp$ collision rate, but that is only one parton among many.) This is actually a well-established "folklore", that all long-lived constituents of a hadron must travel at approximately the same velocity for the hadron to stick together. It is a crucial aspect of the "intrinsic charm" hypothesis [66], where a long-lived $\mathrm{c\bar{c}}$ fluctuation in a proton takes a major fraction of the total momentum. In the inverse direction it has also been used to motivate heavy-flavour hadronization [67,68]. This is the approach we will pursue in the following.

2.4 New parton distribution functions

Most PDF studies have concerned and still concern the proton, not least given the massive influx of HERA and LHC data. Several groups regularly produce steadily improved PDF sets [69–71]. The emphasis of these sets are on physics at high Q^2 and (reasonably) high x to NLO or NNLO precision. In our study the emphasis instead is on inclusive events, dominated by MPIs at scales around $p_{\perp 0}$, ie. a few GeV, and stretching down to low x values. These are regions where NLO/NNLO calculations are notoriously unstable, and LO descriptions are better suited.

Moving away from protons, data is considerably more scarce. There is some for the pion, eg. [72–75], a very small amount for the Kaon [76], and nothing beyond that. There has also been some theoretical PDF analyses, based on data and/or model input, like [77–92] for the pion and [79,82,93–97] for the Kaon. But again nothing for hadrons beyond that, to the best of our knowledge, which prompts our own work on the topic.

In order to be internally consistent, we have chosen to take the work of Glück, Reya and coworkers as a starting point. The basic idea of their "dynamically generated" distributions is to start the evolution at a very low Q_0 scale, where originally the input was assumed purely valence-quark-like [98]. Over the years both gluon and u/d sea distributions have been introduced to allow reasonable fits to more precise data [99–102], but still with ansätze for the PDF shapes at Q_0 that involve a more manageable number of free parameters than modern high-precision (N)NLO ones do. Their LO fits also work well with the PYTHIA MPI framework. To be specific, we will use the GRS99 pion [80] as starting point for meson PDFs, and the GJR07 proton one [102] similarly for baryons. Also the GR97 Kaon one [79] will play some role.

In the LO GRS99 π^+ PDF the up/down valence, sea, and gluon distributions are all parameterized on the form

$$f(x) = N x^a (1-x)^b (1 + A\sqrt{x} + Bx) \tag{14}$$

at the starting scale $Q_0^2 = 0.26\ \mathrm{GeV}^2$. The sea is taken symmetric $\mathrm{u}_{\mathrm{sea}} = \bar{\mathrm{u}} = \mathrm{d} = \bar{\mathrm{d}}_{\mathrm{sea}}$, while $\mathrm{s} = \mathrm{c} = \mathrm{b} = 0$ at Q_0. The GRS97 K^+ PDFs are described by assuming the total valence distribution to be the same as for π^+ (as specified in the same article), but the u PDF is made slightly softer by multiplying it by a factor $(1-x)^{0.17}$. That is, the K^+ valence PDFs are given in terms of the π^+ PDFs as

$$\begin{aligned} v_{\mathrm{u}}^{\mathrm{K}} &= N_{\mathrm{K}/\pi} (1-x)^{0.17} v_{\mathrm{u}}^{\pi}, \\ v_{\bar{\mathrm{s}}}^{\mathrm{K}} &= (v_{\bar{\mathrm{d}}}^{\pi} + v_{\mathrm{u}}^{\pi}) - v_{\mathrm{u}}^{\mathrm{K}}. \end{aligned} \tag{15}$$

The coefficient $N_{\mathrm{K}/\pi}$ is a normalization constant determined by the flavour sum relation

$$\int_0^1 dx\ v(x) = 1. \tag{16}$$

Gluon and sea (both u/d and s) distributions are taken to be the same as for the pion.

In our work, we make the ansatz that hadron PDFs can be parameterized on the form given in Eq. (14) at the initial scale Q_0, but with $A = B = 0$ since there are no data or guiding principles to fix them in the generic case. The a and b parameters are allowed to vary with the particular parton and hadron in question, while N is fixed by Eq. (16) for valence quarks. The deviations introduced by the $A = B = 0$ assumption are illustrated in Fig. 2. In Fig. 2a E615 data [75] are compared with the π PDFs as given by GRV92 [77], by GRS99 [80], and by our simplified description (labeled SU21) where a has been adjusted to give the same $\langle x \rangle$ as for GRS99. In Fig. 2b the $\bar{\mathrm{u}}_v^{\mathrm{K}^-}/\bar{\mathrm{u}}_v^{\pi^-}$ ratio is compared between data [76], GRS97 [79] and our simplified model. In both cases the model differences are comparable with the uncertainty in data.

To further illustrate the changes introduced by setting $A = B = 0$, Fig. 3 shows the number and transverse momentum of MPIs for different (a) proton and (b) pion PDFs, with average values as in Table 2. In both cases, our simplified SU21 ansatz leads to a shift that is comparable to the difference between the two standard PDFs. Thus we feel confident

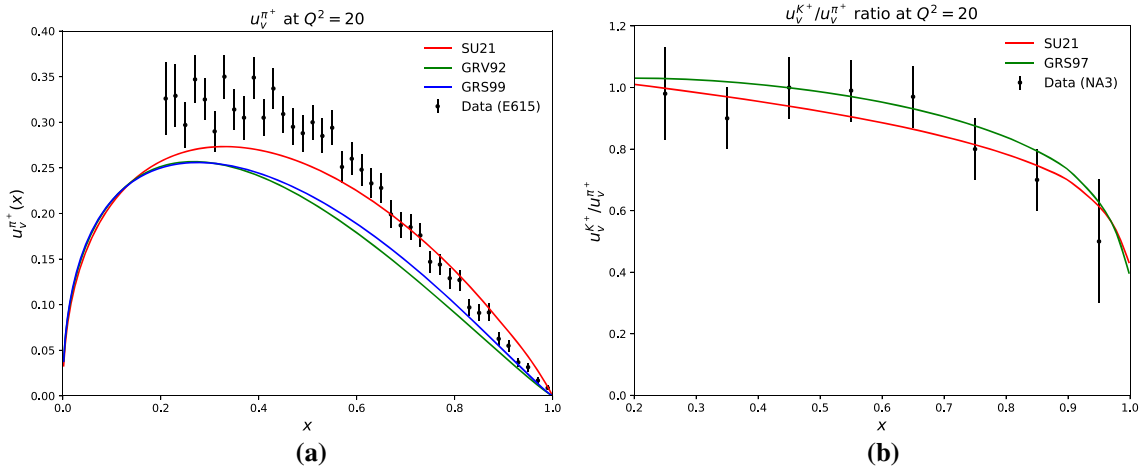


Fig. 2 **a** Pion PDFs, comparing our simplified form to GRV92 and GRS99, and to data. **b** $\bar{u}_v^{K^-}/\bar{u}_v^{\pi^-}$ ratio, comparing our simplified parameterization on the form given in Eq. (14) to the slightly more detailed description of GRS97, and comparing to data. Note that both cases use our simplified $\bar{u}_v^{\pi^-}$ shown in **a**, and only differ in $\bar{u}_v^{K^-}$

that our simplified ansatz is sufficient also for other hadrons, where there are neither data nor detailed theory calculations available. Nevertheless, for accuracy, we use the NNPDF 2.3 QCD+QED LO distribution function for protons and GRS99 LO for pions in our studies, and the SU21 ansatz only for hadrons beyond that.

Given that there is no solid theory for heavy hadron PDFs, the specific choices of a and b necessarily are heuristic. Our guiding principle is that all quarks should have roughly the same velocity, as already mentioned, and thus heavier quarks must have a larger average momentum fraction $\langle x \rangle$, and a smaller b, while gluons and sea u/d must be softer. The $\langle x \rangle$ choices do not exactly agree with the assumed mass ratios in our AQM ansatz, Eq. (1), but are somewhat less uneven than that. This is supported by the Kaon data [76], and also by some modelling [66].

Except for some fine print to come later, our procedure to determine PDFs at the Q_0 starting scale is as follows:

1. Give the valence quark distributions by $Nx^a(1-x)^b$, ie. put $A = B = 0$.
2. Choose sensible b and $\langle x \rangle$ values for each valence quark, based on the principles above.
3. Derive a from $\langle x \rangle = \frac{\int_0^1 dx\, x\, f(x)}{\int_0^1 dx\, f(x)} = (a+1)/(a+b+2)$.
4. Derive N to satisfy Eq. (16).
5. For sea and gluon distributions, pick a d and set $f(x) \propto x^d f^\pi(x)$ (here with A and B values as for the pion).
6. Rescale the gluon and u/d sea distributions by a common factor to satisfy the momentum sum relation

$$\int_0^1 dx \sum_q x f_q(x) = 1.$$

7. The s, c and b contents are zero at the starting scale.

Our choices of b, $\langle x \rangle$ and d are given in Table 3. Excited particles use the same PDFs as their unexcited counterparts. Some PDFs at the initial scale are shown in Fig. 4 for (a) mesons and (b) baryons, which clearly show how heavier quarks are made harder. The baryons are normalized to two u valence quarks, and still the c/b peaks in Σ_c^{++}/Σ_b^+ stand out in the comparison.

Once the initial state has been set up, the DGLAP equations [104–106] describe the evolution towards higher Q^2 scales. Any number of implementations of these equations exist, both private and public, such as QCDNUM [107], HERAFitter/xFitter [108] and APFEL [109], that in principle should be equivalent. We choose to use QCDNUM since we find it well documented and well suited for our purposes. Nevertheless there are some limitations that we had to circumvent.

One such is that the framework is not set up to handle c and b quarks below the respective thresholds Q_c^2 and Q_b^2. To handle their presence, we map some flavours onto others during evolution. Consider eg. a $B^+ = u\bar{b}$ meson, where we wish to evolve the bottom valence $v_{\bar{b}}$ by $\bar{b} \to \bar{b}g$ branchings starting from Q_0^2, but allow $g \to b\bar{b}$ only above Q_b^2. To handle this, we can redefine the initial $\bar{b}$ valence as a contribution eg. to the $\bar{d}$ content, ie. set $\tilde{f}_{\bar{d}}(x, Q_0^2) = f_{sea}(x, Q_0^2) + v_{\bar{b}}(x, Q_0^2)$. Since evolution is linear, this relation also holds for $Q_0^2 \to Q^2 > Q_0^2$, while $f_d(x, Q^2) = f_{sea}(x, Q^2)$. For $Q^2 > Q_b^2$ there will also be a "sea" bottom content $\tilde{f}_{\bar{b}}(x, Q^2)$ from $g \to b\bar{b}$ splittings. Then the correct $\bar{d}$ and $\bar{b}$ contents are reconstructed as

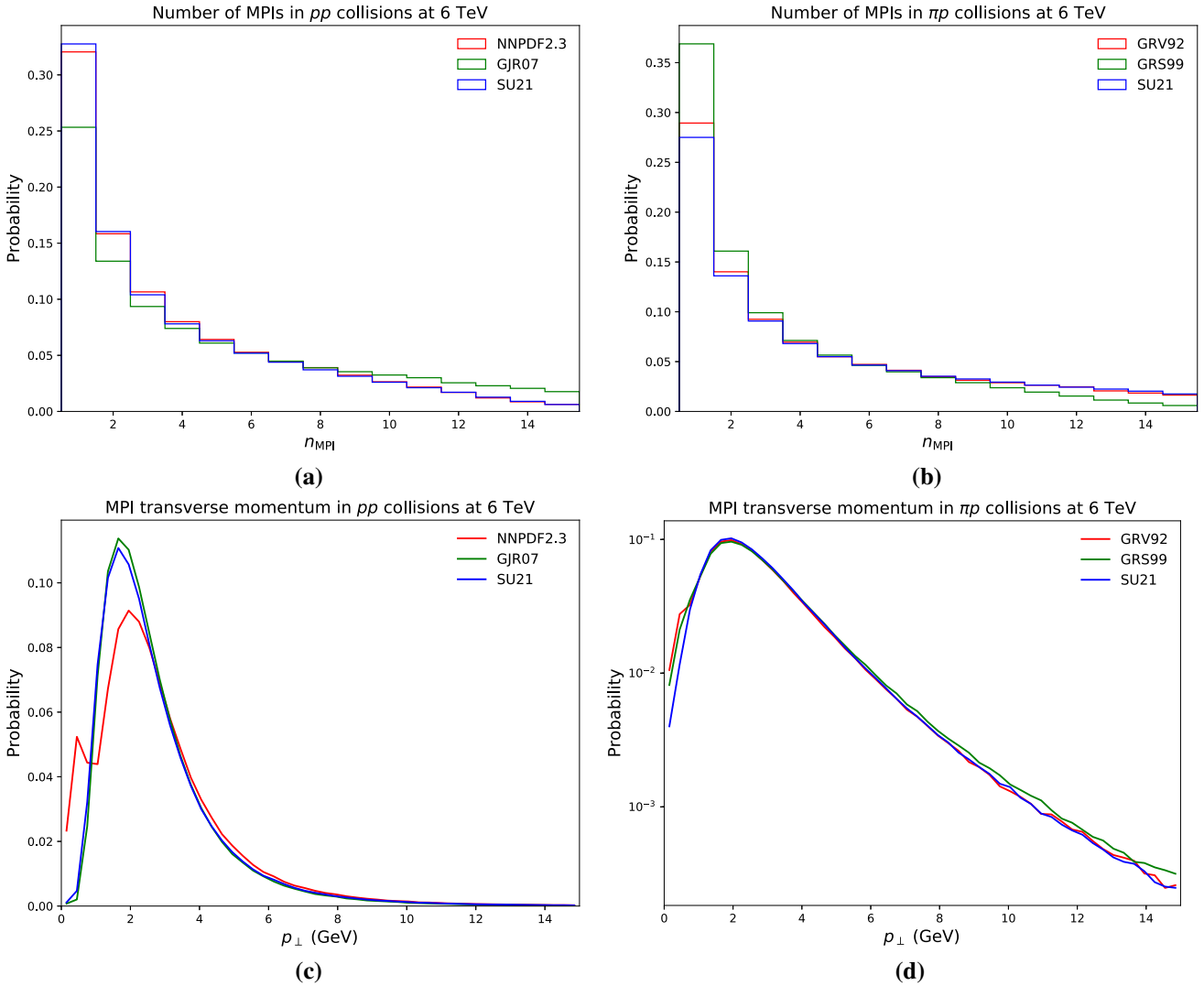


Fig. 3 The **a**, **b** number and **c**, **d** transverse momentum spectrum of MPIs, for **a**, **c** protons and **b**, **d** pions. Each PDF has a cutoff and is considered constant below some Q_0, which leads to the bumps at low $p_\perp$, especially noticeable for NNPDF2.3 distribution in **c**, whose cutoff is $Q = 0.5$ GeV

$$f_{\tilde{\mathrm{b}}}(x, Q^2) = \tilde{f}_{\mathrm{b}}(x, Q^2) + (\tilde{f}_{\bar{\mathrm{d}}}(x, Q^2) - f_{\mathrm{d}}(x, Q^2)),$$
$$f_{\bar{\mathrm{d}}}(x, Q^2) = f_{\mathrm{d}}(x, Q^2). \tag{17}$$

For doubly heavy flavoured mesons, like $\mathrm{B_c}$, we place one valence content in $\bar{\mathrm{d}}$ and the other in u, then use the same procedure. The same trick can be modified to work for flavour-diagonal mesons, like ϕ, J/ψ and Υ, eg. by adding the valence content to d and $\bar{\mathrm{d}}$. Afterward the $\mathrm{u} = \bar{\mathrm{u}} = \mathrm{d} = \bar{\mathrm{d}}$ symmetry of the unmodified sea can be used to shift the heavy flavour content back where it belongs.

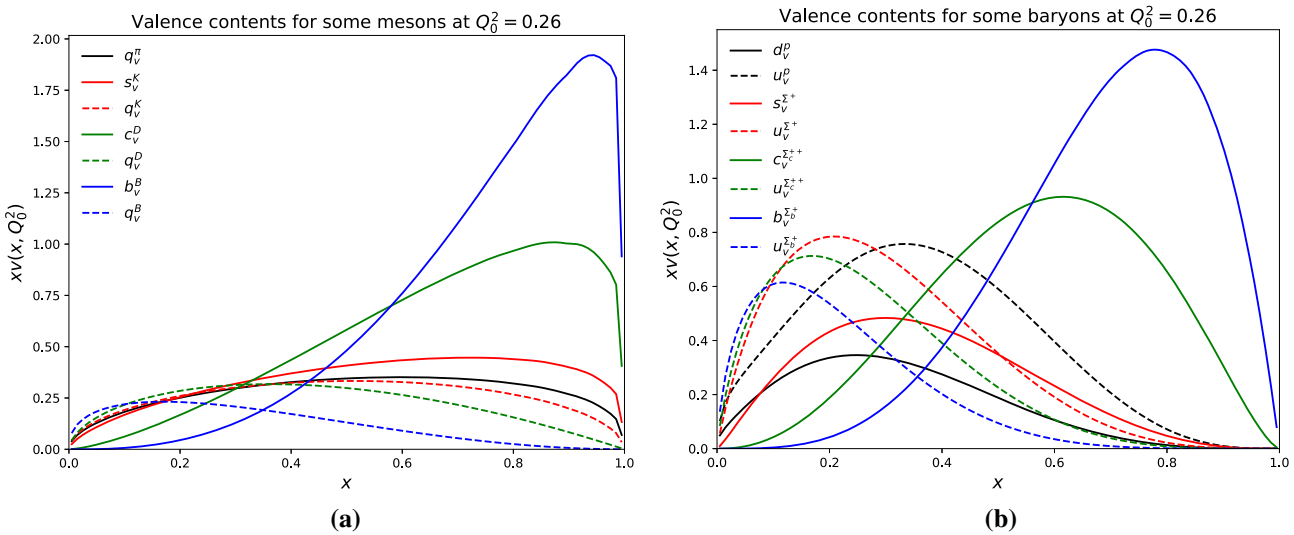
There is a further complication for η and η', which fluctuate between $\mathrm{u}\bar{\mathrm{u}}/\mathrm{d}\bar{\mathrm{d}}/\mathrm{s}\bar{\mathrm{s}}$ valence states. We handle this by treating them as a $\mathrm{u}\bar{\mathrm{u}}$ state during evolution. After a specific quark content is chosen during event generation, the valence part of the evolved $\mathrm{u}\bar{\mathrm{u}}$ is shifted to the corresponding distribution. For simplicity both η and η' are assumed to have the same valence $\langle x \rangle$ values, intermediate between π and ϕ, whether in a $\mathrm{s}\bar{\mathrm{s}}$ state or in a $\mathrm{u}\bar{\mathrm{u}}/\mathrm{d}\bar{\mathrm{d}}$ one.

Table 2 Average number and transverse momentum spectrum for MPIs with different PDFs in pp and πp collisions. The default p PDF in PYTHIA is NNPDF2.3 QCD+QED LO with $\alpha_S(M_Z) = 0.130$ [103]. This default is used for the proton PDF in the πp collisions

	$\langle n_{\mathrm{MPI}} \rangle$	$\langle p_{\perp,\mathrm{MPI}} \rangle$
p, NNPDF2.3	3.27	2.56
p, GJR07	3.88	2.58
p, SU21	3.24	2.54
π, GRV92	3.70	2.67
π, GRS99	3.10	2.68
π, SU21	3.78	2.72

Table 3 Input parameters for the implemented hadron PDFs, as described in the text. Columns are ordered so that heavier quarks appear first. Excited hadrons are also implemented, using the same parameters as for a lighter hadron with the same flavour content

Particle	b_1	$\langle x\rangle_1$	b_2	$\langle x\rangle_2$	b_3	$\langle x\rangle_3$	d
π	0.35	0.28	0.35	0.28	–	–	0.00
K	0.25	0.34	0.52	0.26	–	–	0.17
η	0.32	0.30	0.32	0.30	–	–	0.17
ϕ	0.30	0.32	0.30	0.32	–	–	0.17
D	0.20	0.55	1.00	0.22	–	–	1.00
D_s	0.25	0.53	0.80	0.26	–	–	1.00
J/ψ	0.30	0.43	0.30	0.43	–	–	2.00
B	0.15	0.70	2.00	0.12	–	–	2.00
B_s	0.20	0.68	1.60	0.16	–	–	2.00
B_c	0.25	0.64	1.00	0.24	–	–	3.00
Υ	0.30	0.46	0.30	0.46	–	–	4.00
Σ/Λ	2.8	0.24	3.5	0.17	3.5	0.17	0.17
Ξ	3.0	0.235	3.0	0.235	3.8	0.15	0.17
Ω	3.2	0.22	3.2	0.22	3.2	0.22	0.17
Σ_c/Λ_c	1.5	0.49	4.0	0.14	4.0	0.14	1.00
Ξ_c	1.6	0.475	3.9	0.16	4.5	0.14	1.00
Ω_c	1.7	0.46	3.8	0.16	3.8	0.16	1.00
Σ_b/Λ_b	1.0	0.64	5.0	0.10	5.0	0.10	2.00
Ξ_b	1.1	0.625	4.8	0.12	5.0	0.10	2.00
Ω_b	1.2	0.61	4.8	0.12	4.8	0.12	2.00



(a) **(b)**

Fig. 4 Different valence PDFs at the initial scale $Q_0^2 = 0.26\,\text{GeV}^2$, for (a) π, K, D and B mesons, showing the flavoured valence and the q (= d/u) valence contents; and for (b) uu$q$ baryons for q = d (proton), s (Σ^+), c (Σ_c^{++}) and b (Σ_b^+)

Baryons are treated similarly to mesons, with the obvious exception that they have three valence distributions. In this case the starting point is the GJR07 proton. Despite the known asymmetry of the proton, that the u valence is harder than the d one, we take u and d distributions, where present, to be equal for all other baryons. For heavy-flavoured baryons, the heavy valence is shifted into the d valence in analogy with the meson case. We have not implemented doubly- or triply heavy-flavoured baryons, since these should be produced at a negligible rate, and thus there are no further complications.

After having studied the PDFs resulting from this procedure, we make one additional ad hoc adjustment for the J/ψ, B_c and Υ mesons, ie. the ones that have exclusively c and b valence content. Using only the procedure outlined so far,

the average number of MPIs in interactions involving these particles is much higher than for other hadrons. This comes as no surprise in view of Eq. (13); the absence of light quarks makes for a small total and nondiffractive cross section, while the normal evolution allows a non-negligible gluon and sea to evolve right from the low Q_0 scale. But in real life one should expect heavy quarks to have a reduced emission rate of gluons below their mass scale. To compensate for this, increased Q_0^2 scales of 0.6 GeV^2, 0.75 GeV^2 and 1.75 GeV^2 are used for J/ψ, B_c and Υ, respectively. One could argue that similar shifts should be made for all hadrons containing a c or b quark, but if there are also light valence quarks then there should be some evolution already from small scales. Any mismatch in the emission can then more easily be absorbed in the overall uncertainty of the setup at the Q_0 starting scale.

2.5 The forward region

The fastest particles in the projectile region play a central role for the shower evolution in the atmosphere, so the modelling of this region is a topic of special interest. Traditionally PYTHIA is more aimed towards the modelling of the central region, and there are known issues in the forward region [110–112]. Briefly put, proton/neutron spectra are softer and pion spectra harder than data. In SIBYLL, which normally uses Lund string fragmentation, this has required a separate dedicated handling of leading-baryon formation [13].

We are not here able to report a final resolution of these issues, but a beginning has been made with two new options, both which modify the fragmentation of a diquark in the beam remnant. The first is to disallow popcorn handling [113], ie. the mechanism $q_1q_2 \to q_1\bar{q}_3 + q_2q_3$ whereby a meson can become the leading particle of a jet. (If two valence quarks are kicked out from the proton, which can happen in separate MPIs, the resulting junction topology is unaffected by the popcorn handling.) The second is to set the a and b parameters of the Lund symmetric fragmentation function $f(z) = (1/z)\,(1-z)^a\,\exp(-bm_\perp^2/z)$ separately from those in normal hadronization. There is some support for such a deviation in a few Lund studies [114–116], where it is argued that a drifting-apart of the two quarks of an original dipole indirectly leads to a hardening of the baryon spectrum.

Some first results are shown in Fig. 5. For the nucleon production it can be noted that both steps are about equally important, where $a = 0$ and $b = 2$ GeV^{-2} in the second step. Obviously other a and b values could have given a smaller or larger hardening of the nucleon spectrum. The composition is roughly 65% p and 35% n, over the full phase space. The diffractive peak of protons near $x_F = 1$ has here been removed; in diffraction only the diffracted side of the event is studied. Additional baryon–antibaryon pair production becomes important in the central region, which is why only $x_F > 0.1$ is shown. For pions the major effect comes from removing remnant-diquark popcorn production, while the baryon fragmentation parameters here have a lesser effect.

Further modifications are likely to be necessary, and tuning studies are in progress [117].

2.6 Technical details

The new "SU21" PDFs will be included in an upcoming release of PYTHIA as LHAPDF-compatible [118] files, using the `lhagrid1` format, as a central grid. The grids go down to $x = 10^{-9}$ and up to $Q = 10^4$ GeV.

It is already possible to have a variable energy for the collisions, if switched on at initialization. Then the MPI machinery is initialized at a set of energies up to the maximal one, and later on it is possible to interpolate in tables to obtain relevant MPI values at the current energy. This rather new feature is similar to what has existed a long time for MPIs in diffractive systems, where the diffractive mass varies from event to event even for fixed total energy. Note that it is only implemented for the inclusive processes in the MPI framework, and not for rare processes.

In a future release, it will also be possible to switch between different beam particles on an event-by-event basis. It is assumed that hp and hn cross sections are the same, and that p and n PDFs are related by isospin symmetry. (The latter is not quite true when QED effects are included, but it is close enough for our purposes.) Going one step further, it is also possible to initialize MPIs for the average behaviour of processes that have the same pomeron coefficient X^{AB} in the total cross section, where the PDFs are related by strong isospin, such that the high-energy behaviour should converge. The main example is π^+p, π^-p and π^0p. In detail, the MPI initialization is then based on the average behaviour of σ_{ND} and $d\sigma/dp_\perp$ eg. in Eq. (13), but the total cross section for a collision to occur is still by individual particle combination.

There are other beam combinations that have different total cross sections and PDFs that are not easily related to each other, cf. Tables 1 and 3. In a study where the user wishes to switch between such beam combinations during the run, the PDFs and MPI data grids must be initialized for each individual case. This may take tens of seconds per species, and multiplied by twenty this may be annoyingly long for simple test runs. The future release therefore introduces an option where the MPI initialization data of an instance can be stored on file and reused in a later run. This puts some responsibility on the user, since the new run must then be under the same condition as the original one: same (or only a subset of) allowed processes, same PDFs, same $p_{\perp 0}$, same (or lower) maximum energy, and so on. These features are disabled by default and will only be available if explicitly turned on by the user during initialization.

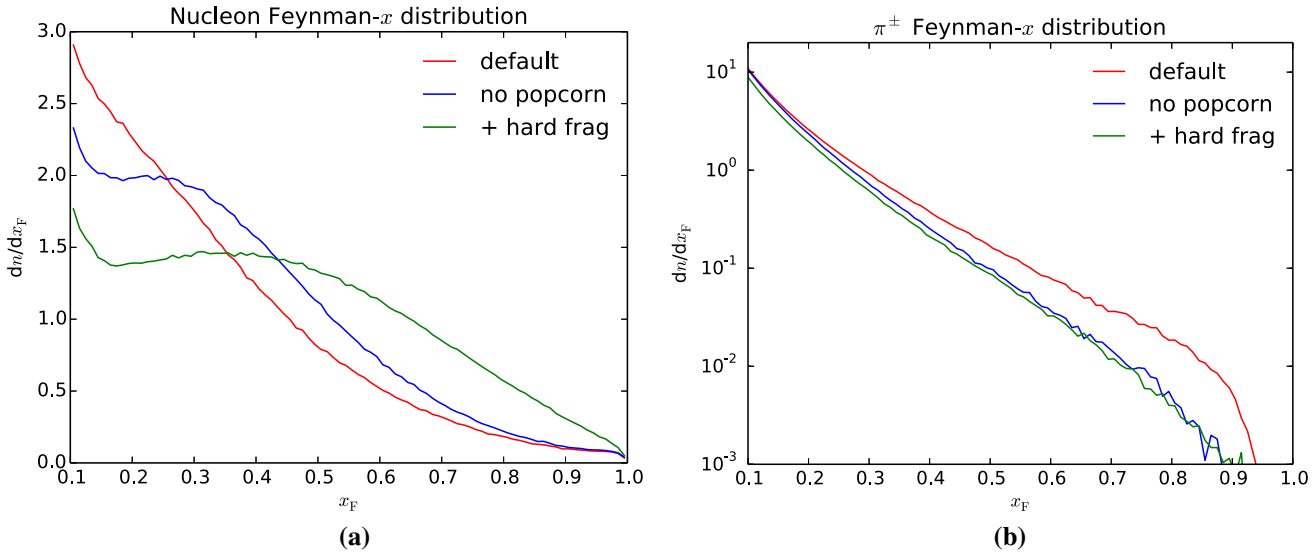


Fig. 5 Feynman-x spectrum of (a) nucleons p/n and (b) $\pi^{\pm}$ for 6 TeV pp collisions with inelastic events, where the quasi-elastic side of single diffraction is not considered

3 Event properties and nuclear effects

With the tools developed in Sect. 2 it is now possible to generate a single hadron-nucleon collision for a wide selection of hadrons and at an almost arbitrary energy. Some comparisons between these hadron-beam options are first presented. But for a realistic simulation of a full cascade, eg. in the atmosphere, we need to consider nuclear effects. Here the ANGANTYR model provides some reference results for fixed topologies. Currently it is not flexible enough for cascade simulation, however, so instead we introduce a simplified approach.

3.1 Hadronic interaction properties

One of the assumptions made above was that the changes in total cross sections and in PDFs would match to some approximation, such that event properties would be comparable over the range of colliding hadrons. In this section we will briefly investigate how this works out by studying non-diffractive hadron–proton collisions at 6 TeV. There is no deep reason for this choice of CM energy, except that any potential proton–oxygen (or proton–nitrogen) run at the LHC is likely to be for a nucleon–nucleon energy in that neighbourhood. The incoming "projectile" hadron will be moving in the $+z$ direction and the "target" proton in the $-z$ ditto.

Distributions for the number and transverse momentum of MPIs are shown in Fig. 6 for a few different hadron types, while Fig. 7 shows charged hadronic multiplicity, $p_\perp$ spectra, and rapidity spectra. We note that, by and large, the various distributions follow suit quite well, and notably the charged multiplicities are comparable. A few key numbers are shown for a larger class of collisions in Table 4, cementing the general picture. This indicates that the joint handling of total cross sections and PDFs are as consistent as can be expected.

Exceptional cases are Υ, B_c^+ and J/ψ where, even after adjusting the initial Q_0^2 scale for the PDF evolution to reduce the number of MPIs, these particles give a higher activity than others. The technical reason is that, even if the MPI cross section is reduced to approximately match the smaller total cross, a non-negligible fraction of the remaining MPIs now come from the heavy valence quarks at large x values. These thereby more easily can produce higher-$p_\perp$ collisions (see $\langle p_{\perp,\mathrm{MPI}} \rangle$ in Table 4), which means more event activity in general.

Studying the rapidity spectra closer, we find asymmetries around zero, depending on how different the PDFs of the projectile are from those of the target proton. That is, a projectile with harder PDFs should give a spectrum shifted towards positive rapidities, and vice versa. Harder valence quarks tend to be counteracted by softer gluons and sea quarks, however. Possible effects also are partly masked by strings being stretched out to the beam remnants, no matter the rapidity of the perturbative subcollision. To better probe larger x values, we also study jet distributions, using the anti-$k_\perp$ algorithm [119,120] with $R = 0.7$ and $p_{\perp\mathrm{jet}} > 50$ GeV. Some distributions are shown in Fig. 8, with average values again given in Table 4. Indeed asymmetries now are quite visible. We also note that the high-$p_\perp$ jet rate, normalized relative to the total nondiffractive cross section, is enhanced in the cases with harder PDFs, even if this is barely noticeable in the average $p_\perp$ of all hadrons.

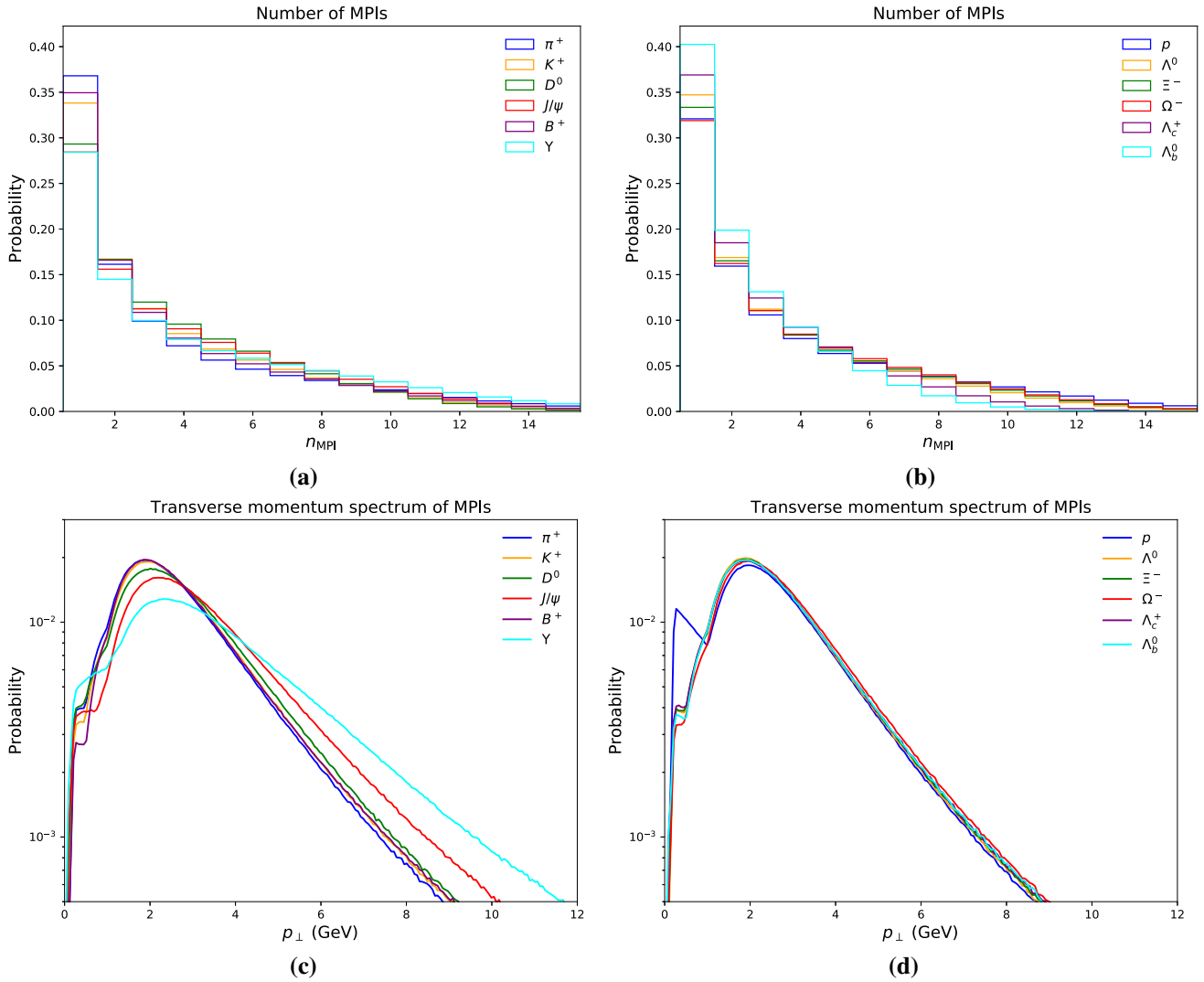


Fig. 6 The **a**, **b** number and **c**, **d** transverse momentum spectrum of MPIs, for a selection of **a**, **c** meson–proton and **b**, **d** baryon–proton collisions at 6 TeV. Labels denote the respective hadron beam. Note that PDFs are taken to be constant below $Q^2 = 0.5$ GeV^2 for protons and below $Q^2 = 0.25$ GeV^2 for the other particles, which is the reason for the strange behaviours at low $p_\perp$

3.2 Nuclear collisions with Angantyr

PYTHIA comes with a built-in model for heavy-ion collisions, ANGANTYR [4], which describes much pA and AA data quite well. It contains a model for the selection of nuclear geometry and impact parameter of collisions. In the Glauber formalism [121] the nucleons are assumed to travel along straight lines, and a binary nucleon–nucleon subcollision can result anytime two such lines pass close to each other. Any nucleon that undergoes at least one collision is called "wounded" [122]. In our case the projectile is a single hadron, so the number of subcollisions equals the number of wounded target nucleons.

In principle all of the components of the total cross section can contribute for each subcollision, but special consideration must be given to diffractive topologies. Notably diffractive excitation on the target side gives rapidity distributions tilted towards that side, a concept used already in the older FRITIOF model [123] that partly has served as an inspiration for ANGANTYR. Alternatively, one can view such topologies as a consequence of the PYTHIA MPI machinery, wherein not all colour strings from several target nucleons are stretched all the way out to the projectile beam remnant, but some tend to get "short-circuited". If such colour connections occur flat in rapidity, then this is equivalent to a $\mathrm{d}M^2/M^2$ diffractive mass spectrum. To first approximation, an ANGANTYR hadron–nucleus collision can be viewed as one "normal" subcollision plus a variable number of diffractive events on the target side.

In its basic form, ANGANTYR event generation is quite fast, about as fast as ordinary PYTHIA hadron–hadron col-

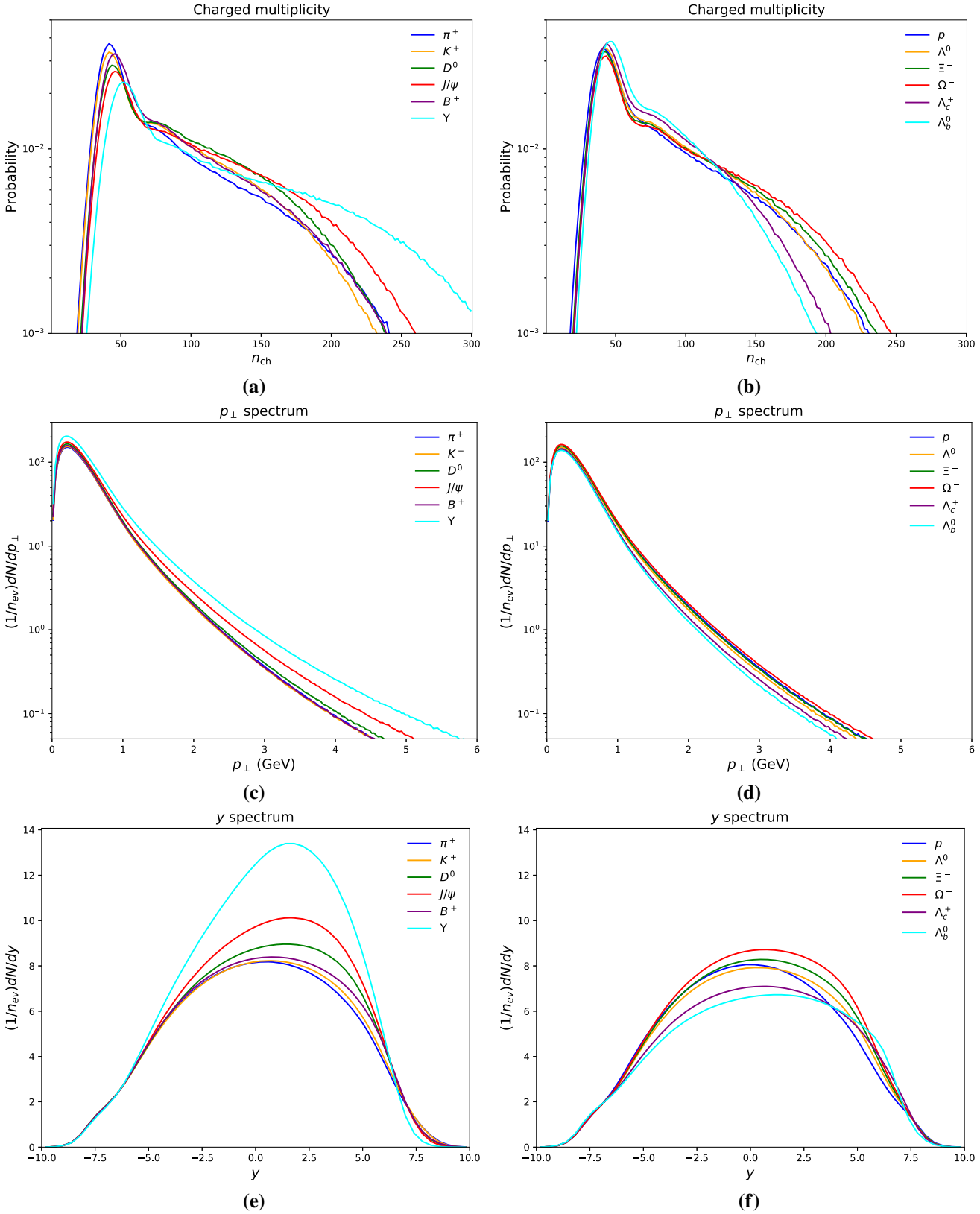


Fig. 7 Charged-hadron **a**, **b** multiplicity distributions, **c**, **d** $p_\perp$ spectra and **e**, **f** rapidity distributions for a selection of **a**, **c**, **e** meson–proton and **b**, **d**, **f** baryon–proton collisions. All results at a 6 TeV collision energy, and for nondiffractive events only. Labels denote the respective hadron beam

Table 4 Some basic numbers for various non-diffractive hadron–proton collisions at 6 TeV: total and nondiffractive cross sections (in mb); average number and $p_\perp$ of MPIs in nondiffractive events; average charged multiplicity, charged $p_\perp$ and y; and jet cross section (in μb) and $p_\perp$ and y of jets with $p_{\perp,\min} = 50$ GeV

beam	σ_{total}	σ_{ND}	$\langle n_{\text{MPI}} \rangle$	$\langle p_{\perp,\text{MPI}} \rangle$	$\langle n_{\text{ch}} \rangle$	$\langle p_\perp \rangle$	$\langle y \rangle$	$\sigma_{\text{jet}}(\mu\text{b})$	$\langle p_{\perp,\text{jet}} \rangle$	$\langle y_{\text{jet}} \rangle$
π^+	55.6	42.9	3.67	2.90	88.1	0.49	0.17	93.6	63.7	0.07
η	49.7	38.5	3.55	3.01	91.7	0.48	0.30	75.9	64.9	0.40
K^+	48.2	37.6	3.57	2.97	89.3	0.48	0.23	77.6	64.8	0.35
ϕ	40.8	32.2	3.75	3.02	93.6	0.48	0.26	75.2	64.6	0.45
D^0	34.6	28.0	3.74	3.07	96.4	0.48	0.36	68.4	64.9	0.53
D_s^+	27.2	22.2	4.04	3.13	99.8	0.49	0.29	72.5	64.6	0.52
J/ψ	13.6	11.3	4.04	3.40	103.9	0.50	0.41	63.1	64.8	0.70
B^+	31.4	26.1	3.60	2.98	92.0	0.48	0.29	65.7	64.9	0.46
B_s^0	22.8	19.1	4.22	3.05	104.4	0.50	0.30	65.7	64.7	0.51
B_c^+	9.2	7.8	4.25	3.50	110.2	0.52	0.43	55.1	64.8	0.74
Υ	4.8	4.1	4.37	3.73	120.1	0.52	0.47	53.7	65.5	0.86
p	88.5	68.3	3.96	2.77	85.9	0.49	0.00	92.9	63.2	0.00
Λ^0	76.7	59.9	3.52	2.88	86.6	0.47	0.15	89.8	64.0	0.14
Ξ^-	64.9	51.3	3.65	2.90	89.8	0.47	0.21	87.1	64.3	0.19
Ω^-	53.1	42.6	3.80	2.98	94.4	0.48	0.26	82.1	64.3	0.17
Λ_c^+	64.9	52.2	3.03	2.90	80.3	0.46	0.27	75.2	64.7	0.32
Ξ_c^0	64.9	52.2	3.06	2.89	81.7	0.46	0.31	79.5	65.0	0.37
Ω_c^0	41.3	33.6	3.96	2.99	97.2	0.48	0.32	76.9	64.8	0.37
Λ_b^0	61.1	50.6	2.69	2.91	78.0	0.45	0.35	71.6	64.9	0.43
Ξ_b^-	61.1	50.6	2.72	2.91	79.5	0.45	0.40	73.9	64.7	0.44
Ω_b^-	37.5	31.7	3.65	2.99	96.0	0.47	0.40	72.0	64.4	0.40

lisions per hadron produced. That is, the overhead from nuclear geometry considerations and energy–momentum sharing between partly overlapping nucleon–nucleon collisions is negligible. The program becomes much slower if the more sophisticated features are switched on, such as ropes [124,125], shove [126,127], and hadronic rescattering [6]. These contribute aspects that only become apparent in a more detailed scrutiny of events, beyond what is needed for our purposes. With minor modifications to the ANGANTYR code itself, ie. on top of the PYTHIA-generic ones we have already introduced in this article, it is also possible to allow any hadron to collide with a nucleus.

There are two severe limitations, however. Firstly, a time-consuming recalculation of hadronic geometry parameters is required anytime the collision energy or incoming hadron species is to be changed in ANGANTYR. Potentially this could be fixed in the future, eg. by interpolation in a grid of initializations at different energies, but it appears to be less simple than what we have introduced for the MPI framework. And secondly, ANGANTYR is only intended to be valid for nucleon–nucleon collision energies above roughly 100 GeV.

A less severe limitation is that the handling of nuclear remnants is very primitive. All non-wounded nucleons are lumped into a new nucleus, without any possibility for it to break up into smaller fragments. For a fixed target the new nucleus is essentially at rest, however, which means that it does not contribute to the continued evolution of the hadronic cascade. Therefore even the simple approach is good enough. In a cascade initiated by a primary nucleus with a fixed energy, it would have been conceivable to handle at least the primary collision using the full ANGANTYR, if it was not for this last limitation.

Even given the limitations, ANGANTYR offers a useful reference when next we come up with a simplified framework. Firstly, the number of subcollisions in pN collision roughly follows a geometric series, Fig. 9a. This is largely a consequence of geometry, where peripheral collisions are common and usually only give one subcollision, while central ones are rare but give more activity. To reach the highest multiplicities one also relies on rare chance alignments of target nucleons along the projectile trajectory. The deviations from an approximate geometric series are larger if one instead considers pPb collisions, Fig. 9b, but not unreasonably so.

An approximate geometric behaviour is also observed for other hN and hPb collisions. The average number of subcollisions depends on the hadron species and the collision energy, but mainly via the total cross section, as can be seen from Fig. 9c,d. Here the results are shown for seven differ-

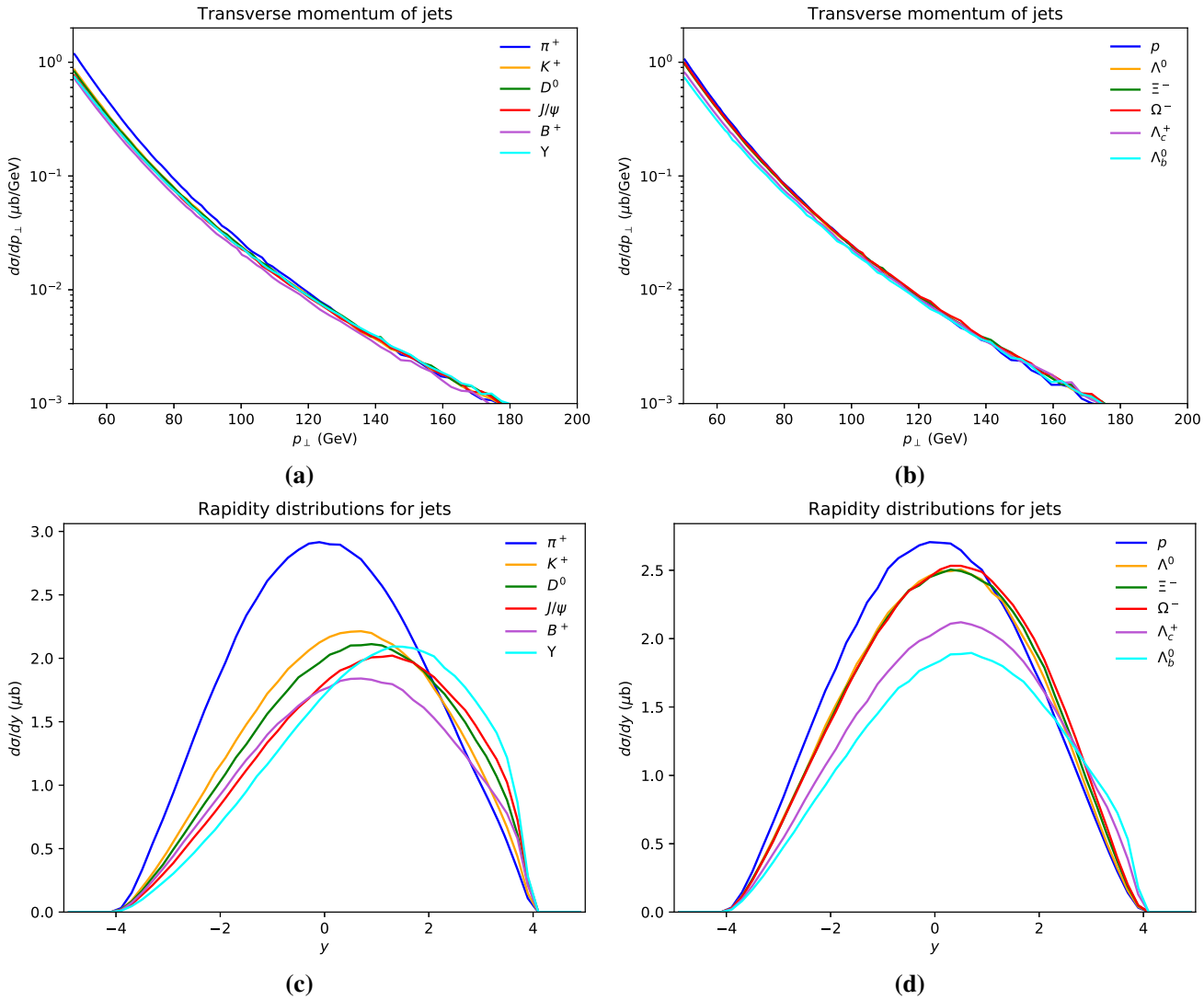


Fig. 8 **a**, **b** Jet $p_\perp$ and **c**, **d** rapidity differential cross sections (in units of μb), with $p_{\perp,\min} = 50$ GeV. Average values are shown in Table 4. Labels denote the respective hadron beam

ent incoming hadrons (p, π^+, K^+, ϕ^0, Λ^0, Ξ^0, Ω^-) at five different energies (10, 100, 1000, 10,000, 100,000 GeV). In the limit of $\sigma_{\text{total}} \to 0$ one would never expect more than one subcollision. Given this constraint, a reasonable overall description is obtained as

$$\langle n_{\text{sub}}^{h\text{N}} \rangle = \begin{cases} 1 + 0.017\sigma_{\text{tot}} & \text{for } \sigma_{\text{tot}} < 31 \\ 1.2 + 0.0105\sigma_{\text{tot}} & \text{else} \end{cases} \tag{18}$$

$$\langle n_{\text{sub}}^{h\text{Pb}} \rangle = \begin{cases} 1 + 0.07\sigma_{\text{tot}} & \text{for } \sigma_{\text{tot}} < 40 \\ 1.8 + 0.05\sigma_{\text{tot}} & \text{else} \end{cases} \tag{19}$$

with σ_{tot} in mb, as plotted in the figures.

Secondly, ANGANTYR may also be used as a reference for expected final-state properties, such as the charged rapidity distribution, $\mathrm{d}n_{\text{ch}}/\mathrm{d}y$. As a starting point, Fig. 10a compares the ANGANTYR pN distribution with the PYTHIA pp one at a 6 TeV collision energy. The PYTHIA curve has beeen scaled up by a factor of 2.05, which corresponds to the average number of subcollisions in pN. While the total charged multiplicities are comparable, there are three differences of note. (1) The ANGANTYR distribution is shifted into the target region, while the PYTHIA one by construction is symmetric around $y = 0$. (2) PYTHIA has a large peak at around $y \approx 8.7$ from elastic scattering, and some single diffraction, that is much smaller in ANGANTYR. (3) At $y \approx -8.7$ instead ANGANTYR has a narrow peak from the not wounded nucleons that together create a new nucleus.

The description of nuclear effects on hadronization is nontrivial. At ANGANTYR initialization the relative composition of different subprocesses is changed. This means eg. that the elastic rate in pN with one single subcollision is reduced relative to PYTHIA pp, Fig. 10b,c. Each further subcollision in

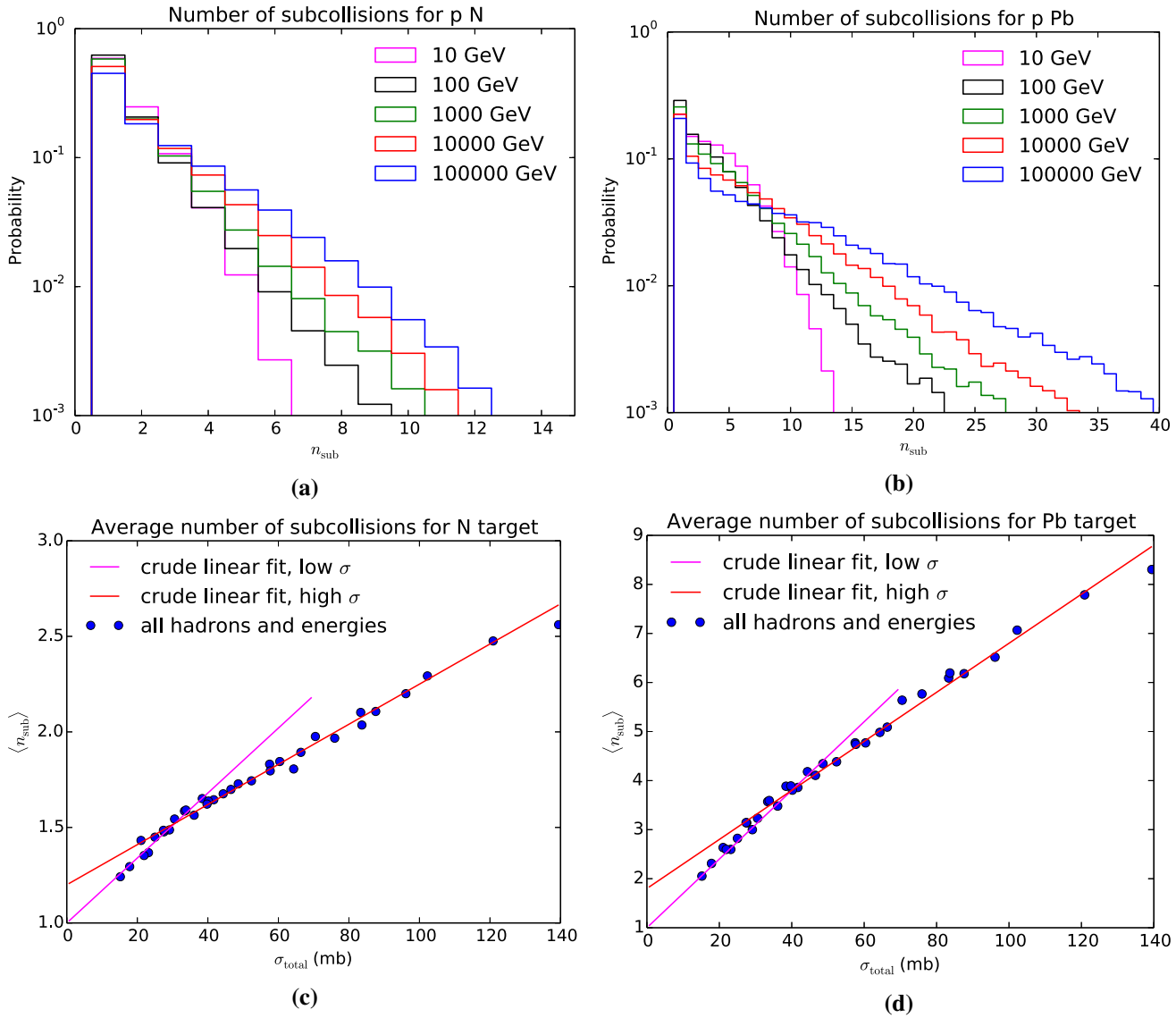
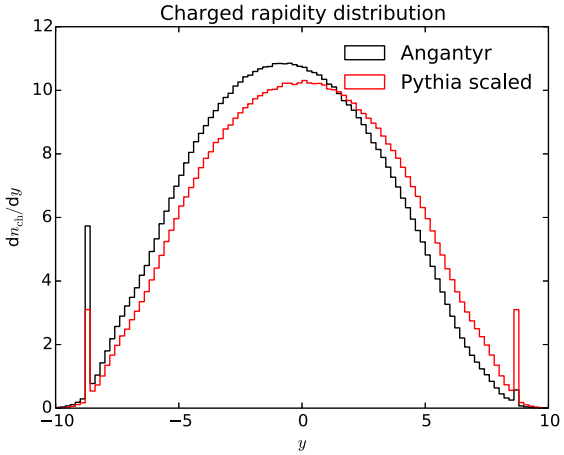


Fig. 9 Number of subcollisions in **a** proton–nitrogen and **b** proton–lead collisions at five different subcollision energies. The average subcollision number in **c** hadron–nitrogen and **d** hadron–lead as a function of the total cross section, with some fits, see text for details

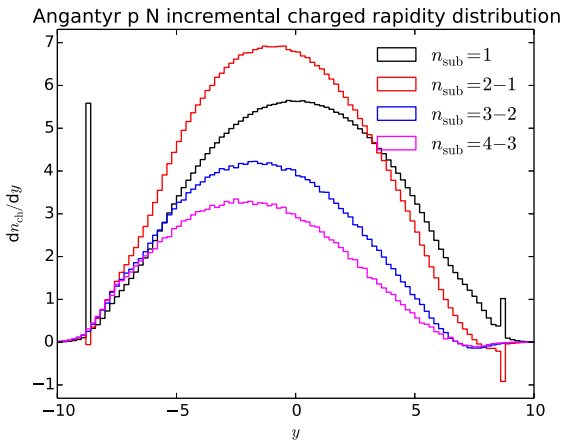
ANGANTYR involves the addition of a diffractive-like system on the target side, but the step from 1 to 2 also includes eg. a drop in the elastic fraction, Fig. 10b, and the correlations arising from nucleons being assumed to have fluctuating sizes event-to-event. The diffractive systems have a higher activity and are more symmetric than corresponding PYTHIA ones, Fig. 10c. As already mentioned the ANGANTYR mechanism is not quite equivalent with that of ordinary diffraction, which is reflected in the choice of PDF for the "pomeron", as used for the MPI activity inside the diffractive system. In ANGANTYR a rescaled proton PDF is used, while by default the H1 2006 Fit B LO [128] pomeron PDF is used in PYTHIA. If MPIs (and parton showers) are switched off, the diffractive systems become quite similar, and have a marked triangular shape, as expected for a $\mathrm{d}M^2/M^2$ diffractive mass distribution. That is, without MPIs the asymmetry of ANGANTYR events is dramatically larger. It is possible to use the rescaled-proton approach also in the PYTHIA simulation to obtain better agreement, but this is not quite good enough, so next we will present a slightly different solution.

3.3 Simplified nuclear collisions

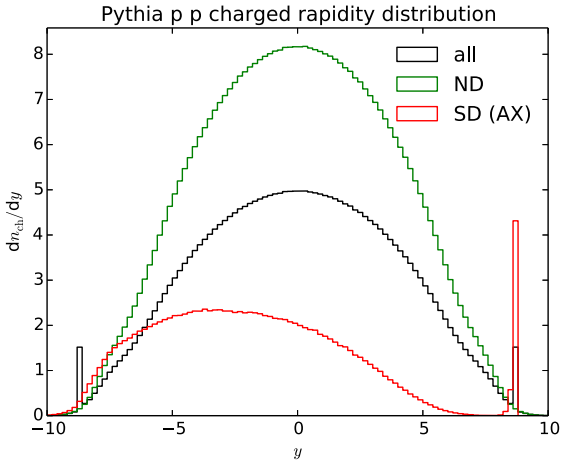
In this section we introduce a simple framework to handle the interactions of a hadron h impinging on an N target at rest. ANGANTYR is used as a reference, but does not allow the rapid switching of beam particle and energy needed to trace the evolution of a cascade, which instead is the area



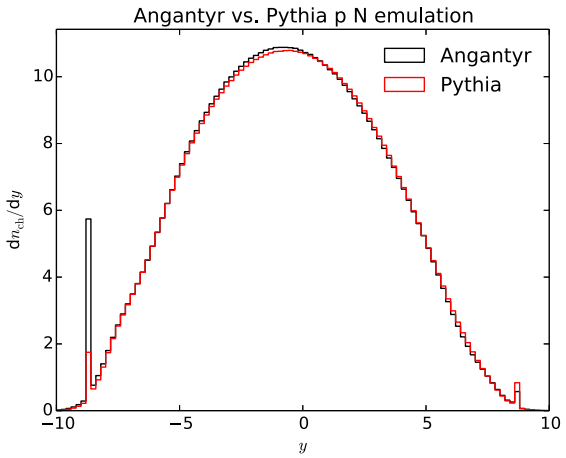
(a)



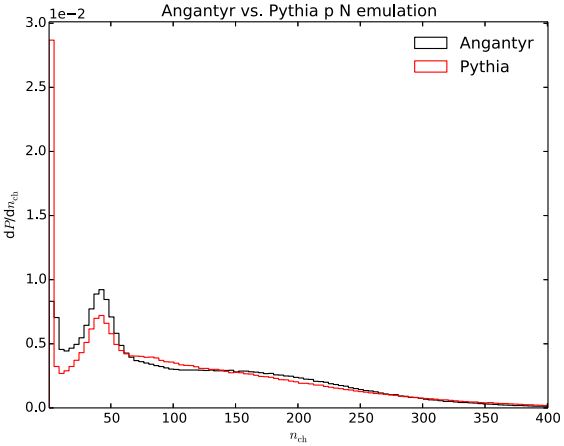
(b)



(c)



(d)



(e)

Fig. 10 Charged rapidity distributions $\mathrm{d}n_{\mathrm{ch}}/\mathrm{d}y$ for a nucleon–nucleon collision energy of 6 TeV. **a** Inclusive ANGANTYR pN events relative to PYTHIA events scaled up with the average number of subcollisions in ANGANTYR. **b** ANGANTYR for different number of subcollisions, relative to results for one less subcollision. **c** PYTHIA inclusive, the nondiffractive (ND) component only, and the target-side single diffractive one (SD (AX)) in pp collisions. **d** The PYTHIA emulation of pN. **e** As **d**, but for the charged multiplicity distribution

where the framework developed in this article excels. The new algorithm works as follows.

1. Calculate the invariant mass for the hp, which is the same as the one for hn to first approximation.
2. Evaluate the hp total cross section $\sigma_{\rm tot}^{hp}$ as already described; again assumed equal to the hn one.
3. Evaluate $\langle n_{\rm sub}^{h\rm N}\rangle$ from Eq. (18).
4. Define $r = 1 - 1/\langle n_{\rm sub}^{h\rm N}\rangle$, such that the geometric distribution $P_n = r^{n-1}/\langle n_{\rm sub}^{h\rm N}\rangle$, $n \geq 1$ has $\sum P_n = 1$ and $\langle n\rangle = \langle n_{\rm sub}^{h\rm N}\rangle$.
5. Decide with equal probability that a proton or a neutron in the target is wounded, and correspondingly generate an inclusive hp or hn event. Do not do any decays, however.
6. Continue the generation with probability r. If not go to point 10. Also go there if there are no more nucleons that can be wounded, or if another user-set upper limit has been reached.
7. Find the newly produced hadron that has the largest longitudinal momentum along the direction of the mother hadron h. Redefine h to be this newfound hadron.
8. Pick a new target proton or neutron among the remaining ones, and generate a corresponding hp or hn event. Below 10 GeV, all low energy processes are allowed. Above 10 GeV, only allow a mix between nondiffractive and target single diffractive topologies, the former with probability $P_{\rm ND} = 0.7$. Again omit decays at this stage.
9. Loop back to point 6.
10. At the end allow all unstable hadrons to decay.

The procedure is to be viewed as a technical trick, not as a physical description. Obviously there is no time for hadronization during the passage of the original h through the N. Rather the idea is that the new h in each step represents the original one, only with some loss of momentum. This is not too dissimilar from how a p projectile in ANGANTYR has to give up some of its momentum for each subcollision it is involved in, even if that particular loss is calculated before hadronization. The new procedure also leads to the central rapidity of each further subcollision being shifted towards the target region.

The specific value of $P_{\rm ND}$ in step 8 has no deep physical meaning, but is "tuned" such that this mix of the relevant curves in Fig. 10c, combined with the rapidity shift procedure just explained, gives the same average behaviour as the consecutive steps in Fig. 10b. Notably a high $P_{\rm ND}$ is needed to reproduce the high activity and low $\langle y\rangle$ shift in ANGANTYR "diffractive" systems.

A simplified test is shown in Fig. 10d, where the new hadron h and the target nucleon is always assumed to be a proton, such that only one collision kind is needed for the simulation at this stage. We have checked that an almost equally good description is obtained also for a lead target, and for a range of collision energies. A warning is in place, however, that the picture is a bit more impressive than warranted. Specifically, the charged multiplicity distributions show non-negligible discrepancies, Fig. 10e, where the PYTHIA machinery gives somewhat more low-multiplicity events, reflected in the forward elastic peak region, which then is compensated elsewhere to give the same average. It should be good enough to get some reasonable understanding of nuclear effects on cascade evolution, however. Later on, for such studies, we will use the full framework.

The Angantyr model is not intended to be applied at very low energies, so there we have no explicit guidance. The same approach with successive subcollisions is still used, with the hardest hadron allowed to go on to the next interaction, meaning that the CM frame is gradually shifted in the target direction. But there are two modifications. Firstly, below 10 GeV in the CM frame, all allowed (low-energy) processes are mixed in their normal fractions. And secondly, no further subcollisions are considered once the kinetic energy of the hardest hadron falls below $E_{\rm kin,min}$.

As a final comment, if each initial hp or hn collision results in an average of $\langle n_{\rm sub}^{h\rm N}\rangle$ subcollisions, then the effective $\sigma_{\rm tot}^{hp}$ cross section must be reduced by the same factor. That is, relative to a gas of free p/n, the incoming hadron will travel longer in a same-nucleon-density N gas before interacting, but produce more subcollisions each time it interacts. More generally, for a nucleus A with atomic number A, the ansatz is that $\sigma_{\rm tot}^{h\rm A}\,\langle n_{\rm sub}^{h\rm A}\rangle = A\,\sigma_{\rm tot}^{hp}$. Nontrivial nuclear effects, such as the fluctuating nuclear sizes or shadowing, could modify this. For the 28 cases studied in Fig. 9 the ratio $\sigma_{\rm tot}^{h\rm A}\,\langle n_{\rm sub}^{h\rm A}\rangle/(A\,\sigma_{\rm tot}^{hp})$ lands in the range $1-1.2$ for N and $1.2-1.4$ for Pb, with no obvious pattern. For now such a possible correction factor is left aside.

4 Modelling hadronic cascades

In the previous sections we have developed and tested the tools needed to described hadron–nucleon interactions, and in an approximate manner hadron–nucleus ones, over a wide range of energies. In this section we will present some simple studies making use of the resulting framework. To this end we introduce a toy model of the atmosphere and study the evolution of a cascade. At the end we also study a cascade in a slab of lead, to go from a dilute to a dense medium, and from a light to a heavy nucleus. These examples are intended to point to possibilities rather than to give any definitive answers.

4.1 Medium density

The simplest possible medium density distribution is a uniform density ρ. While not accurate for the atmosphere, it may be applicable eg. for simulating solid particle detectors. In such a medium, the mean free path of a particle is

$$l_0 = \frac{1}{\sigma\rho}, \tag{20}$$

where σ is the cross section for an interaction between the beam particle and a medium particle. The distance traveled before an interaction then follows an exponential distribution with mean value l_0. In PYTHIA, lengths are given in units of mm and cross sections in units of mb. For particle densities, we use the common standard g/cm^3. The conversion reads

$$l_0 = 1.78266 \cdot 10^4 \text{ mm } \frac{1\text{ mb}}{\sigma} \frac{1\text{ g/cm}^3}{\rho} \frac{mc^2}{1\text{ GeV}}, \tag{21}$$

where m is the target nucleus mass.

If decays of long-lived particles can be neglected, the evolution of the cascade is only a function of the g/cm^2 interaction depth traversed, and we will use this as standard horizontal axis along which to present several results. When decays are to be included, however, it becomes important to model density variations. Let the atmospheric density $\rho(z)$ depend on height z above the surface. The incoming particle enters the medium at $z = z_0$ with a zenith angle θ, and the earth curvature is neglected. Then the naive probability for an interaction at height z is given by

$$\frac{\mathrm{d}\mathcal{P}_{\text{naive}}(z)}{\mathrm{d}z} = \frac{\sigma\rho(z)}{\cos\theta} \equiv f(z). \tag{22}$$

But we are only interested in the first time this particle interacts, and this gives the conventional "radioactive decay" exponential damping, that the particle must not have interacted between z_0 and z:

$$\frac{\mathrm{d}\mathcal{P}(z)}{\mathrm{d}z} = f(z)\ \exp\left(-\int_z^{z_0} f(z')\,\mathrm{d}z'\right). \tag{23}$$

If $f(z)$ has an invertible primitive function $F(z)$ then the Monte Carlo solution for the selection of z is

$$z = F^{-1}(F(z_0) - \log\mathcal{R}), \tag{24}$$

where $\mathcal{R}$ is a random number uniformly distributed between 0 and 1. If not, then the veto algorithm [1] can be used. Once the first interaction has been picked at a height z_1, then the same algorithm can be used for each of the particles produced in it, with z_0 replaced by z_1 as starting point, and each particle having a separate σ and θ. This is iterated as long as needed. For unstable particles also a decay vertex is selected, and the decay wins if it happens before the interaction. A particle reaches the ground if the selected $z < 0$.

In our simple study, we model the atmospheric density at altitude h starting from the ideal gas law, $\rho(h) = pM/RT$, where T and p are temperature and pressure at h, M is the molar mass of dry air, and R is the ideal gas constant. Assuming a linear drop-off for temperature (as is the case for the troposphere [129]), we have $T = T_0 - Lh$, where L is the temperature lapse rate. From the hydrostatic equation, $dp/dh = -g\rho$, the pressure is

$$p = p_0\,(1 - Lh/T_0)^{gM/RL}\,. \tag{25}$$

Then

$$\rho(h) = \frac{p_0 M}{RT_0}\left(1 - \frac{Lh}{T_0}\right)^{\frac{gM}{RL}-1} \approx \rho_0 e^{-h/H}, \tag{26}$$

where the approximation holds for $Lh \ll T_0$, and

$$\frac{1}{H} = \frac{gM}{RT_0} - \frac{L}{T_0}. \tag{27}$$

Using International Standard Atmosphere values for the atmospheric parameters (ISO 2533:1975 [130]) gives $\rho_0 = 1.225$ g/m^3 and $H = 10.4$ km. This approximation is good up until around $L = 18$ km (near the equator), but in our simplified framework, we assume the entire atmosphere follows this shape. In this case, it is possible to sample z according to Eq. (24), specifically,

$$z = -H \log\left(e^{-z_0/H} - \frac{\cos\theta}{H\sigma\rho_0}\log\mathcal{R}\right). \tag{28}$$

In our model, we make the additional assumption that the particle will never interact above $z_0 = 100$ km. This is a good approximation since, applying the exponential approximation in Eq. (26) to infinity, the probability of such an interaction is of order 10^{-5}. Dedicated programs use a more detailed description, eg. CORSIKA has an atmosphere with five different layers.

4.2 Some atmospheric studies

In this section we study the cascade initiated by a proton hitting the model atmosphere above. In the future it would be useful to study also eg. an incoming iron nucleus, but here it is still not clear how to handle the nuclear remnants. The simulation includes hadronic cascades and decays, plus muon decays. But there is no simulation of electromagnetic cascades, nor electromagnetic energy loss of charged particles, nor bending in the earth magnetic field, nor a multitude of other effects. Photoproduction of hadronic states

could be added, either for primary cosmic-ray photons or for secondary ones, but would require some more work, and is not considered important enough for now. The atmosphere is assumed to consist solely of nitrogen, which is a reasonable first approximation, given that eg. oxygen has almost the same atomic number.

Four atmospheric scenarios will be compared:

1. A constant-density atmosphere, like on earth surface, consisting of free protons and neutrons. It starts 10.4 km up, so has the same total interaction depth as the normal atmosphere at vanishing zenith angle.
2. Also a constant-density atmosphere of same height, but using the already described emulation of collisions with nitrogen.
3. An exponentially attenuated "nitrogen" atmosphere, with vanishing zenith angle. Upper cutoff at 100 km.
4. An exponentially attenuated "nitrogen" atmosphere, with 45° zenith angle. This means a factor $\sqrt{2}$ larger interaction depth than the other three scenarios. Again a 100 km upper cutoff.

The first atmosphere is included as a reference, such that the second one can gauge the impact of nuclear effects on the cascade evolution, as follows. In a p/n atmosphere, or between the separate N nuclei in an N one, production is multiplicative: each hadron formed in a first collision can go on and produce new particles in a separate second collision. But inside a nucleus new hadrons do not have time to form until after all the hp/hn subcollisions have occured. Then the hadronic multiplicity is approximately proportional to the number of nucleons wounded by the original incoming hadron, ie. is additive. Therefore one expects a slower start of the cascade evolution in an N atmosphere than in an equivalent-nucleon-density p/n one. In reality differences can be reduced by energy–momentum conservation effects, and by a smaller ability for low-energy particles to produce further hadrons. As a toy example how large effects are possible, we have studied the case where each of the products of a primary 6 TeV pp event, represented by a 19,200 TeV beam on a fixed target, can interact with one further p in the target, Fig. 11. For this toy study the cascade stops after the second step, and hadrons are only allowed to interact if the kinetic energy of the collision is larger than 0.2 GeV.

What is expected to differ between the second and the later atmospheres is the competition between decays and secondary interactions. An early interaction high up in an attenuated atmosphere gives the produced hadrons more time to decay before they can interact.

For simplicity we assume an incoming proton energy of 10^8 GeV. This corresponds to a pp collision CM energy of 13.7 TeV, ie. just at around the maximal LHC energy. Only hadrons with a kinetic energy above $E_{\mathrm{kin,min}} = 0.2$ GeV are

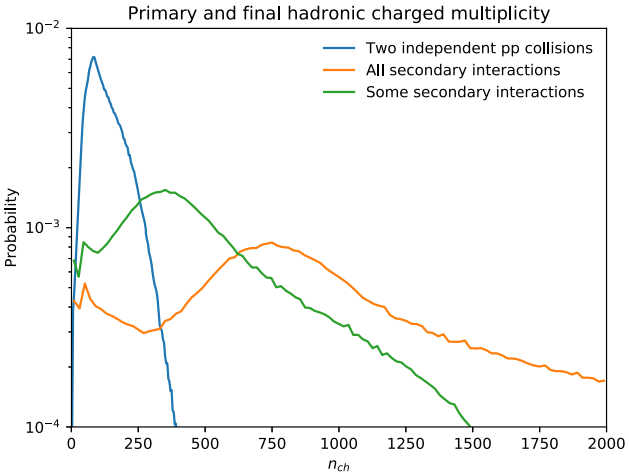


Fig. 11 Toy study on the effect of clustering nucleons into nuclei, as reflected in the charged multiplicity distribution after an initial proton has passed through two "layers" of target nucleons. In one scenario these layers are closely spaced, as inside a (nitrogen) nucleus, so that no hadrons from the first collision have time to form before the second one. This is approximated by the sum of two independent pp collisions. In the other scenario the two "layers" are well separated, as they would be in a hydrogen atmosphere. Then all of the hadrons of the first pp collision have time to form and spread out, and each such hadron can undergo a second independent collision on a target proton. In one suboption ("All") all such hadrons are forced to undergo a second collision. In another, more realistic ("Some"), each interacts with a probability given by the ratio of the individual hadron-proton cross section to the original pp one. See further explanations in the text

allowed to interact. Below that scale they can still decay, but those that do not are assumed to have dissipated and will not be counted in our studies below.

The evolution of the cascade is shown in Fig. 12 and Fig. 13, and the energy spectra and transverse spread of particles reaching the surface in Fig. 14.

Considering Fig. 12a, the hadron–nucleon interaction rate as a function of atmospheric depth, we note that interactions begin earlier in the p/n atmosphere than in the N one, but then also peters out earlier, when most hadrons have low energies. Moving on to the exponential atmosphere, more hadrons (notably pions) decay before they can interact, which reduces the interaction rate. Even more so in the case of a 45° zenith angle, where more of the early evolution takes place in a thin atmosphere. The production rate of hadrons, Fig. 12b, correlates rather well with the interaction rate, although the number of hadrons produced per collision would gradually decrease as each hadron gets to have a lower energy. The early hadron production is higher for the free p/n atmosphere, consistent with expectations but not quite as dramatic as Fig. 11 might have led one to expect. Most of the hadrons decay reasonably rapidly, leaving mainly protons and neutrons to carry on. Fig. 12c shows how the number of such undecayed hadrons increases, following the pattern of the pre-

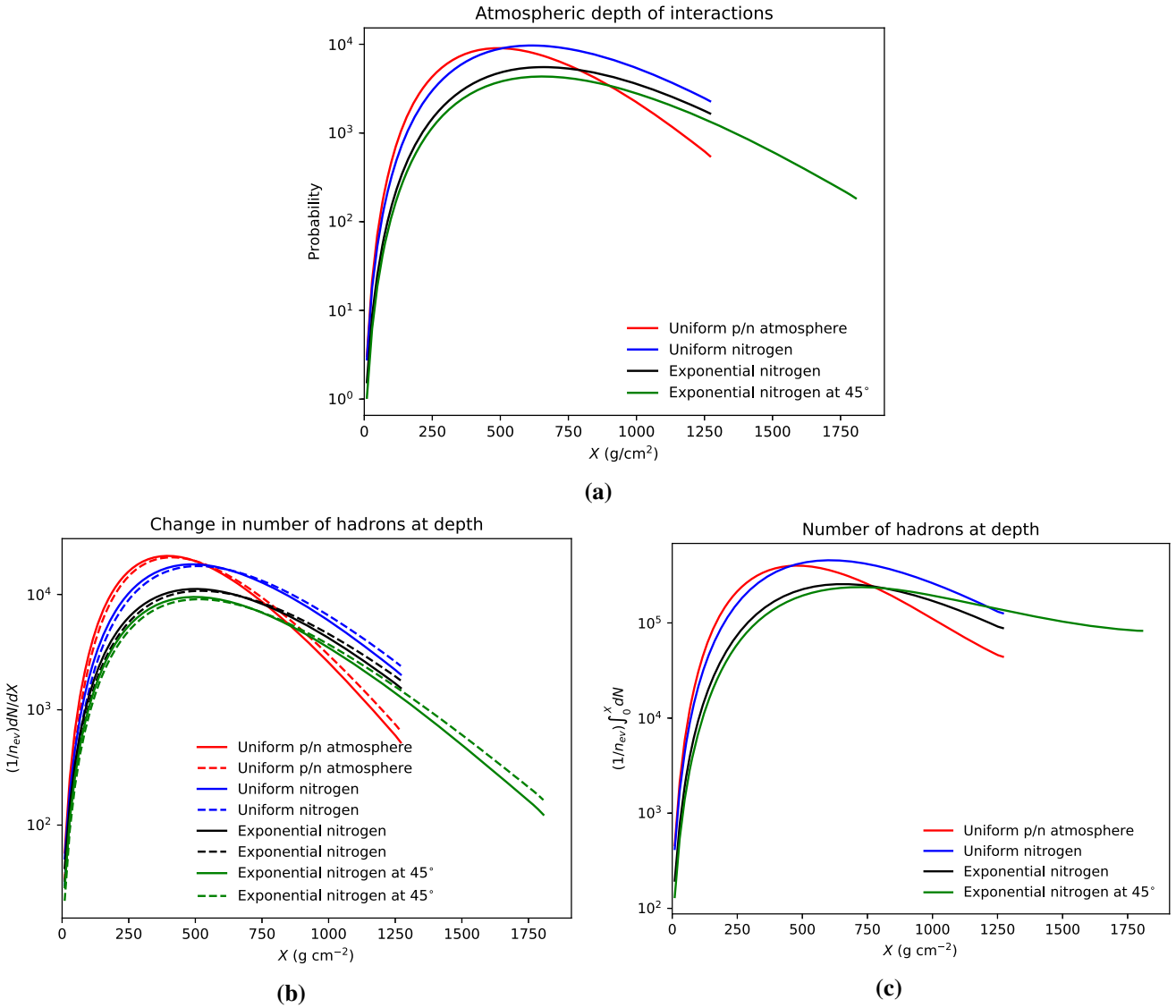


Fig. 12 Evolution of a cascade initiated by a 10^8 GeV proton travelling through four different simple models of the atmosphere, as described in the text, as a function of atmospheric depth in g/cm^2. The models stop when reaching the surface. Shown is the number of **a** interactions, **b** hadrons produced (full) and decayed (dashed), and **c** hadrons remaining. Hadrons that fall below the $E_{\mathrm{kin,min}}$ threshold are removed from the numbers in **c**, but have not been counted as decays in **b**

vious plots. Specifically, the exponential atmospheres give a reduced number of final hadrons.

The long lifetime of muons, $c\tau = 659$ m, means that muon decays lag behind production, Fig. 13a. The number of muons reaches a plateau, where production and decay roughly balance, Fig. 13b. Recall that electromagnetic cascades, largely initiated by the $\pi^0 \to \gamma\gamma$ decays, are not studied here, and thus their potential contribution to muon production is not included. The total number of muons follows the same pattern between the four atmospheric scenarios as noted for hadrons. The production of muon and electron neutrinos, Fig. 13c, d, is dominated by pion and muon decays, but also receives contributions from other weak decays. Neutrino oscillations are not considered here.

The bulk of hadrons that apparently reach the ground have very low kinetic energies, even given the cut $E_{\mathrm{kin}} = E - m > E_{\mathrm{kin,min}}$, Fig. 14a. In reality most of these would be stopped by ionization energy losses or bent away by the earth magnetic field, and those that still reach the ground would do it with reduced energies, so the figure should be viewed as a study of the consequences of hadronic cascades on their own. The uniform and exponential nitrogen atmospheres have comparable rates of higher-energy hadrons. These hadrons are dominated by p and n, which are not

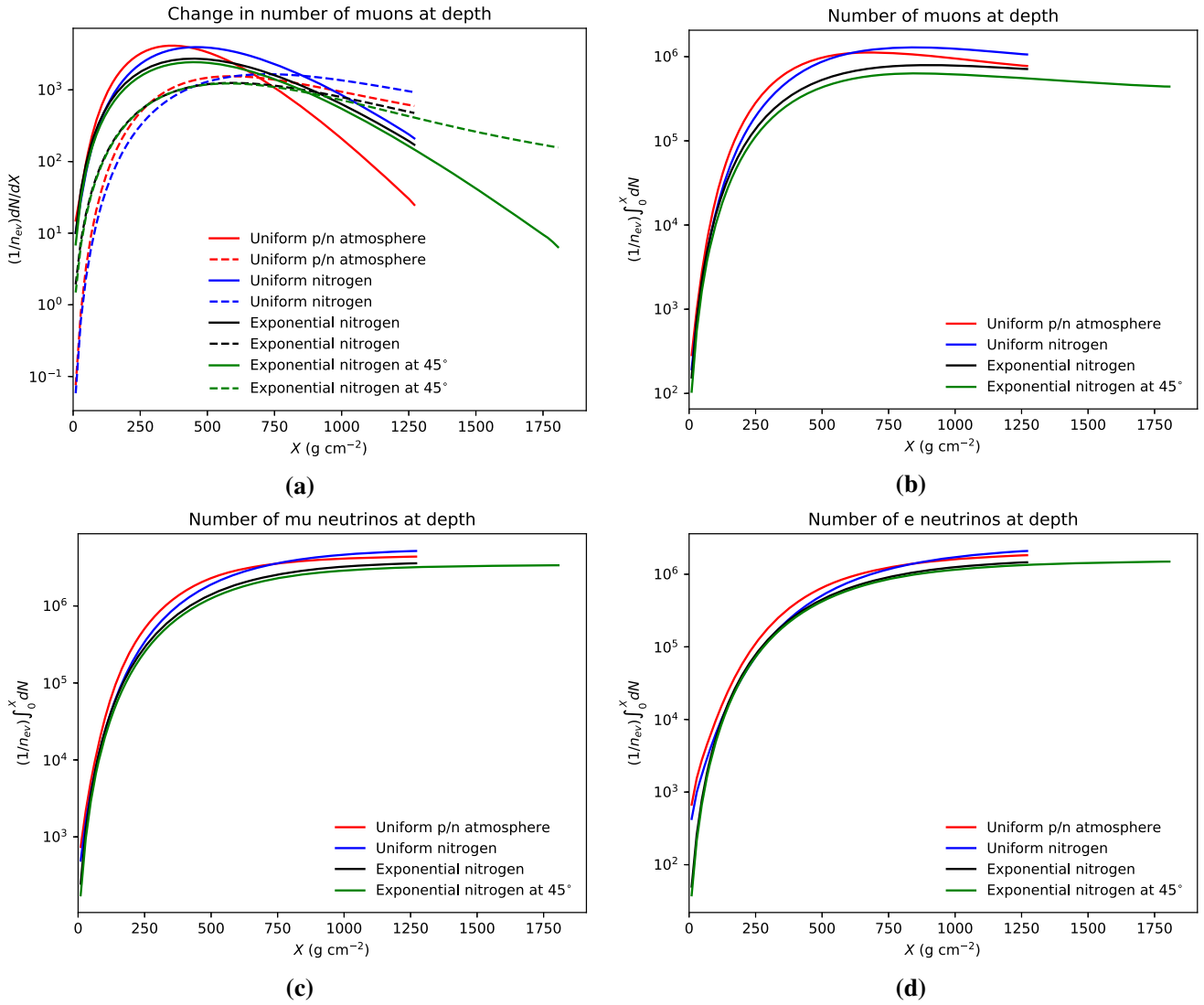


Fig. 13 Evolution of muons and neutrinos for the same cascades as in Fig. 12. Shown is the number of **a** muons produced (full) and decayed (dashed), **b** muons remaining, **c** muon neutrinos remaining and **d** electron neutrinos remaining

affected by decays. That the higher-energy hadron rate is reduced for a non-vanishing zenith angle is to be expected. Also the p/n atmosphere gives a lower rate, presumably as a consequence of the faster split of the original energy into several lower-energy collision chains. The kinetic energy spectra of muons and neutrinos, Fig. 14b, again are peaked at lower energies, though not quite as dramatically. The four atmospheric models also come closer to each other for leptons, though the p/n one remains an outlier.

The cascades disperse particles in quite different directions, implying large footprints on the earth surface. In Fig. 14c, d we show the distributions of hadrons or muons/neutrinos as a function of the distance $r_\perp$ away from the point where the original proton would have hit if it had not interacted. Recall that the relevant area element is $\mathrm{d}^2 r_\perp = 2\pi\, r_\perp\, \mathrm{d}r_\perp$, while Fig. 14c, d plots $\mathrm{d}n/\mathrm{d}r_\perp$, so the number of particles per area is strongly peaked around $r_\perp = 0$. The area argument is also the reason why two of the curves can turn upwards at large $r_\perp$. Not unexpectedly a non-vanishing zenith angle increases the spread, both by having interactions further away and by the elongation of a fictitious shower cone hitting the surface at a tilt. Conversely, the uniform nitrogen gives less spread, by virtue of cascades starting closer to the surface. It should be mentioned that kinetic-energy-weighted distributions (not shown) are appreciably more peaked close to $r_\perp = 0$, as could be expected. Occasionally an event can have a large energy spike close to but a bit displaced from the origin. We have not studied this phenomenon closer, but assume it relates to an early branching where a high-energy particle is produced with a non-negligible transverse kick relative to the event axis.

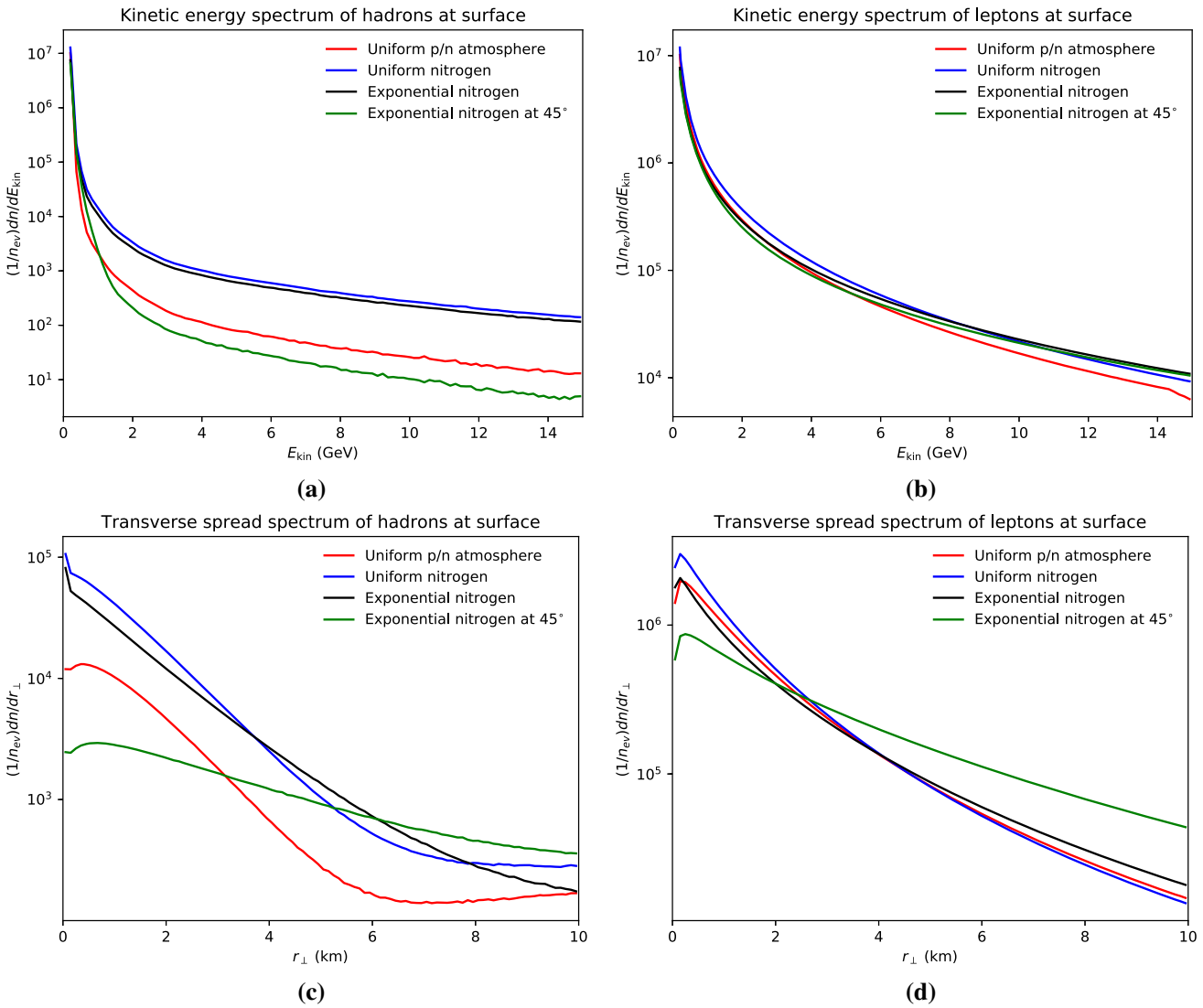


Fig. 14 Kinetic energy spectra (in **a**, **b**) and transverse spread (in **c**, **d**) of the particles that reach the surface for the same cascades as in Fig. 12. Shown are the spectra of **a**, **c** hadrons and **b**, **d** muons and neutrinos

4.3 A lead study

The new code can also be used to track a cascade through a solid material. We have taken lead as an example of a heavy element that is used in some detectors, with rather different properties than the light elements and low density of air. Here the decays of longer-lived particles, such as $\pi^{\pm}$, $\mathrm{K}^{\pm}$, $\mathrm{K_L}$ and $\mu^{\pm}$, do not play as significant a role as in the atmosphere, given the shorter distances a particle travels through a detector. The maximal primary hadronic energy is also lower than for cosmic rays. Taking LHC as example, the 7 TeV maximum translates into collision CM energies below 115 GeV. When we now study the cascades in lead, only hadronic interactions are considered, as before, ie. leptons and photons are free-streaming. Some illustrative results are shown in Fig. 15, for a $p_z = 1$ TeV initial hadron of different kinds. The density of lead is $\rho = 11.35$ g/cm^3, so an interaction depth of 4000 g/cm^2 corresponds to 3.5 m. Hadrons below $E_{\mathrm{kin,min}} = 0.2$ GeV are assumed to stop in the matter and not interact any further. Thus the number of hadrons vanishes after som depth.

The main conclusion of Fig. 15 is that the different incoming hadrons give rise to rather similar cascades. This is largely owing to the rapid multiplication into a fairly similar set of secondary hadrons. Baryons tend to have larger cross sections than mesons, and the proton the largest of them all, so it is understandable why the proton cascade starts somewhat earlier and also dies down earlier. Strange particles have somewhat lower cross sections than their non-strange counterparts, which explains why the K^+ curve starts slower than the π^+ one. But also other factors may be relevant, like how the leading-particle spectrum of a collision affects the nature

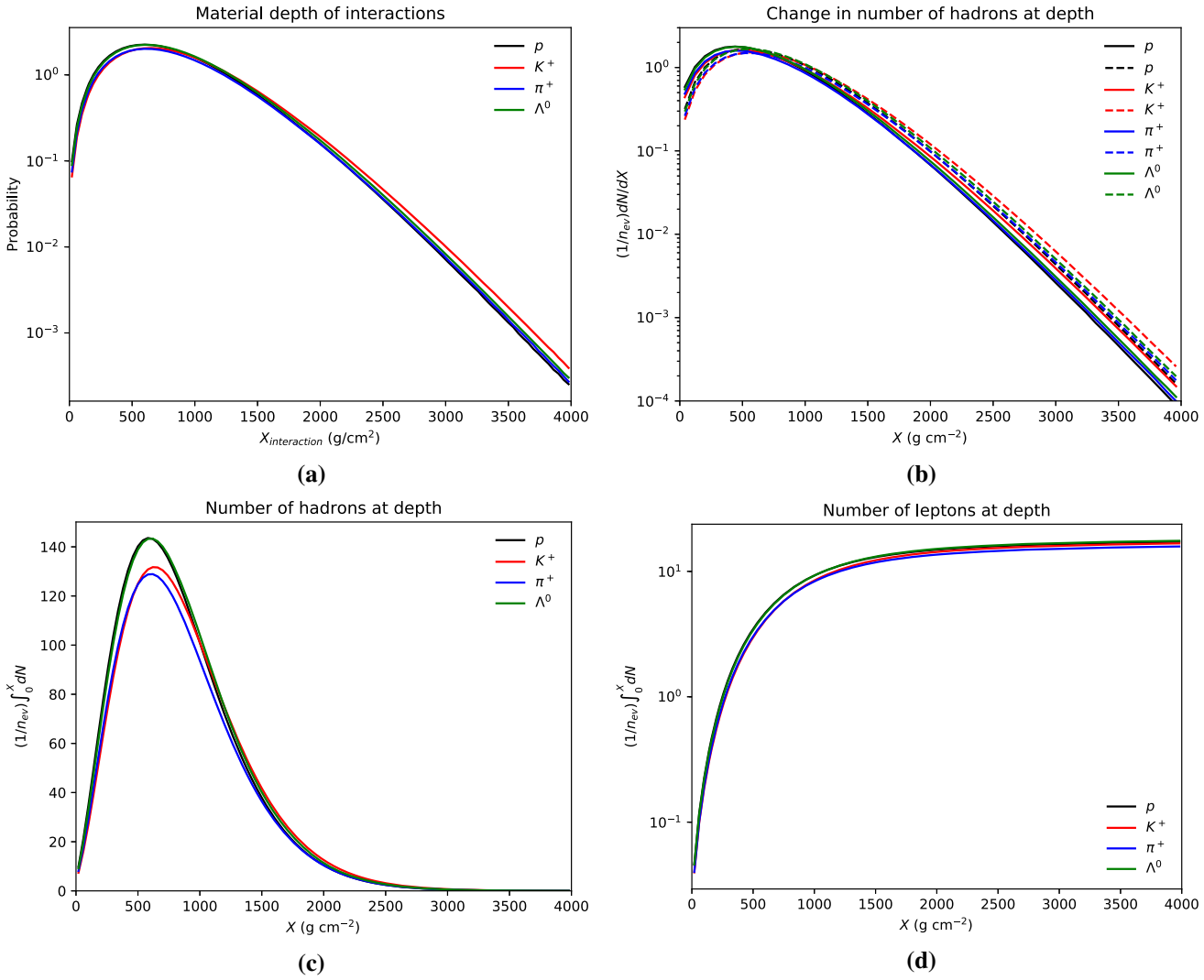


Fig. 15 Evolution of a cascade initiated by a 1000 GeV proton, π^+, K^+ or Λ^0 passing through a 3.5 m thick slab of lead. Shown is the number of **a** interactions, **b** hadrons produced (full) and decayed (dashed), **c** hadrons remaining and **d** muons and neutrinos remaining. Hadrons that fall below the $E_{\mathrm{kin,min}}$ threshold are removed from the numbers in **c**, but have not been counted as decays in **b**

of subsequent collisions. Here we expect a baryon beam to give a harder leading hadron than a meson, and a strange hadron a harder spectrum than a non-strange one, within the context of normal string fragmentation. This could partly compensate for the cross section differences. Further studies will be needed to disentangle these and other factors that may contribute to the small differences observed.

5 Summary and outlook

In this article we have extended the existing hadron–hadron interaction framework of the PYTHIA event generator. Traditionally it has been centered around pp and p$\bar{\mathrm{p}}$ collisions. A few extensions to some meson–meson collision types have been implemented as part of the Vector Meson Dominance scenario of a photon fluctuating to and interacting like a flavour-diagonal vector meson.

Now we have made a deeper study of almost all possible hadron–nucleon collision types. This includes deriving new total and partial cross sections at medium-to-high collision energies, based on the DL and SaS ansätze, extended with the help of the Additive Quark Model and Reggeon systematics where no data is available. It also includes producing some twenty new PDF sets, here denoted SU21. One key assumption has been that heavier valence quarks start out with a larger fraction of the total hadron momentum, at the expense of lighter quarks and gluons, so that all hadron constituents have comparable average velocities. The same constituent-quark-mass ratios as used in the AQM therefore

come to characterize our new PDFs. A consistency check then is that the average number of multiparton interactions is comparable in all collision types. This average is the ratio of the integrated (mini)jet cross section, which directly relates to the PDFs used, and the total (nondiffractive) cross section. Both these numbers should reduce at comparable rates when light quarks are replaced by heavier ones.

Event properties nevertheless are not and should not be identical. This is visible eg. in the rapidity distributions of charged particles, which tend to peak in the hemisphere of the heavier hadron, with its (partly) harder PDFs, and for the same reason such hadronic collisions tend to give somewhat harder $p_\perp$ spectra. Such differences should be explored further and, to the extent data is or becomes available, it would be interesting to compare.

It would also be interesting to explore the sensitivity of the cascade to the different components of the full PYTHIA event simulation. Considerable effort has gone into the separate modelling of different hadron species, but how much of that actually affects the end result? Is it important to use PDFs tailormade for each hadron, or would one proton/baryon and one pion/meson PDF have been enough? And what is the impact of minijets with its initial- and final-state radiation? Jets are key features for LHC physics, where PYTHIA likely is more developed and better tuned than many cosmic-ray generators, but where effects may be overshadowed eg. by the beam-remnant description in the forward direction. (The latter is the subject of a separate ongoing study.) If one wants to study how a charm or bottom hadron interacts on its way through matter, on the other hand, a tailormade description may be relevant.

We do not claim any fundamentally new results in this article, but still present some nice studies that point to the usefulness of the framework. We show how hadronic cascades evolve in the atmosphere, spanning energy scales from 10^8 GeV (or higher if wanted) to 0.2 GeV, how the energy rapidly is spread among many hadrons with low energy each, how hadron decays give muon and neutrino fluxes, how the kinematics and dynamics leads to a wide spread of particles that hit the ground, and more. Note that a complete record of all particles is kept, so it is possible to ask rather specific questions, such as eg. whether hard-jet production in the primary interaction correlate with isolated energy/particle clusters on the ground. We also show, for the solid-target case, how hadrons with larger cross sections also begin their cascades earlier, evolve faster and peter out sooner.

In the current article we have put emphasis on the applications to full cascade evolution, in the atmosphere or in solid matter, rather than on the single collision. One reason is that the full cascade offers further technical challenges on top of modelling the individual collision, which forces us to extend the capabilities of the PYTHIA code. Previously it has not been feasible to switch collision energy or beam type event by event, at least not without each time doing a complete reinitialization, which then slows down event generation times by orders of magnitude. The other reason is that we would like to be able to benefit from and contribute to the understanding of hadronic collisions in different environments. Currently there is one set of event generators that is mainly used for LHC pp physics, such as HERWIG [131], SHERPA [132] or PYTHIA, and another one for cosmic rays, see the Introduction, with only EPOS as an example of a code used in both environments.

Nevertheless, we are aware that we have not presented a full framework for hadronic cascades. One would need to extend the ANGANTYR framework for nuclear collisions so that it could also switch between different collision beams and energies within a manageable time. Ideally it would be validated at lower energies and, for the handling of iron and other heavy cosmic rays, include a model of the nuclear breakup region. This is a tall order, that is beyond our control. In the current study we have instead introduced a quick-and-dirty fix, tuned to reproduce some of the simpler ANGANTYR phenomenology, to handle hadron–nucleus but not nucleus–nucleus collisions.

Furthermore, hadronic cascades is not the end of the story, but must be part of a larger framework that encompasses all relevant processes, and provides a more detailed modelling of the atmosphere. The hope is that the code will find use in larger frameworks, such as CORSIKA 8 for cosmic rays and GEANT4 for detector simulation. The framework ought to be further tested and compared to data before it is used in a state-of-the-art cascade simulation, but even at this point, we at the very least offer a far more powerful replacement to the older PYTHIA 6 code currently used in some such frameworks. In the future we could also take on some other related tasks, such as photoproduction in the cascades.

The PYTHIA generator is under active development in a number of directions. This article should not be viewed as an endpoint but hopefully as a step on the way towards making PYTHIA even more useful for a number of physics studies.

Acknowledgements Thanks to Christian Bierlich for useful discussions on ANGANTYR. Work supported in part by the Swedish Research Council, contract number 2016-05996, and in part by the MCnetITN3 H2020 Marie Curie Innovative Training Network, Grant agreement 722104.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: All new plots are generated with the Pythia event generator. There is no experimental data to be deposited.]

Funded by SCOAP[3].

References

1. T. Sjöstrand, S. Mrenna, P.Z. Skands, PYTHIA 6.4 physics and manual. JHEP **05**, 026 (2006). arXiv:hep-ph/0603175
2. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C.O. Rasmussen, P.Z. Skands, An introduction to PYTHIA 8.2. Comput. Phys. Commun. **191**, 159–177 (2015). arXiv:1410.3012
3. C. Bierlich, G. Gustafson, L. Lönnblad, Diffractive and non-diffractive wounded nucleons and final states in pA collisions. JHEP **10**, 139 (2016). arXiv:1607.04434
4. C. Bierlich, G. Gustafson, L. Lönnblad, H. Shah, The Angantyr model for heavy-ion collisions in PYTHIA8. JHEP **10**, 134 (2018). arXiv:1806.10820
5. T. Sjöstrand, M. Utheim, A framework for hadronic rescattering in pp collisions. Eur. Phys. J. C **80**(10), 907 (2020). arXiv:2005.05658
6. C. Bierlich, T. Sjöstrand, M. Utheim, Hadronic rescattering in pA and AA collisions. arXiv:2103.09665
7. E.M. Levin, L.L. Frankfurt, The Quark hypothesis and relations between cross-sections at high-energies. JETP Lett. **2**, 65–70 (1965)
8. H.J. Lipkin, Quarks for pedestrians. Phys. Rep. **8**, 173–268 (1973)
9. J. Engel, T.K. Gaisser, T. Stanev, P. Lipari, Nucleus–nucleus collisions and interpretation of cosmic ray cascades. Phys. Rev. D **46**, 5013–5025 (1992)
10. R.S. Fletcher, T.K. Gaisser, P. Lipari, T. Stanev, SIBYLL: an event generator for simulation of high-energy cosmic ray cascades. Phys. Rev. D **50**, 5710–5731 (1994)
11. E.-J. Ahn, R. Engel, T.K. Gaisser, P. Lipari, T. Stanev, Cosmic ray interaction event generator SIBYLL 2.1. Phys. Rev. D **80**, 094003 (2009). arXiv:0906.4113
12. R. Engel, F. Riehn, A. Fedynitch, T.K. Gaisser, T. Stanev, The hadronic interaction model Sibyll: past, present and future. EPJ Web Conf. **145**, 08001 (2017)
13. 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). arXiv:1912.03300
14. N.N. Kalmykov, S.S. Ostapchenko, The nucleus–nucleus interaction, nuclear fragmentation, and fluctuations of extensive air showers. Phys. Atom. Nucl. **56**, 346–353 (1993)
15. N.N. Kalmykov, S.S. Ostapchenko, A.I. Pavlov, EAS and a quark–gluon string model with jets. Bull. Russ. Acad. Sci. Phys. **58**, 1966–1969 (1994)
16. S. Ostapchenko, Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: I. QGSJET-II model. Phys. Rev. D **83**, 014018 (2011). arXiv:1010.1869
17. S. Ostapchenko, LHC data on inelastic diffraction and uncertainties in the predictions for longitudinal extensive air shower development. Phys. Rev. D **89**(7), 074009 (2014). arXiv:1402.5084
18. S. Ostapchenko, QGSJET-III model: physics and preliminary results. EPJ Web Conf. **208**, 11001 (2019)
19. J. Ranft, The dual parton model at cosmic ray energies. Phys. Rev. D **51**, 64–84 (1995)
20. S. Roesler, R. Engel, J. Ranft, The Monte Carlo event generator DPMJET-III, in *International Conference on Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications* (MC 2000), 12 (2000). arXiv:hep-ph/0012252
21. A. Fedynitch, R. Engel, Revision of the high energy hadronic interaction models PHOJET/DPMJET-III
22. K. Werner, Strings, pomerons, and the venus model of hadronic interactions at ultrarelativistic energies. Phys. Rep. **232**, 87–299 (1993)
23. H.J. Drescher, M. Hladik, S. Ostapchenko, T. Pierog, K. Werner, Parton based Gribov–Regge theory. Phys. Rep. **350**, 93–289 (2001). arXiv:hep-ph/0007198
24. K. Werner, F.-M. Liu, T. Pierog, Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at RHIC. Phys. Rev. C **74**, 044902 (2006). arXiv:hep-ph/0506232
25. 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). arXiv:1306.0121
26. T. Pierog, B. Guiot, I. Karpenko, G. Sophys, M. Stefaniak, K. Werner, EPOS 3 and air showers. EPJ Web Conf. **210**, 02008 (2019)
27. D. Heck, J. Knapp, J.N. Capdevielle, G. Schatz, T. Thouw, CORSIKA: a Monte Carlo code to simulate extensive air showers
28. W.R. Nelson, H. Hirayama, D.W.O. Rogers, The Egs4 code system
29. H. Fesefeldt, The simulation of hadronic showers: physics and applications
30. S.A. Bass et al., Microscopic models for ultrarelativistic heavy ion collisions. Prog. Part. Nucl. Phys. **41**, 255–369 (1998). arXiv:nucl-th/9803035
31. M. Bleicher et al., Relativistic hadron hadron collisions in the ultrarelativistic quantum molecular dynamics model. J. Phys. G **25**, 1859–1896 (1999). arXiv:hep-ph/9909407
32. A. Ferrari, P.R. Sala, A. Fasso, J. Ranft, FLUKA: a multi-particle transport code (Program version 2005)
33. S.J. Sciutto, AIRES: a system for air shower simulations. arXiv:astro-ph/9911331
34. S.J. Sciutto, The AIRES system for air shower simulations: an update, in *27th International Cosmic Ray Conference*, 6 (2001). arXiv:astro-ph/0106044
35. R. Engel, D. Heck, T. Huege, T. Pierog, M. Reininghaus, F. Riehn, R. Ulrich, M. Unger, D. Veberič, 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). arXiv:1808.08226
36. CORSIKA 8 Collaboration, H.P. Dembinski, L. Nellen, M. Reininghaus, R. Ulrich, Technical Foundations of CORSIKA 8: new concepts for scientific computing. *PoS* **ICRC2019** (2020) 236
37. T. Bergmann, R. Engel, D. Heck, N.N. Kalmykov, S. Ostapchenko, T. Pierog, T. Thouw, K. Werner, One-dimensional hybrid approach to extensive air shower simulation. Astropart. Phys. **26**, 420–432 (2007). arXiv:astro-ph/0606564
38. D. d'Enterria, T. Pierog, G. Sun, Impact of QCD jets and heavy-quark production in cosmic-ray proton atmospheric showers up to 10^{20} eV. Astrophys. J. **874**, 152 (2019). arXiv:1809.06406
39. I.T. Storehaug, Improving the atmospheric neutrino flux estimation in IceCube. Master thesis, University of Copenhagen (2019)
40. R. Brun, F. Bruyant, F. Carminati, S. Giani, M. Maire, A. McPherson, G. Patrick, L. Urban, GEANT detector description and simulation tool
41. GEANT4 Collaboration, S. Agostinelli et al., GEANT4—a simulation toolkit. Nucl. Instrum. Methods A **506**, 250–303 (2003)
42. J. Allison et al., Geant4 developments and applications. IEEE Trans. Nucl. Sci. **53**, 270 (2006)

43. J. Allison et al., Recent developments in Geant4. Nucl. Instrum. Methods A **835**, 186–225 (2016)
44. G. Battistoni, A. Margiotta, S. Muraro, M. Sioli, FLUKA as a new high energy cosmic ray generator. Nucl. Instrum. Methods A **626–627**, S191–S192 (2011). arXiv:1002.4655
45. P. Paschalis, H. Mavromichalaki, L.I. Dorman, C. Plainaki, D. Tsirigkas, Geant4 software application for the simulation of cosmic ray showers in the Earth's atmosphere. New Astron. **33**, 26–37 (2014)
46. R. Sarkar, A. Roy, S.K. Chakrabarti, Simulation of cosmic rays in the Earth's atmosphere and interpretation of observed counts in an X-ray detector at balloon altitude near tropical region. Adv. Space Res. **65**, 189–197 (2020)
47. P.D.B. Collins, *An Introduction to Regge Theory and High-Energy Physics. Cambridge Monographs on Mathematical Physics* (Cambridge University Press, Cambridge, 2009), p. 5
48. J.R. Forshaw, D.A. Ross, *Quantum Chromodynamics and the Pomeron*, vol. 9 (Cambridge University Press, Cambridge, 2011), p. 1
49. S. Donnachie, H.G. Dosch, O. Nachtmann, P. Landshoff, *Pomeron Physics and QCD*, vol. 19 (Cambridge University Press, Cambridge, 2004), p. 12
50. V. Barone, E. Predazzi, *High-Energy Particle Diffraction, Texts and Monographs in Physics*, vol. 565 (Springer, Berlin, 2002)
51. C.O. Rasmussen, T. Sjöstrand, Models for total, elastic and diffractive cross sections. Eur. Phys. J. C **78**(6), 461 (2018). arXiv:1804.10373
52. A. Donnachie, P.V. Landshoff, Total cross-sections. Phys. Lett. B **296**, 227–232 (1992). arXiv:hep-ph/9209205
53. TOTEM Collaboration, G. Antchev et al., First determination of the ρ parameter at $\sqrt{s} = 13$ TeV: probing the existence of a colourless C-odd three-gluon compound state. Eur. Phys. J. C **79**(9), 785 (2019). arXiv:1812.04732
54. D0, TOTEM Collaboration, V.M. Abazov et al., Comparison of pp and $p\bar{p}$ differential elastic cross sections and observation of the exchange of a colorless C-odd gluonic compound. arXiv:2012.03981
55. L. Lukaszuk, B. Nicolescu, A possible interpretation of p p rising total cross-sections. Lett. Nuovo Cim. **8**, 405–413 (1973)
56. G.A. Schuler, T. Sjöstrand, Towards a complete description of high-energy photoproduction. Nucl. Phys. B **407**, 539–605 (1993)
57. G.A. Schuler, T. Sjöstrand, A scenario for high-energy gamma gamma interactions. Z. Phys. C **73**, 677–688 (1997). arXiv:hep-ph/9605240
58. S. Okubo, Phi meson and unitary symmetry model. Phys. Lett. **5**, 165–168 (1963)
59. G. Zweig, An SU(3) model for strong interaction symmetry and its breaking. Version 1
60. J. Iizuka, Systematics and phenomenology of meson family. Prog. Theor. Phys. Suppl. **37**, 21–34 (1966)
61. G.A. Schuler, T. Sjöstrand, Hadronic diffractive cross-sections and the rise of the total cross-section. Phys. Rev. D **49**, 2257–2267 (1994)
62. B. Andersson, G. Gustafson, G. Ingelman, T. Sjöstrand, Parton fragmentation and string dynamics. Phys. Rep. **97**, 31–145 (1983)
63. T. Sjöstrand, M. van Zijl, A multiple interaction model for the event structure in hadron collisions. Phys. Rev. D **36**, 2019 (1987)
64. T. Sjöstrand, The development of MPI modeling in Pythia. Adv. Ser. Direct. High Energy Phys. **29**, 191–225 (2018). arXiv:1706.02166
65. C.O. Rasmussen, T. Sjöstrand, Hard diffraction with dynamic gap survival. JHEP **02**, 142 (2016). arXiv:1512.05525
66. S.J. Brodsky, P. Hoyer, C. Peterson, N. Sakai, The intrinsic charm of the proton. Phys. Lett. B **93**, 451–455 (1980)
67. M. Suzuki, Fragmentation of hadrons from heavy quark partons. Phys. Lett. B **71**, 139–141 (1977)
68. J.D. Bjorken, Properties of hadron distributions in reactions containing very heavy quarks. Phys. Rev. D **17**, 171–173 (1978)
69. NNPDF Collaboration, R.D. Ball et al., Parton distributions from high-precision collider data. Eur. Phys. J. C **77**(10), 663 (2017). arXiv:1706.00428
70. T.-J. Hou et al., New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC. Phys. Rev. D **103**(1), 014013 (2021). arXiv:1912.10053
71. S. Bailey, T. Cridge, L.A. Harland-Lang, A.D. Martin, R.S. Thorne, Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs. Eur. Phys. J. C **81**(4), 341 (2021). arXiv:2012.04684
72. NA3 Collaboration, J. Badier et al., Experimental determination of the pi meson structure functions by the Drell–Yan mechanism. Z. Phys. C **18**, 281 (1983)
73. NA10 Collaboration, B. Betev et al., Differential cross-section of high mass muon pairs produced by a 194-GeV/$c\pi^-$ beam on a tungsten target. Z. Phys. C **28**, 9 (1985)
74. NA10 Collaboration, P. Bordalo et al., Nuclear effects on the nucleon structure functions in hadronic high mass dimuon production. Phys. Lett. B **193**, 368 (1987)
75. J.S. Conway et al., Experimental study of muon pairs produced by 252-GeV pions on tungsten. Phys. Rev. D **39**, 92–122 (1989)
76. Saclay-CERN-College de France-Ecole Poly-Orsay Collaboration, J. Badier et al., Measurement of the K^-/π^- structure function ratio using the Drell–Yan process. Phys. Lett. B **93**, 354–356 (1980)
77. M. Glück, E. Reya, A. Vogt, Pionic parton distributions. Z. Phys. C **53**, 651–656 (1992)
78. P.J. Sutton, A.D. Martin, R.G. Roberts, W.J. Stirling, Parton distributions for the pion extracted from Drell–Yan and prompt photon experiments. Phys. Rev. D **45**, 2349–2359 (1992)
79. M. Glück, E. Reya, M. Stratmann, Mesonic parton densities derived from constituent quark model constraints. Eur. Phys. J. C **2**, 159–163 (1998). arXiv:hep-ph/9711369
80. M. Glück, E. Reya, I. Schienbein, Pionic parton distributions revisited. Eur. Phys. J. C **10**, 313–317 (1999). arXiv:hep-ph/9903288
81. B. Clerbaux, M.V. Polyakov, Partonic structure of pi and rho mesons from data on hard exclusive production of two pions off nucleon. Nucl. Phys. A **679**, 185–195 (2000). arXiv:hep-ph/0001332
82. R.M. Davidson, E.R. Arriola, Parton distributions functions of pion, kaon and eta pseudoscalar mesons in the NJL model. Acta Phys. Pol. B **33**, 1791–1808 (2002). arXiv:hep-ph/0110291
83. F. Bissey, J.R. Cudell, J. Cugnon, M. Jaminon, J.P. Lansberg, P. Stassart, A model for the pion structure function. Phys. Lett. B **547**, 210–218 (2002). arXiv:hep-ph/0207107
84. W. Detmold, W. Melnitchouk, A.W. Thomas, Parton distribution functions in the pion from lattice QCD. Phys. Rev. D **68**, 034025 (2003). arXiv:hep-lat/0303015
85. M. Aicher, A. Schäfer, W. Vogelsang, Soft-gluon resummation and the valence parton distribution function of the pion. Phys. Rev. Lett. **105**, 252003 (2010). arXiv:1009.2481
86. R.J. Holt, C.D. Roberts, Distribution functions of the nucleon and pion in the Valence region. Rev. Mod. Phys. **82**, 2991–3044 (2010). arXiv:1002.4666
87. C. Han, H. Xing, X. Wang, Q. Fu, R. Wang, X. Chen, Pion valence quark distributions from maximum entropy method. Phys. Lett. B **800**, 135066 (2020). arXiv:1809.01549
88. A. Watanabe, C.W. Kao, K. Suzuki, Meson cloud effects on the pion quark distribution function in the chiral constituent quark model. Phys. Rev. D **94**(11), 114008 (2016). arXiv:1610.08817
89. P.C. Barry, N. Sato, W. Melnitchouk, C.-R. Ji, First Monte Carlo global QCD analysis of pion parton distributions. Phys. Rev. Lett. **121**(15), 152001 (2018). arXiv:1804.01965

90. A. Watanabe, T. Sawada, M. Huang, Extraction of gluon distributions from structure functions at small x in holographic QCD. Phys. Lett. B **805**, 135470 (2020). arXiv:1910.10008
91. M. Ding, K. Raya, D. Binosi, L. Chang, C.D. Roberts, S.M. Schmidt, Symmetry, symmetry breaking, and pion parton distributions. Phys. Rev. D **101**(5), 054014 (2020). arXiv:1905.05208
92. I. Novikov et al., Parton distribution functions of the charged pion within the xFitter framework. Phys. Rev. D **102**(1), 014040 (2020). arXiv:2002.02902
93. A. Watanabe, T. Sawada, C.W. Kao, Kaon quark distribution functions in the chiral constituent quark model. Phys. Rev. D **97**(7), 074015 (2018). arXiv:1710.09529
94. A. Watanabe, T. Sawada, C.W. Kao, Meson cloud effects on kaon quark distribution functions and the SU(3) flavor symmetry. Nucl. Part. Phys. Proc. **300–302**, 121–125 (2018). arXiv:1810.04032
95. C. Shi, C. Mezrag, H.-S. Zong, Pion and kaon valence quark distribution functions from Dyson–Schwinger equations. Phys. Rev. D **98**(5), 054029 (2018). arXiv:1806.10232
96. J. Lan, C. Mondal, S. Jia, X. Zhao, J.P. Vary, Pion and kaon parton distribution functions from basis light front quantization and QCD evolution. Phys. Rev. D **101**(3), 034024 (2020). arXiv:1907.01509
97. Z.-F. Cui, M. Ding, F. Gao, K. Raya, D. Binosi, L. Chang, C.D. Roberts, J. Rodríguez-Quintero, S.M. Schmidt, Kaon and pion parton distributions. Eur. Phys. J. C **80**(11), 1064 (2020)
98. M. Glück, R.M. Godbole, E. Reya, Dynamically generated parton distributions for high-energy collisions. Z. Phys. C **41**, 667 (1989)
99. M. Glück, E. Reya, A. Vogt, Parton distributions for high-energy collisions. Z. Phys. C **53**, 127–134 (1992)
100. M. Glück, E. Reya, A. Vogt, Dynamical parton distributions of the proton and small x physics. Z. Phys. C **67**, 433–448 (1995)
101. M. Glück, E. Reya, A. Vogt, Dynamical parton distributions revisited. Eur. Phys. J. C **5**, 461–470 (1998). arXiv:hep-ph/9806404
102. M. Glück, P. Jimenez-Delgado, E. Reya, Dynamical parton distributions of the nucleon and very small-x physics. Eur. Phys. J. C **53**, 355–366 (2008). arXiv:0709.0614
103. NNPDF Collaboration, R.D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N.P. Hartland, J. Rojo, Parton distributions with QED corrections. Nucl. Phys. B **877**, 290–320 (2013). arXiv:1308.0598
104. V.N. Gribov, L.N. Lipatov, Deep inelastic e p scattering in perturbation theory. Sov. J. Nucl. Phys. **15**, 438–450 (1972)
105. Y.L. Dokshitzer, Calculation of the structure functions for deep inelastic scattering and e+ e- annihilation by perturbation theory in quantum chromodynamics. Sov. Phys. JETP **46**, 641–653 (1977)
106. G. Altarelli, G. Parisi, Asymptotic freedom in parton language. Nucl. Phys. B **126**, 298–318 (1977)
107. M. Botje, QCDNUM: fast QCD evolution and convolution. Comput. Phys. Commun. **182**, 490–532 (2011). arXiv:1005.1481
108. S. Alekhin et al., HERAFitter. Eur. Phys. J. C **75**(7), 304 (2015). arXiv:1410.4412
109. V. Bertone, S. Carrazza, J. Rojo, APFEL: a PDF evolution library with QED corrections. Comput. Phys. Commun. **185**, 1647–1668 (2014). arXiv:1310.1394
110. LHCf Collaboration, O. Adriani et al., Measurement of forward neutral pion transverse momentum spectra for $\sqrt{s}$ = 7TeV proton–proton collisions at LHC. Phys. Rev. D **86**, 092001 (2012). arXiv:1205.4578
111. LHCf Collaboration, O. Adriani et al., Measurement of very forward neutron energy spectra for 7 TeV proton–proton collisions at the Large Hadron Collider. Phys. Lett. B **750**, 360–366 (2015). arXiv:1503.03505
112. V. Kireyeu, I. Grishmanovskii, V. Kolesnikov, V. Voronyuk, E. Bratkovskaya, Hadron production in elementary nucleon-nucleon reactions from low to ultra-relativistic energies. Eur. Phys. J. A **56**(9), 223 (2020). arXiv:2006.14739
113. B. Andersson, G. Gustafson, T. Sjöstrand, Baryon production in jet fragmentation and Υ decay. Phys. Scr. **32**, 574 (1985)
114. B. Andersson, G. Gustafson, I. Holgersson, O. Mansson, A model for the reaction mechanism and the baryon fragmentation distributions in low p_T hadronic interactions. Nucl. Phys. B **178**, 242–262 (1981)
115. P. Edén, G. Gustafson, Baryon production in the string fragmentation picture. Z. Phys. C **75**, 41–49 (1997). arXiv:hep-ph/9606454
116. P. Edén, A Program for baryon generation and its applications to baryon fragmentation in DIS. arXiv:hep-ph/9610246
117. F. Kling, Forward neutrino fluxes at the LHC. arXiv:2105.08270
118. A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, G. Watt, LHAPDF6: parton density access in the LHC precision era. Eur. Phys. J. C **75**, 132 (2015). arXiv:1412.7420
119. M. Cacciari, G.P. Salam, G. Soyez, The anti-k_t jet clustering algorithm. JHEP **04**, 063 (2008). arXiv:0802.1189
120. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. C **72**, 1896 (2012). arXiv:1111.6097
121. R.J. Glauber, Cross-sections in deuterium at high-energies. Phys. Rev. **100**, 242–248 (1955)
122. A. Bialas, M. Bleszynski, W. Czyz, Multiplicity distributions in nucleus–nucleus collisions at high-energies. Nucl. Phys. B **111**, 461–476 (1976)
123. B. Andersson, G. Gustafson, B. Nilsson-Almqvist, A model for low p(t) hadronic reactions, with generalizations to hadron–nucleus and nucleus–nucleus collisions. Nucl. Phys. B **281**, 289–309 (1987)
124. C. Bierlich, G. Gustafson, L. Lönnblad, A. Tarasov, Effects of overlapping strings in pp collisions. JHEP **03**, 148 (2015). arXiv:1412.6259
125. C. Bierlich, Rope hadronization and strange particle production. EPJ Web Conf. **171**, 14003 (2018). arXiv:1710.04464
126. C. Bierlich, G. Gustafson, L. Lönnblad, Collectivity without plasma in hadronic collisions. Phys. Lett. B **779**, 58–63 (2018). arXiv:1710.09725
127. C. Bierlich, S. Chakraborty, G. Gustafson, L. Lönnblad, Setting the string shoving picture in a new frame. JHEP **03**, 270 (2021). arXiv:2010.07595
128. H1 Collaboration, A. Aktas et al., Measurement and QCD analysis of the diffractive deep-inelastic scattering cross-section at HERA. Eur. Phys. J. C **48**, 715–748 (2006). arXiv:hep-ex/0606004
129. National Weather Service, Layers of the atmosphere. https://www.weather.gov/jetstream/layers (2019)
130. ISO, "Standard atmosphere," ISO 2533:1975, International Organization for Standardization, Geneva, Switzerland (2007)
131. J. Bellm et al., Herwig 7.2 release note. Eur. Phys. J. C **80**(5), 452 (2020). arXiv:1912.06509
132. Sherpa Collaboration, E. Bothmann et al., Event Generation with Sherpa 2.2. SciPost Phys. **7**(3), 034 (2019). arXiv:1905.09127