THE EUROPEAN
PHYSICAL JOURNAL A

Hadronic rescattering in pA and AA collisions

Christian Bierlich, Torbjörn Sjöstrand, Marius Utheim[a]

Theoretical Particle Physics, Department of Astronomy and Theoretical Physics, Lund University, Sölvegatan 14A, 223 62 Lund, Sweden

Abstract In a recent article we presented a model for hadronic rescattering, and some results were shown for pp collisions at LHC energies. In order to extend the studies to pA and AA collisions, the ANGANTYR model for heavy-ion collisions is taken as the starting point. Both these models are implemented within the general-purpose Monte Carlo event generator PYTHIA, which makes the matching reasonably straightforward, and allows for detailed studies of the full space–time evolution. The rescattering rate is significantly higher than in pp, especially for central AA collisions, where the typical primary hadron rescatters several times. We study the impact of rescattering on a number of distributions, such as $p_\perp$ and η spectra, and the space–time evolution of the whole collision process. Notably rescattering is shown to give a significant contribution to elliptic flow in XeXe and PbPb, and to give a nontrivial impact on charm production.

1 Introduction

Heavy-ion experiments at RHIC and LHC have produced convincing evidence that a Quark-Gluon Plasma (QGP) is formed in high-energy nucleus-nucleus (AA) collisions. The discussion therefore has developed into one of understanding the underlying detailed mechanisms, such as the nature of the initial state, the early thermalization, the subsequent hydrodynamical expansion, and the transition back to a hadronic state. Numerous models have been and are being developed to study such issues.

The standard picture of heavy ion collisions, separates the evolution of the QGP phase into three or four stages, outlined in the following.

The first < 1 fm after the collision, is denoted the "initial state". It consists of dense matter, highly out of equilibrium. Most QGP-based models seek to calculate an energy density

(or a full energy-momentum tensor) from a model of the evolution of the initial stage. The simplest approaches are based purely on geometry, and are denoted Glauber models [1]. Here, the energy density in the transverse plane is determined purely from the distributions of nucleons in the incoming nuclei. Going beyond nucleonic degrees of freedom, some of the more popular choices includes either introducing constituent quarks [2], or invoking the more involved formalism known the Colour Glass Condensate [3]. In the latter case, the so-called IP-Glasma [4] program is often used, as it allows for computations with realistic boundary conditions.

The initial state, glasma or not, will then transition into a plasma. Recently, progress has been made to describe the transition from an out-of-equilibrium initial state to a hydrodynamized plasma, using kinetic theory [5]. In such cases, the pre-equilibration will describe the dynamics between ≈ 0.1–1 fm.

Between 1–10 fm, the plasma evolves according to relativistic viscous hydrodynamics [6–8]. Hydrodynamics is a long wavelength effective theory, able to describe interactions at low momentum, when the mean free path of particles is much smaller than the characteristic size of the system. As such, its use has been criticised in small collision systems, but nevertheless seems to be able to describe flow observables reasonably well even there [9].

Finally, after 10 fm, the QGP freezes out to hadronic degrees of freedom. In a nutshell, this paper is about the physics involved after this freeze-out, but with the significant difference that we reach this state without invoking any QGP-based models.

Paradoxically, one of the key problems is that the QGP picture has been too successful. QGP formation was supposed to be unique to AA collisions, while pA and pp collisions would not involve volumes and time scales large enough for it. And yet QGP-like signals have been found in these as well. One key example is the observation of a non-isotropic particle flow, in the form of a "ridge" at the same

[a] e-mail: marius.utheim@thep.lu.se (corresponding author)

azimuthal angle as a trigger jet [10–12] or of non-vanishing v_2 azimuthal flow coefficients [11–13]. Another example is that the fraction of strange hadrons, and notably multi-strange baryons, is smoothly increasing from low-multiplicity to high-multiplicity pp, on through pA to saturate for AA multiplicities [14].

The most obvious way out is to relax the large-volume requirement, and accept that a QGP, or at least a close-to-QGP-like state, can be created in smaller systems. An excellent example of this approach is the core–corona model [15], implemented in the EPOS event generator [16], wherein the high-density core of a system hadronizes like a plasma, while the outer lower-density corona does not. The evolution from low-multiplicity pp to AA is then a consequence of an increasing core fraction.

Another approach is to ask what physics mechanisms, not normally modelled in pp collisions, would be needed to understand pp data without invoking QGP formation. And, once having such a model, one could ask what consequences that would imply for pA and AA collisions. More specifically, could some of the signals attributed to QGP formation have alternative explanations? If nothing else, exploring these questions could help sharpen experimental tests, by providing a straw-man model. At best, we may actually gain new insights.

This is the road taken by the ANGANTYR model [17,18]. It is based on and contained in the PYTHIA event generator [19,20], which successfully describes many/most features of LHC pp events. ANGANTYR adds a framework wherein pA and AA collisions can be constructed as a superposition of simpler binary collisions, in the spirit of the old FRITIOF model [21,22]. Such a framework is already sufficient to describe many simple pA and AA distributions, such as $dn_{charged}/d\eta$. Beyond that, it also offers a platform on top of which various collective non-QGP phenomena can be added. One example is shoving [23–25], whereby closely overlapping colour fields repel each other, to give a collective flow. Another is colour rope formation [26], wherein overlapping colour fields can combine to give a higher field strength, thus enhancing strangeness production relative to the no-overlap default.

In this article we will study a third mechanism, that of hadronic rescattering, without assuming (a) any partonic component of final state interactions (such as ropes or shoving mentioned above), nor (b) any thermalized QGP. The basic idea here is that the standard fragmentation process produces a region of closely overlapping hadrons, that then can collide with each other as the system expands. Each single such collision on its own will give negligible effects, but if there are many of them then together they may give rise to visible physics signals. Rescattering is often used as an afterburner to the hadronization of the QGP, commonly making use of programs such as UrQMD [27], SMASH [28], HSD

[29] or AMPT [30] (for a comparison between some existing rescattering codes, see e.g. [31]). In order to use a rescattering framework as an afterburner to ANGANTYR, a first step is to describe the space–time structure of hadronization in PYTHIA, which was worked out in [32]. This picture can easily be extended from pp to pA and AA using the nuclear geometry set up in ANGANTYR. Thereby the road is open to add rescattering e.g. with UrQMD, which was done by Ref. [33]. One significant difference between using ANGANTYR and a QGP-based model is that in the former case, hadronization occurs much sooner than the corresponding process in the latter, producing a denser hadronic state, and therefore rescattering can hypothetically give more significant effects.

Using two different programs is cumbersome, however. It requires the user to learn to use each individual framework, and they have to convert the output from the first program into a format that can be input to the second. A related issue arises if the two programs represent event records differently, so that it might be impossible to trace the full particle history. A desire for convenience is one of the main motivations behind a recently developed framework for hadronic rescattering, implemented natively in PYTHIA [34]. This framework was inspired by UrQMD in the way it handles some of the processes and cross, sections. With this framework, rescattering can be enabled with just a single additional line of code, which is a trivial task for anyone already familiar with PYTHIA. In addition, this framework also introduces physics features not found in some other frameworks like UrQMD, such as a basic model for charm and bottom hadrons in rescattering, and with PYTHIA being in active development, there is a low threshold for making further improvements in the future.

In [34], initial studies using the framework were limited to implications for pp collisions, which not unexpectedly were found to be of moderate size. That is, while visible enough in model studies, generally they are less easy to pin down experimentally, given all other uncertainties that also exist. The goal of this article is to extend the rescattering studies to pA and AA collisions via ANGANTYR, where effects are expected to be larger. Indeed, as we shall see, the outcome confirms this expectation. The number of rescatterings rises faster than the particle multiplicity, such that the fraction of not-rescattered hadrons is small in PbPb collisions. Rescatterings are especially enhanced at lower masses, but the process composition at a given mass is universal. Obviously the primary production volume increases from pp and pA to AA, and thus so does the range of rescatterings. Transverse momentum spectra are significantly more deformed by rescattering in AA. There is a clear centrality dependence on particle production rates, e.g. a J/ψ depletion in central collisions. The most interesting result is a clear signal of elliptic flow induced by rescatterings, that even matches experimen-

tal PbPb numbers at large multiplicities, to be contrasted with the miniscule effects in pp.

The outline of the article is as follows. In Sect. 2 we describe the main points of the model, from the simulation of the nuclear collision, through the modelling of individual nucleon–nucleon sub-collisions and on to the rescattering framework proper. In Sect. 3 effects in this model are tested on its own, while Sect. 4 shows comparisons with data. Some conclusions and an outlook are presented in Sect. 5. Finally, technical aspects related to computation time for rescattering are discussed in the appendix.

Natural units are assumed throughout the article, i.e. $c = \hbar = 1$. Energy, momentum and mass are given in GeV, space and time in fm, and cross sections in mb.

2 The model

In this section we will review the framework used to simulate nuclear collisions. Initially the ANGANTYR framework is used to set the overall nucleus–nucleus (AA) collision geometry and select colliding nucleon-nucleon (NN) pairs. Then the multiparton interactions (MPI) concept is used to model each single NN collision. The resulting strings are fragmented to provide the primary setup of hadrons, that then can begin to decay and rescatter. All of these components are described in separate publications, where further details may be found, so only the key aspect are collected here is to describe how it all hangs together.

2.1 ANGANTYR

The ANGANTYR part of the modelling is responsible for setting up the AA collision geometry, and selecting the number and nature of the ensuing NN collisions [18].

Take the incoming high-energy nucleons to be travelling along the $\pm z$ directions. By Lorentz contraction all the NN collisions then occur in a negligibly small range around $t = z = 0$, and the nucleon transverse (x, y) positions can be considered frozen during that time. The nucleon locations inside a nucleus are sampled according to a two-dimensional Woods–Saxon distribution in the GLISSANDO parametrisation [35,36], applicable for heavy nuclei with $A > 16$, and with a nuclear repulsion effect implemented algorithmically as a "hard core" radius of each nucleon, below which two nucleons cannot overlap. The AA collision impact parameter provides an offset $\pm b_{AA}/2$, e.g. along the x axis. Up to this point, this is a fairly standard Glauber model treatment, where one would then combine the geometry with measured cross sections (usually total and/or inelastic non-diffractive), to obtain the amount of participating or wounded nucleons, and the number of binary sub-collisions (see e.g. Ref. [1] for a review). In ANGANTYR, a distinction between nucleons

wounded inelastic non-diffractively, diffractively or elastically is desired, along with a dependence on the nucleon-nucleon impact parameter. To this end, a parametrization of the nucleon-nucleon elastic amplitude in impact parameter space ($T(\mathbf{b})$) is used. It allows for the calculation of the amplitude $T_{kl}(\mathbf{b})$ for any combination of projectile and target state, k and l respectively. All parameters of the parametrization can be estimated from proton-proton total and semi-inclusive cross sections, and varies with collision energy. The input cross sections used are the ones available in PYTHIA, with the SaS model [37] being the default choice. The parametrization of $T(\mathbf{b})$ thus adds no new parameters beyond the ones already present in the model for hadronic cross sections.

Inelastic non-diffractive collisions involve a colour exchange between two nucleons. In the simplest case, where each incoming nucleon undergoes at most one collision, the traditional PYTHIA collision machinery can be used essentially unchanged. The one difference is that the nuclear geometry has already fixed the NN impact parameter b_{NN}, whereas normally this would be set only in conjunction with the hardest MPI.

The big extension of ANGANTYR is that it also handles situations where a given nucleon A interacts inelastic non-diffractively with several nucleons $B_1, B_2, \ldots B_n$ from the other nucleus. Colour fields would then be stretched from A to each B_i. It would be rare for all the fields to stretch all the way out to A, however, but rather matching colour–anticolour pairs would "short-circuit" most of the colour flow out to the remnants. Such a mechanism is already used for MPIs in a single NN collision, but here it is extended to the full set of interconnected nucleons. Therefore only one AB_i collision is handled as a normal NN one, while the other AB_j, $j \neq i$ ones will produce particles over a smaller rapidity range. This is analogous to the situation encountered in single diffraction $AB_j \rightarrow AX_j$. If we further assume that the short-circuiting can occur anywhere in rapidity with approximately flat probability distribution, this translates into an excited mass spectrum like $dM_{X_j}^2/M_{X_j}^2$, again analogous to diffraction. To this end, $n-1$ carrier particles with vacuum quantum numbers $\mathbb{P}_j$ (denoted $\mathbb{P}$ for the similarity with pomerons) are emitted, with fractions x_j of the incoming A (lightcone) momentum picked according to dx_j/x_j, subject to momentum conservation constraints, with a leftover x_i that usually should represent the bulk of the A momentum. Thereby the complexity of the full problem is reduced to one of describing one regular AB_i collision, at a slightly reduced energy, and $n-1$ $\mathbb{P}_j B_j$ collisions, at significantly reduced energies, similar to diffraction. The pomeron-like objects have no net colour or flavour, but they do contain partons and the full MPI machinery can be applied to describe also these collisions. As the particles are not true pomerons, the PDFs can be different from the pomeron ones measured at HERA, and the transverse size is that of the

original nucleon rather than the smaller one expected for a pomeron.

In a further step of complexity, the nucleons on side A and B may be involved in multiply interrelated chains of interactions. Generalizing the principles above, it is possible to reduce even complex topologies to a set of decoupled NN, N$\mathbb{P}$, and $\mathbb{PP}$ collisions, to be described below. The reduction is not unique, but may be chosen randomly among the allowed possibilities.

One current limitation is that there is no description of the breakup of the nuclear remnant. Rather, all non-wounded nucleons of a nucleus are collected together into a single fictitious new nucleus, that is not considered any further.

Other approaches, such as HIJING [38], exist to combine the PYTHIA nucleon-nucleon scatterings with a nucleus geometry. Like ANGANTYR, HIJING also draws inspiration from the even earlier FRITIOF model. They both take a Glauber calculation as their starting points, and then use PYTHIA to generate hard partonic sub-collisions and PYTHIA string fragmentation for hadronisation. The main difference between ANGANTYR and HIJING is the treatment of fluctuations and coherence effects. While ANGANTYR accounts for fluctuations in the individual states of the nucleons (denoted k, l above), and introduces the concept of secondary wounded nucleons as a dynamical, Regge-based model for interactions between one projectile and several target nucleons, HIJING does neither. Instead HIJING introduces a geometry depending "shadowing factor", reducing the nucleon-nucleon cross section by a constant. This allows HIJING to reproduce average multiplicities by reducing the nucleon-nucleon cross section with up to 50% in AuAu collisions at $\sqrt{s_{NN}} = 200$ GeV. For the purpose of tracing out and investigating the effect of hadronic rescattering depending on detailed event geometry, we find the dynamical modelling of secondary wounded nucleons in ANGANTYR more appealing than a rather ad hoc scaling factor. We refer the reader to Ref. [18], where a more detailed comparison between ANGANTYR, HIJING and other models is carried out.

2.2 Multiparton interaction vertices

At the end of the ANGANTYR modelling, a set of separate hadron–hadron (HH) interactions have been defined inside an AA collision, where the hadron can be either a nucleon or a pomeron-like object as discussed above. The locations of the HH collisions in the transverse plane is also fixed.

When two Lorentz-contracted hadrons collide inelastically with each other, a number of separate (semi-)perturbative parton–parton interactions can occur. These are modelled in a sequence of falling transverse momenta $p_\perp$, as described in detail elsewhere [39,40]. The MPI vertices are spread over a transverse region of hadronic size, but in the past it was not necessary to assign an explicit location for every single MPI. Now it is. The probability for an interaction at a given transverse coordinate (x, y) can be assumed proportional to the overlap of the parton densities of the colliding hadrons in that area element. A few possible overlap function options are available in PYTHIA, where the Gaussian case is the simplest one. If two Gaussian-profile hadrons pass with an impact parameter b_{HH}, then the nice convolution properties gives a total overlap that is a Gaussian in b_{HH}, and the distribution of MPI vertices is a Gaussian in (x, y). Specifically note that there is no memory of the collision plane in the vertex distribution.

This property is unique to Gaussian convolutions, however. In general, the collision region will be elongated either out of or in to the collision plane. The former typically occurs for a distribution with a sharper proton edge, e.g. a uniform ball, which gives rise to the almond-shaped collision region so often depicted for heavy-ion collisions. The latter shape instead occurs for distributions with a less pronounced edge, such as an exponential. The default PYTHIA behaviour is close to Gaussian, but somewhat leaning towards the latter direction. Even that is likely to be a simplification. The evolution of the incoming states by initial-state cascades is likely to lead to "hot spots" of increased partonic activity, see e.g. [41]. A preliminary study in [34] showed that azimuthal anisotropies in the individual HH collision give unambiguous, but miniscule flow effects, and furthermore the many HH event planes of an AA collision point in random directions, further diluting any such effects. In the end, it is the asymmetries related to the AA geometry that matter for our studies.

Only a fraction of the full nucleon momentum is carried away by the MPIs of an HH collision, leaving behind one or more beam remnants [42]. These are initially distributed according to a Gaussian shape around the center of the respective hadron. By the random fluctuations, and by the interacting partons primarily being selected on the side leaning towards the other beam hadron, the "center of gravity" will not agree with the originally assumed origin. All the beam remnants will therefore be shifted so as to ensure that the energy-weighted sum of colliding and remnant parton locations is where it should be. Shifts are capped to be at most a proton radius, so as to avoid extreme spatial configurations, at the expense of a perfectly aligned center of gravity.

Not all hadronizing partons are created in the collision moment $t = 0$. Initial-state radiation (ISR) implies that some partons have branched off already before this, and final-state radiation (FSR) that others do it afterwards. These partons then can travel some distance out before hadronization sets in, thereby further complicating the space–time picture, even if the average time of parton showers typically is a factor of five below that of string fragmentation [32]. We do not trace the full shower evolution, but instead include a smearing of the transverse location in the collision plane that a parton

points back to. No attempt is made to preserve the center of gravity during these fluctuations.

The partons produced in various stages of the collision process (MPIs, ISR, FSR) are initially assigned colours according to the $N_C \rightarrow \infty$ approximation, such that different MPI systems are decoupled from each other. By the beam remnants, which have as one task to preserve total colour, these systems typically become connected with each other through the short-circuiting mechanism already mentioned. Furthermore, colour reconnection (CR) is allowed to swap colours, partly to compensate for finite-N_C effects, but mainly that it seems like nature prefers to reduce the total string length drawn out when two nearby strings overlap each other. When such effects have been taken into account, what remains to hadronize is one or more separate colour singlet systems.

2.3 Hadronization

Hadronization is modelled in the context of the Lund string fragmentation model [43]. In it, a linear confinement is assumed, i.e. a string potential of $V = \kappa r$, where the string tension $\kappa \approx 1$ GeV/fm and r is the separation between a colour triplet–antitriplet pair. For the simplest possible case, that of a back-to-back $q\bar{q}$ pair, the linearity leads to a straightforward relationship between the energy–momentum and the space–time pictures:

$$\left| \frac{\mathrm{d}p_{z,\mathrm{q/\bar{q}}}}{\mathrm{d}t} \right| = \left| \frac{\mathrm{d}p_{z,\mathrm{q/\bar{q}}}}{\mathrm{d}z} \right| = \left| \frac{\mathrm{d}E_{\mathrm{q/\bar{q}}}}{\mathrm{d}t} \right| = \left| \frac{\mathrm{d}E_{\mathrm{q/\bar{q}}}}{\mathrm{d}z} \right| = \kappa. \qquad (1)$$

If there is enough energy, the string between an original $q_0\bar{q}_0$ pair may break by producing new $q_i\bar{q}_i$ pairs, where the intermediate q_i ($\bar{q}_i$) are pulled towards the $\bar{q}_0$ (q_0) end, such that the original colour field is screened. This way the system breaks up into a set of n colour singlets $q_0\bar{q}_1 - q_1\bar{q}_2 - q_2\bar{q}_3 - \ldots - q_{n-1}\bar{q}_0$, that we can associate with the primary hadrons. By (1) the location of the breakup vertices in space–time is linearly related to the energy–momentum of the hadrons produced between such vertices [32].

When quarks with non-vanishing mass or $p_\perp$ are created, they have to tunnel out a distance before they can end up on mass shell. This tunnelling process gives a suppression of heavier quarks, like s relative to u and d ones, and an (approximately) Gaussian distribution of the transverse momenta. Effective equivalent massless-case production vertices can be defined. Baryons can be introduced e.g. by considering diquark–antidiquark pair production, where a diquark is a colour antitriplet and thus can replace an antiquark in the flavour chain.

Having simultaneous knowledge of both the energy–momentum and the space–time picture of hadron production

violates the Heisenberg uncertainty relations. In this sense the string model should be viewed as a semiclassical one. The random nature of the Monte Carlo approach will largely mask the issue, and smearing factors are introduced in several places to further reduce the tension.

A first hurdle is to go on from a simple straight string to a longer string system. In the limit where the number of colours is large, the $N_C \rightarrow \infty$ approximation [44], a string typically will be stretched from a quark end via a number intermediate gluons to an antiquark end, where each string segment is stretched between a matching colour-anticolour pair. To first approximation each segment fragments as a boosted copy of a simple $q\bar{q}$ system, but the full story is more complicated, with respect to what happens around each gluon. Firstly, if a gluon has time to lose its energy before it has hadronized, the string motion becomes more complicated. And secondly, even if not, a hadron will straddle each gluon kink, with one string break in each of the two segments it connects. A framework to handle energy and momentum sharing in such complicated topologies was developed in Ref. [45], and was then extended to reconstruct matching space–time production vertices in [32]. This includes many further details not covered here, such as a transverse smearing of breakup vertices, to represent a width of the string itself, and various safety checks.

In addition to the main group of open strings stretched between $q\bar{q}$ endpoints, there are two other common string topologies. One is a closed gluon loop. It can be brought back to the open-string case by a first break somewhere along the string. The other is the junction topology, represented by three quarks moving out in a different directions, each pulling out a string behind itself. These strings meet at a common junction vertex, to form a Y-shaped topology. This requires a somewhat more delicate extensions of the basic hadronization machinery.

One complication is that strings can be stretched between partons that do not originate from the same vertex. In the simplest case, a q connected with a $\bar{q}$ from a different MPI, the vertex separation could be related to a piece of string already at $t = 0$. At the small distances involved it is doubtful whether the full string tension is relevant, in particular since the net energy associated with such initial strings should not realistically exceed the proton mass. Since this energy is then to be spread over many of the final-state hadrons, the net effect on each hardly would be noticeable, and is not modelled.

For the space–time picture we do want to be somewhat more careful about the effects of the transverse size of the original source. Even an approximate description would help smear the hadron production vertices in a sensible manner. To begin, consider a simple $q\bar{q}$ string, where the relevant length of each hadron string piece is related to its energy. For a given hadron, define E_{hq} ($E_{\mathrm{h\bar{q}}}$) as half the energy of the hadron plus the full energy of all hadrons lying between it and the q ($\bar{q}$)

end, and use this as a measure of how closely associated a hadron is with the respective endpoint. Also let $\mathbf{r}_{\perp q}$ ($\mathbf{r}_{\perp \bar{q}}$) be the (anti)quark transverse production coordinates. Then define the hadron production vertex offset to be

$$
\begin{aligned}
\Delta \mathbf{r}_{\perp h} &= \frac{E_{h\bar{q}} \, \mathbf{r}_{\perp q} + E_{hq} \, \mathbf{r}_{\perp \bar{q}}}{E_{hq} + E_{h\bar{q}}} \\
&= \frac{(E_{tot} - E_{hq}) \, \mathbf{r}_{\perp q} + E_{hq} \, \mathbf{r}_{\perp \bar{q}}}{E_{tot}},
\end{aligned}
\tag{2}
$$

relative to what a string motion started at the origin would have given.

This procedure is then generalized to more complicated string topologies. Again energy is summed up from one string end, for partons and hadrons alike, to determine which string segment a given hadron is most closely associated with, and how the endpoints of that segment should be mixed. Note that, although energy is not a perfect measure of location along the string, the comparison between parton and hadron energies is only mildly Lorentz-frame dependent, which is an advantage. More complicated string topologies, like junction ones, require further considerations not discussed here. Again we stress that the main point is not to provide a perfect location for each individual hadron, but to model the average effects.

2.4 The hadronic rescattering formalism

By the procedure outlined so far, each primary produced hadron has been assigned a production vertex $x_0 = (t_0, \mathbf{x}_0)$ and a four-momentum $p = (E, \mathbf{p})$. The latter defines its continued motion along straight trajectories $\mathbf{x}(t) = \mathbf{x}_0 + (t - t_0) \mathbf{p}/m$. Consider now two particles produced at x_1 and x_2 with momenta p_1 and p_2. Our objective is to determine whether these particles will scatter and, if so, when and where. To this end, the candidate collision is studied in the center-of-momentum frame of the two particles. If they are not produced at the same time, the position of the earlier one is offset to the creation time of the later one. Particles moving away from each other already at this common time are assumed unable to scatter.

Otherwise, the probability P of an interaction is a function of the impact parameter b, the center-of-mass energy E_{CM}, and the two particle species A and B. There is no solid theory for the b dependence of P, so a few different options are implemented, such as a black disk, a grey disk or a Gaussian. In either case the normalization is such that $\int P(b) \, d^2 b = \sigma_{AB}(E_{CM})$. To first approximation all options thus give the same interaction rate, but the drop of hadronic density away from the center in reality means fewer interactions for a broader distribution.

If it is determined that the two particles will interact, the interaction time is defined as the time of closest approach in the rest frame, which is similar to the approach used by UrQMD [27]. The spatial component of the interaction vertex depends on the character of the collision. Elastic and diffractive processes can be viewed as t-channel exchanges of a pomeron (or reggeon), and then it is reasonable to let each particle continue out from its respective location at the interaction time. For other processes, where either an intermediate s-channel resonance is formed or strings are stretched between the remnants of the two incoming hadrons, an effective common interaction vertex is defined as the average of the two hadron locations at the interaction time. In cases where strings are created, be it by s-channel processes or by diffraction, the hadronization starts around this vertex and is described in space–time as already outlined. This means that after the interaction there is a delay before the new hadrons are formed and can begin to interact. For processes where the outgoing particles are produced directly, such as elastic scattering or an intermediate resonance decay, the particles are not allowed to rescatter again before some time has passed. This "effective formation time" is chosen according to an exponential $\exp(\tau/\tau_{formation})$, where $\tau_{formation} = 1$ fm by default. Alternatively, this can be expressed as a reduced rescattering cross section at early times.

In actual events with many hadrons, each hadron pair is checked to see if it fulfils the interaction criteria and, if it does, the interaction time for that pair (in the CM frame of the event) is recorded in a time-ordered list. Furthermore, unstable particles can decay during the rescattering phase. For these, an invariant lifetime τ is picked at random according to an exponential $\exp(-\tau/\tau_0)$, where $\tau_0 = 1/\Gamma_0$ is the inverse of the mass-independent width. The resulting decay times are inserted into the same list. Then the scattering or decay that is first in time order is simulated, unless the particles involved have already interacted/decayed. This produces new hadrons that are checked for rescatterings or decays, and any such are inserted into the time-ordered list. This process is repeated until there are no more potential interactions.

There are some obvious limitations to the approach as outlined so far:

- The procedure is not Lorentz invariant, since the time-ordering of interactions is defined in the CM frame. We do not expect this to be a major issue. This has been studied and confirmed within existing rescattering approaches [27,28,46], and reconfirmed in our pp studies.
- Currently only collisions between two incoming hadrons are considered, even though in a dense environment one would also expect collisions involving three or more hadrons. This is a more relevant restriction, that may play a role for some observables, and to be considered in the future.
- Since traditional PYTHIA tunes do not include rescattering effects, some retuning to pp events has to be made

before the model is applied to AA ones. For now, only the simplest possible one is used, wherein the $p_{\perp 0}$ parameter of the MPI framework is increased slightly so as to restore the same average charged multiplicity in proton collisions at LHC energies as without rescattering.

- All modelled subprocesses are assumed to share the same hadronic impact-parameter profile. In a more detailed modelling the t-channel elastic and diffractive processes should be more peripheral than the rest, and display an approximately inverse relationship between the t and b values.
- The model only considers the effect of hadrons colliding with hadrons, not those of strings colliding/overlapping with each other or with hadrons. An example of the former is the already-introduced shoving mechanism. Both shoving and rescattering act to correlate the spatial location of strings/hadrons with a net push outwards, giving rise to a radial flow. Their effects should be combined, but do not add linearly since an early shove leads to a more dilute system of strings and primary hadrons, and thereby less rescattering.

2.5 Hadronic rescattering cross sections

A crucial input for deciding whether a scattering can occur is the total cross section. Once a potential scattering is selected, it also becomes necessary to subdivide this total cross section into a sum of partial cross sections, one for each possible process, as these are used to represent relative abundances for each process to occur. A staggering amount of details enter in such a description, owing to the multitude of incoming particle combinations and collision processes. To wit, not only "long-lived" hadrons can collide, i.e. π, K, η, η', p, n, Λ, Σ, Ξ, Ω, and their antiparticles, but also a wide selection of short-lived hadrons, starting with ρ, K*, ω, ϕ, Δ, Σ^* and Ξ^*. Required cross sections are described in detail in Ref. [34], and we only provide a summary of the main concepts here.

Of note is that most rescatterings occur at low invariant masses, typically only a few GeV. Therefore the descriptions are geared to this mass range, and cross sections are not necessarily accurate above 10 GeV. Furthermore event properties are modelled without invoking any perturbative activity, i.e. without MPIs. We will see in Sect. 3.2 that the number of interactions above 10 GeV is small enough that these discrepancies can safely be disregarded.

For this low-energy description, the following process types are available:

- Elastic interactions are ones where the particles do not change species, i.e. $AB \rightarrow AB$. In our implementation, these are considered different from elastic scattering through a resonance, e.g. $\pi^+\pi^- \rightarrow \rho^0 \rightarrow \pi^+\pi^-$,

although the two could be linked by interference terms. In experiments, usually all $AB \rightarrow AB$ events are called elastic because it is not possible to tell which underlying mechanism is involved.
- Resonance formation typically can be written as $AB \rightarrow R \rightarrow CD$, where R is the intermediate resonance. This can only occur when one or both of A and B are mesons. It is the resonances that drive rapid and large cross-section variations with energy, since each (well separated) resonance should induce a Breit-Wigner peak.
- Annihilation is specifically aimed at baryon–antibaryon collisions where the baryon numbers cancel out and gives a mesonic final state. It is assumed to require the annihilation of at least one $q\bar{q}$ pair. This is reminiscent of what happens in resonance formation, but there the final state is a resonance particle, while annihilation forms strings between the outgoing quarks.
- Diffraction of two kinds are modelled here: single $AB \rightarrow XB$ or $AB \rightarrow AX$ and double $AB \rightarrow X_1X_2$. Here X represents a massive excited state of the respective incoming hadron, and there is no net colour or flavour exchange between the two sides of the event.
- Excitation can be viewed as the low-mass limit of diffraction, where either one or both incoming hadrons are excited to a related higher resonance. It can be written as $AB \rightarrow A^*B$, $AB \rightarrow AB^*$ or $AB \rightarrow A^*B^*$. Here A^* and B* are modelled with Breit-Wigners, as opposed to the smooth mass spectra of the X diffractive states. In our description, this has only been implemented in nucleon-nucleon interactions.
- Non-diffractive topologies are assumed to correspond to a net colour exchange between the incoming hadrons, such that colour strings are stretched out between them after the interaction.

Some examples of input used for the modelling of these total and partial cross sections are as follows.

- Cross sections are invariant when all particles are replaced by their antiparticles.
- In some cases good enough data exists that interpolation works.
- $\pi\pi$ and Kπ cross sections are found using the calculations of Peláez et al. [47–49], which partly are based on Chiral Perturbation Theory.
- The neutral Kaon system is nontrivial, with strong interactions described by the $K^0/\overline{K}^0$ states and weak decays by the K^0_S/K^0_L ones. Cross sections for a K^0_S/K^0_L with a hadron are given by the mean of the cross section for K^0 and $\overline{K}^0$ with that hadron. When a collision occurs, the $K_{S,L}$ is converted into either K^0 or $\overline{K}^0$, where the prob-

ability for each is proportional to the total cross section for the interaction with that particle.

- Several total cross sections are described by the HPR_1R_2 parameterization [50], consisting of one fixed term, one "pomeron" $\ln^2 s$ ($s = E_{CM}^2$) and two "reggeon" $s^{-\eta}$ ones.
- NN and Nπ elastic cross sections are partly covered by the CERN/HERA data parameterizations [51].
- The UrQMD program [27] has a complete set of total and partial cross sections for all light hadrons, and in several cases we make use of these expressions.
- Intermediate resonance formation can be modelled in terms of (non-relativistic) Breit-Wigners, given a knowledge of mass and (partial) width of the resonance. The widths are made mass-dependent using the ansatz in UrQMD.
- The annihilation cross section is the difference between the total and the elastic ones near threshold, and above the inelastic threshold it is based on a simple parameterization by Koch and Dover [52].
- Differential diffractive cross sections are described by the SaS (Schuler and Sjöstrand) ansatz [37,53], and their integrated cross sections are parameterized with special attention to achieving the relevant threshold behaviour.
- Excitation into explicit higher resonances is implemented for NN collisions, using the UrQMD expressions. For other collision types the low-mass diffraction terms of SaS are included instead.
- Inelastic non-diffractive events are represented by the cross section part that remains when everything else is removed. Typically it starts small near the threshold, but then grows to dominate at higher energies.
- The additive quark model (AQM) [54,55] assumes that total cross sections scales like the product of the number of valence quarks in the two incoming hadrons. The contribution of heavier quarks is scaled down relative to that of a u or d quark, presumably by mass effects giving a narrower wave function. Assuming that quarks contribute inversely proportionally to their constituent masses, this gives an effective number of interacting quarks in a hadron of approximately

$$n_{q,AQM} = n_u + n_d + 0.6\,n_s + 0.2\,n_c + 0.07\,n_b. \quad (3)$$

For lack of alternatives, many unmeasured cross sections are assumed to scale in proportion to this, relative to known ones. For heavier particles, notably charm and bottom ones, it is also necessary to correct the collision energy relative to the relevant mass threshold.

2.6 Hadronic rescattering events

The choice of subprocess is not enough to specify the resulting final state. In some cases only a few further variable choices are needed. For elastic scattering the selection of the Mandelstam t is sufficient, along with an isotropic φ variable. Resonances are assumed to decay isotropically, as are the low-mass excitations related to diffraction. For inelastic non-diffractive events, higher-mass diffractive ones, and annihilation processes, generically one one would expect strings to form and hadronize. For diffraction these strings would be stretched inside a diffractively excited hadron, while for the other two cases the strings would connect the two original hadrons.

To illustrate the necessary steps, consider an inelastic non-diffractive event. Each of the incoming hadrons first has to be split into a colour piece, q or $\overline{qq}$, and an anticolour ditto, $\overline{q}$ or qq. For a baryon, $SU(6)$ flavour×spin factors are used to pick the diquark spin. Then the lightcone momentum $p^+(p^-)$ is split between the two pieces of incoming hadron $A(B)$ moving along the $+z(-z)$ direction, in such a way that a diquark is likely to carry the major fraction. The pieces also are given a relative $p_\perp$ kick. Including (di)quark masses, the transverse masses $m_{\perp A1}$ and $m_{\perp A2}$ of the two A hadron pieces are defined. The p_{Ai}^- can now be obtained from $p^+p^- = m_\perp^2$, and combined to give an effective mass m_A^*, and similarly an m_B^* is calculated. Together, the criterion $m_A^* + m_B^* < E_{CM}$ must be fulfilled, or the whole selection procedure has to be restarted. Once an acceptable pair (m_A^*, m_B^*) has been found, it is straightforward first to construct the kinematics of A^* and B^* in the collision rest frame, and thereafter the kinematics of their two constituents.

Since the procedure has to work at very small energies, some additional aspects should be mentioned. At energies very near the threshold, the phase space for particle production is limited. If the lightest hadrons that can be formed out of each of the two new singlets together leave less than a pion mass margin up to the collision CM energy, then a simple two-body production of those two lightest hadrons is (most likely) the only option and is thus performed. There is then a risk to end up with an unintentional elastic-style scattering. For excesses up to two pion masses, instead an isotropic three-body decay is attempted, where one of the strings breaks up by the production of an intermediate u$\overline{u}$ or d$\overline{d}$ pair. If that does not work, then two hadrons are picked as in the two-body case and a π^0 is added as third particle.

Even when the full collision energy is well above threshold, either one or both of the strings individually may have a small mass, such that only one or at most two hadrons can be produced from it. It is for cases like this that the ministring framework has been developed, where it is allowed for a string to collapse into a single hadron, with liberated excess momentum shuffled to the other string. In a primary high-energy collisions, low-mass strings are rare, and typically surrounded by higher-mass ones that easily can absorb the recoil. At lower energies it is important to try harder to find working solutions, and several steps of different kinds

have been added to the sequence of tries made. The new setup still can fail occasionally to find an acceptable final state, but far less than before the new measures were introduced.

3 Model tests

In this section we will study the rescattering model in pp, pPb and PbPb collisions. All collision energies are set to 5.02 TeV per nucleon-nucleon system. This includes pp, for comparison reasons; results at the more standard 13 TeV pp energy have already been presented elsewhere [34].

3.1 Multiplicities

The current lack of $3 \rightarrow 2$ processes in our model, to partly balance the $2 \rightarrow 3$ ones, means that rescattering will increase the charged hadron multiplicity. Effects are modest for pp but, to compensate, the $p_{\perp 0}$ parameter of the MPI framework is increased slightly when rescattering is included. Thus the number of MPIs is reduced slightly, such that the pp charged multiplicity distribution is restored to be in reasonable agreement with experimental data. We have used the same value for this parameter also for the pPb and PbPb rescattering cases. Then rescattering increases the final charged multiplicities by about 4% and 20%, respectively, due to a larger relative amount of rescattering in larger systems. Implementing $3 \rightarrow 2$ processes (by checking whether further hadrons are in the neighbourhood of a candidate $2 \rightarrow 2$ process) would partly mitigate this effect, but even then, we still expect rescattering to increase charged multiplicity due to entropy production. Thus, to simultaneously restore the multiplicity for all cases, a retune also of ANGANTYR parameters would be necessary. This is beyond the scope of the current article, and for now we accept some mismatch.

Charged multiplicity distributions are shown in Fig. 1a, split into hadrons that have or have not been affected by rescattering. Particles with a proper lifetime $\tau_0 > 100$ fm have been considered stable, and multiplicities are reported without any cuts on η or $p_{\perp}$. Moving from pp to pPb to PbPb we see how the fraction of particles that do not rescatter drops dramatically. In absolute numbers there still are about as many unrescattered in pPb as in pp, and about twice as many in PbPb. A likely reason is that many collisions are peripheral, and even when not there are particles produced at the periphery.

The total charged multiplicity is also compared with and without rescattering. As foretold, the pp case has there been tuned to show no difference, whereas rescattering enhances the high-multiplicity tail in pPb and PbPb. Rescattering also changes the relative abundances of different particle types. In particular, baryon-antibaryon annihilation depletes the baryon rate, by 7.5% for pp, 9.9% for pPb and 23.4% for

PbPb, compared to the baryon number with a retuned $p_{\perp 0}$. The retuning itself gives in all cases a $\sim 2\%$ reduction, that should be kept separate in the physics discussion. The observed strange-baryon enhancement [14,56] thus has to be explained by other mechanisms, such as the rope model [26] or other approaches that give an increased string tension [57].

3.2 Rescattering rates

One of the most basic quantities of interest is the number of rescatterings in an event. The average number of rescatterings as a function of the final charged multiplicity n_{ch} is shown in Fig. 1b. The number of potential interactions at the beginning of rescattering is proportional to $n^2_{primary}$, where the number of primary hadrons $n_{primary} \simeq n_{ch}$. The scaling is different in practice however, due to the fact that some particles rescatter several times, while others do not rescatter at all. As a first approximation one might still expect the number of rescatterings to increase as n_{ch}^p for some power p. As seen in Fig. 1b, this relation appears to hold remarkably well, with $p = 1.37$ for pp, $p = 1.47$ for pPb, and $p = 1.43$ for PbPb. Interestingly, the exponent is highest for the intermediate case pPb, but the rescattering activity as such is still highest for PbPb. A possible explanation could be that in PbPb, high multiplicity corresponds to more central events with a larger volume, and thus higher multiplicity does not necessarily mean higher density in this case. We have also studied other pA and AA cases for a wide variety of sizes of A, including Li, O, Cu and Xe. While there is some A dependence in the exponent, this variation is less significant than the overall difference between the pA and AA cases, and in all instances the respective p numbers for pPb and PbPb provide a reasonable description.

The invariant mass distributions of rescatterings are shown in Fig. 2a by incoming particle kind and in Fig. 2b by rescattering type. For increasingly large systems the fraction of low-mass rescatterings goes up. A likely reason for this is rescattering causes a greater multiplicity increase in the larger systems, reducing the average energy of each particle. The composition of collision types at a given mass is the same (within errors), as could be expected. Our rescattering model is based on a non-perturbative framework intended to be reasonably accurate up to around ~ 10 GeV. It would have to be supplemented by perturbative modelling if a significant fraction of the collisions were well above 10 GeV, but clearly that is not the case. As an aside, the bump around 5.5 GeV comes from interactions involving bottom hadrons.

3.3 Transverse momentum spectra

The $p_{\perp}$ spectra for pions, kaons, nucleons and charm mesons, with and without rescattering, are shown in Fig. 3a, b, and the ratios with/without are shown in Fig. 3c ,d. The effects

(a)

(b)
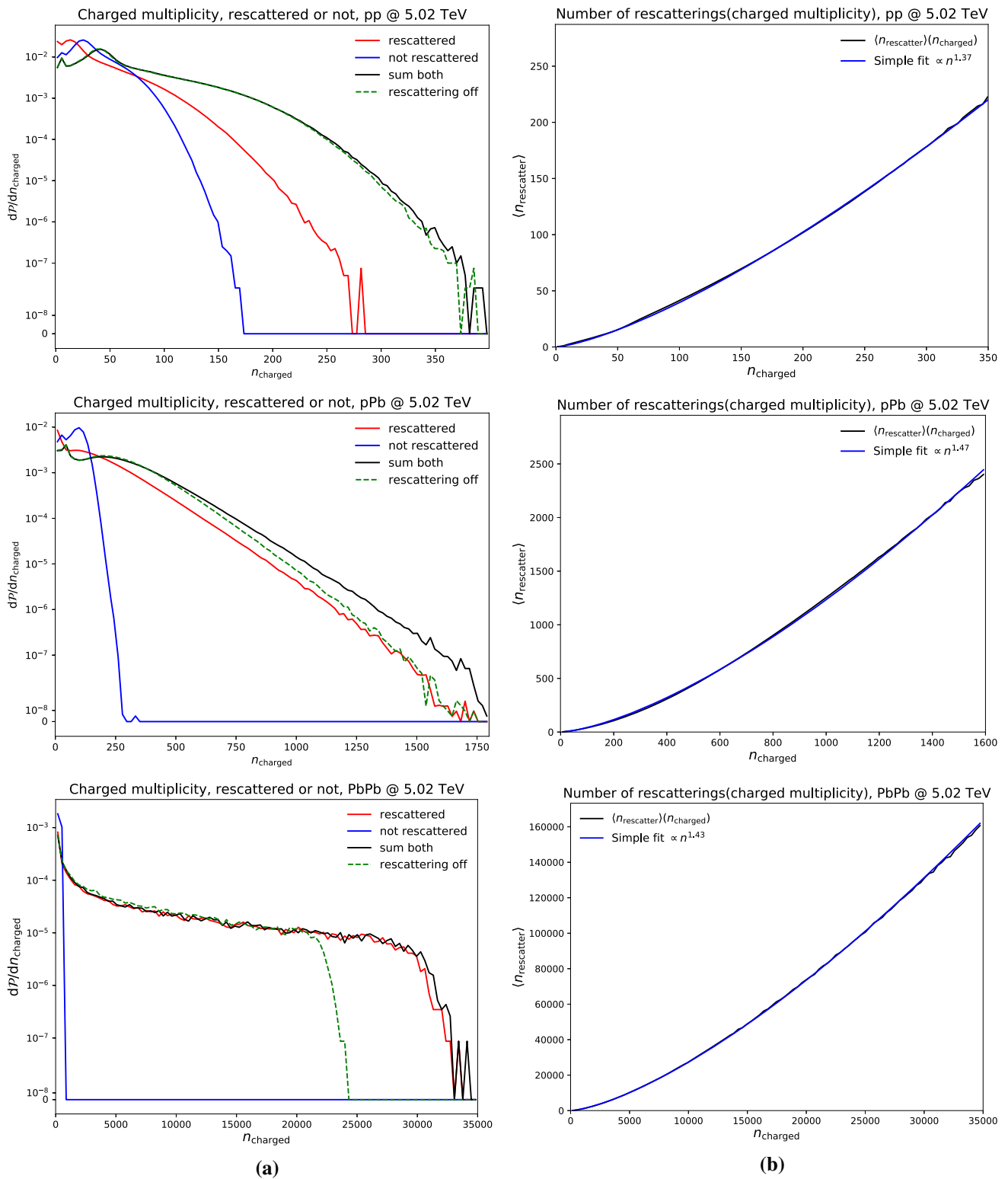
Fig. 1 **a** Probability distributions for the total number of charged hadrons, with and without rescattering, as well as the former number split in those where the final charged hadrons have been affected (directly or indirectly) by rescattering and those where not. **b** Average number of rescatterings as a function of the charged hadron multiplicity, together with a simple fit proportional to n_{ch}^{p}
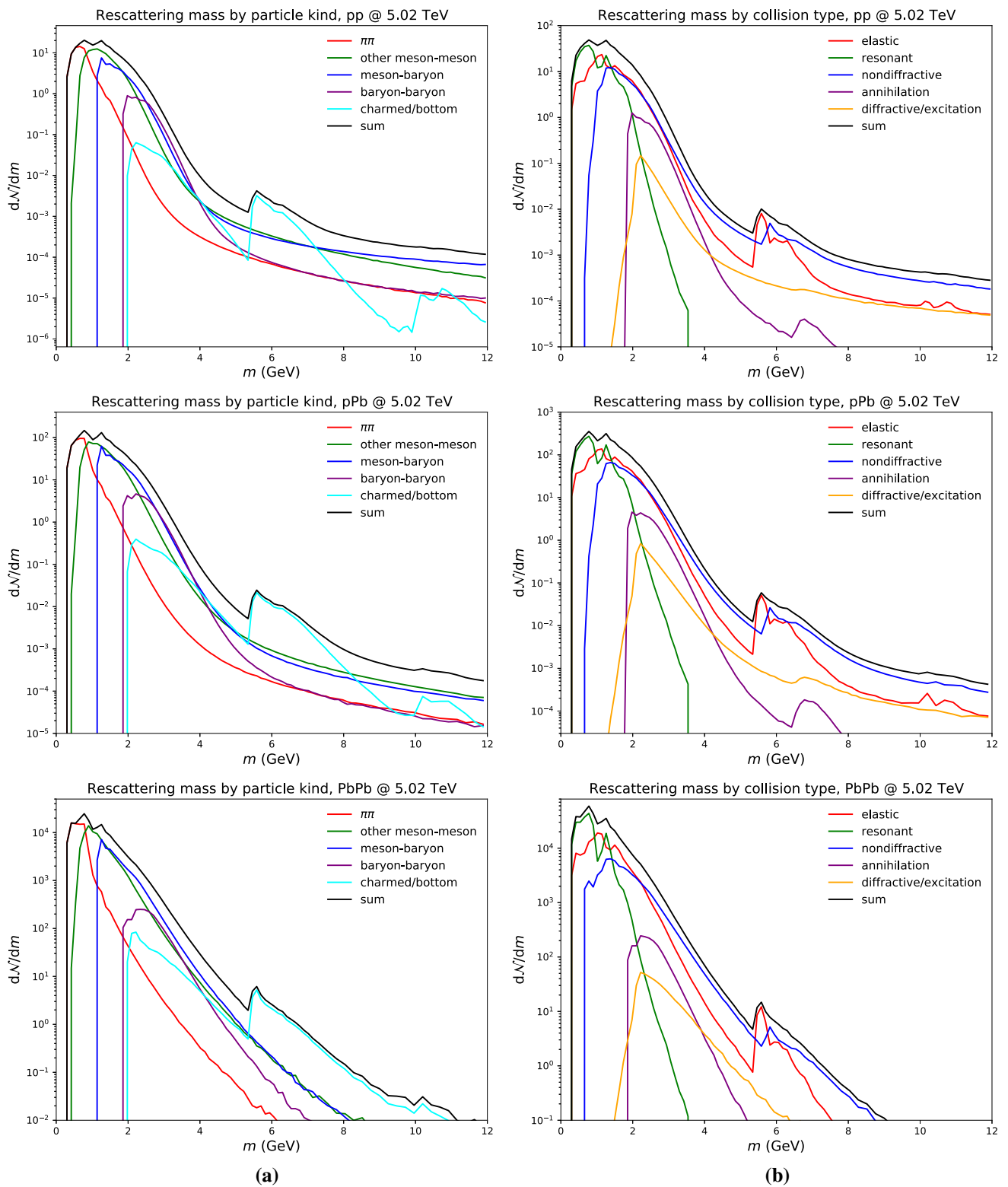
Fig. 2 Invariant masses for rescatterings, **a** by particle kind and **b** by rescattering process type
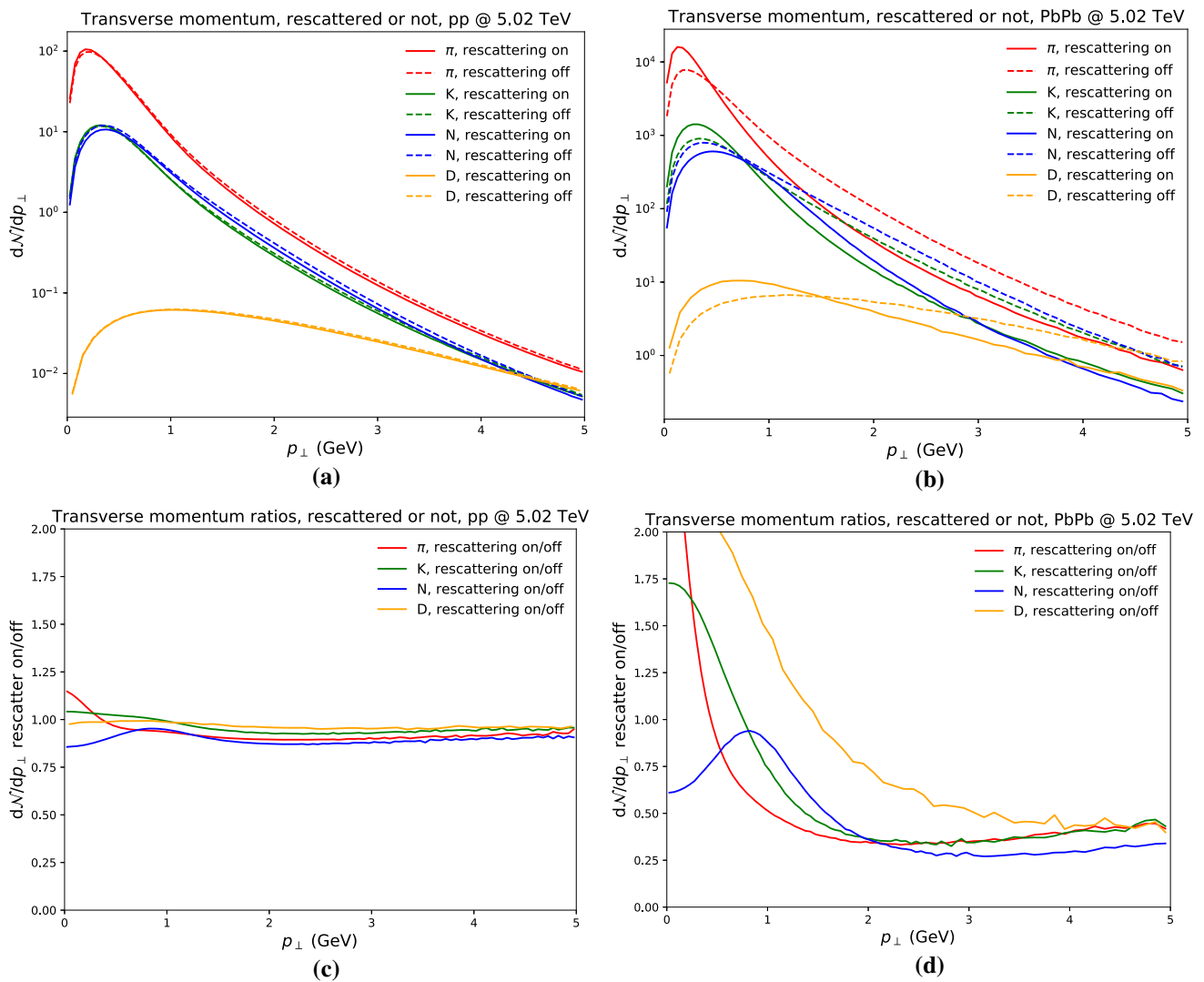
Fig. 3 $p_\perp$ spectra for pions, kaons, nucleons and D mesons, for **a** pp and **b** PbPb, together with ratios between the spectra with to without rescattering, for **c** pp and **d** PbPb

are qualitatively similar for pp and PbPb, but more prominent for the latter case. Pions get pushed to lower $p_\perp$, which is consistent with the expectation that lighter particles will lose momentum due to the "pion wind" phenomenon, where lighter particles move faster than heavier and push the latter ones from behind. We remind that all primary hadrons types are produced with the same $p_\perp$ distribution in string fragmentation, if the string is stretched parallel with the collision axis. Rapid ρ and K* decays decrease the average pion $p_\perp$, but initially indeed pions have the largest velocities.

The effect is similar for kaons, which unfortunately is inconsistent with measurement [56]. Our studies indicate that a significant contribution to the loss of $\langle p_\perp \rangle$ for kaons comes from inelastic interactions, and that the $\langle p_\perp \rangle$ increases if all rescatterings are forced to be elastic. We believe this effect can be somewhat ameliorated by implementing $3 \rightarrow 2$ and

related processes. For nucleons we note an overall loss in the rescattering scenario, which comes mainly from baryon–antibaryon annihilation, as already mentioned. The $\langle p_\perp \rangle$ is shifted upwards by the aforementioned pion wind phenomenon.

D mesons are enhanced at low $p_\perp$, all the way down to threshold. At first glance this appears inconsistent with the pion wind phenomenon, since D mesons are heavy. One key difference is that charm quarks are not produced in string fragmentation, but only in perturbative processes. Therefore D mesons start out at higher $p_\perp$ values than ordinary hadrons, and can lose momentum through rescattering. Nevertheless, the overall shift is still somewhat towards higher momenta if only elastic rescatterings are permitted, as for kaons.

Overall we see a rather significant effect on $p_\perp$ spectra, and this is to be kept in mind for other distributions. Espe-

cially for pions, where the choice of a lower $p_\perp$ cut in experimental studies strongly affects the (pseudo)rapidity spectrum deformation by rescattering, among others.

3.4 Spacetime picture of rescattering

In this section we study the spacetime distributions of rescatterings. Specifically, we consider the transverse production distance, $r_\perp^2 = x^2 + y^2$, and longitudinal invariant time, $\tau_L^2 = t^2 - z^2$. The two Lorentz-contracted "pancake" nuclei are set to collide at $t = z = 0$, with the center of collision at $x = y = 0$, but with sub-collisions spread all over the (x, y) overlap region. Thus the squared invariant time $\tau^2 = t^2 - x^2 - y^2 - z^2 = \tau_L^2 - r_\perp^2$ tends to have a large tail out to large negative values, so it is not a suitable measure for heavy-ion collisions. The $r_\perp$ and τ_L distributions are shown in Fig. 4, separately for particles involved or not in rescattering. For the latter it is the location of the last rescattering that counts. Particle decays are included for particles with proper lifetimes $\tau_0 < 100$ fm, so that a "final" pion could be bookkept at the decay vertex of for instance a ρ.

The overall observation is that rescattering reduces particle production at very early and at late times, as is especially clear in the τ_L distribution for PbPb. Particles produced at early times are more likely to participate in rescattering and get assigned new τ_L values on the way out. With this in mind, it may seem paradoxical that the $r_\perp$ distributions are comparably broad for rescattered and unrescattered particles. Hadrons produced in the periphery of the collision are more likely to evade rescattering than central ones, however, so this introduces a compensating bias towards larger $r_\perp$ for the unrescattered. In this respect the τ_L distribution more follows the expected pattern, with the unrescattered particles having comparable average values in all three collision scenarios, whereas the rescattered ones are shifted further out. Maybe somewhat unexpectedly, particle production at late times and large $r_\perp$ is also reduced with rescattering on. Our studies indicate that there is some rescattering activity at late times ($\gtrsim 50$ fm), but the number of rescatterings here is roughly a factor of three smaller than the number of decays. Now, since rescattering produces more particles early, it tends to reduce the average particle mass, which increases the number of stable particles produced early and reduces the number of decaying ones in the 50–100 fm range. Furthermore, unstable particles often have lower $p_\perp$ and hence smaller Lorentz factors, leading them to decay at lower $r_\perp$ values.

While the exact time of a rescattering cannot be measured directly, phenomena such as resonance suppression can give an indication of the duration of the hadronic phase [58,59]. Experimentally, a suppression of the K*/K yield ratio at higher multiplicities has been observed, but not of the ϕ/K yield ratio. The interpretation of this observation is as follows: after the K* decays, the outgoing π and K are likely to participate in rescattering because of their large cross sections, which disturbs their correlation and suppresses the original K* signal. The fact that the ϕ signal is not suppressed in this way indicates that they tend to decay only after most rescattering has taken place. With the K* and ϕ lifetimes being 3.9 fm and 46.3 fm respectively, this places bounds on the duration of the rescattering phase. These bounds seem to be consistent with the spacetime distributions shown in Fig. 4.

With the full event history provided by PYTHIA, it is possible to study the actual number of K* and ϕ that were produced, and to trace what happens to their decay products. A naïve way to approach resonance suppression is to define a K* or ϕ meson as detectable if it decayed and no decay product participated in rescattering. When defining the K* multiplicity in this way, we found that rescattering actually increases the K*/K ratio for larger charged multiplicities. This increase is not observable, however, since it mainly comes from $K\pi \to K^* \to K\pi$. That is, some of the combinatorial background gets to be reclassified as K*, without any change of the overall $K\pi$ mass spectrum. To find the more subtle effects of nontrivial processes requires a detailed fitting of the $K\pi$ mass spectrum. This is outside the scope of this article, but would be interesting to study in the future. Nevertheless, the change in the ϕ/K ratio is much smaller, suggesting that qualitatively, longer-lived resonances are indeed less affected by rescattering.

3.5 Centrality dependent observables

In heavy ion experiments, observables are most often characterized according to collision centrality. The characterization is a sensible one, also for checking the effects of hadronic rescatterings, as this will be the largest in the most central collisions. While experiments employ a centrality definition depending on particle production in the forward or central regions of the experiments, we will in the following sections use the definition adhering to impact parameter. As such, the centrality of a single collision is defined as

$$c = \frac{1}{\sigma_{\text{inel}}} \int_0^b \mathrm{d}b' \frac{\mathrm{d}\sigma_{\text{inel}}}{\mathrm{d}b'}. \tag{4}$$

We note, however, that the results presented for nucleus-nucleus collisions can be transferred directly to experimental centrality measures, as the ANGANTYR model provides a good description of e.g. forward energy, which correlates directly with the theoretical impact parameter.

3.5.1 Particle yields and ratios

In the following we present the effect on identified particle yields in $|y| < 4$ (to avoid the beam region) in XeXe
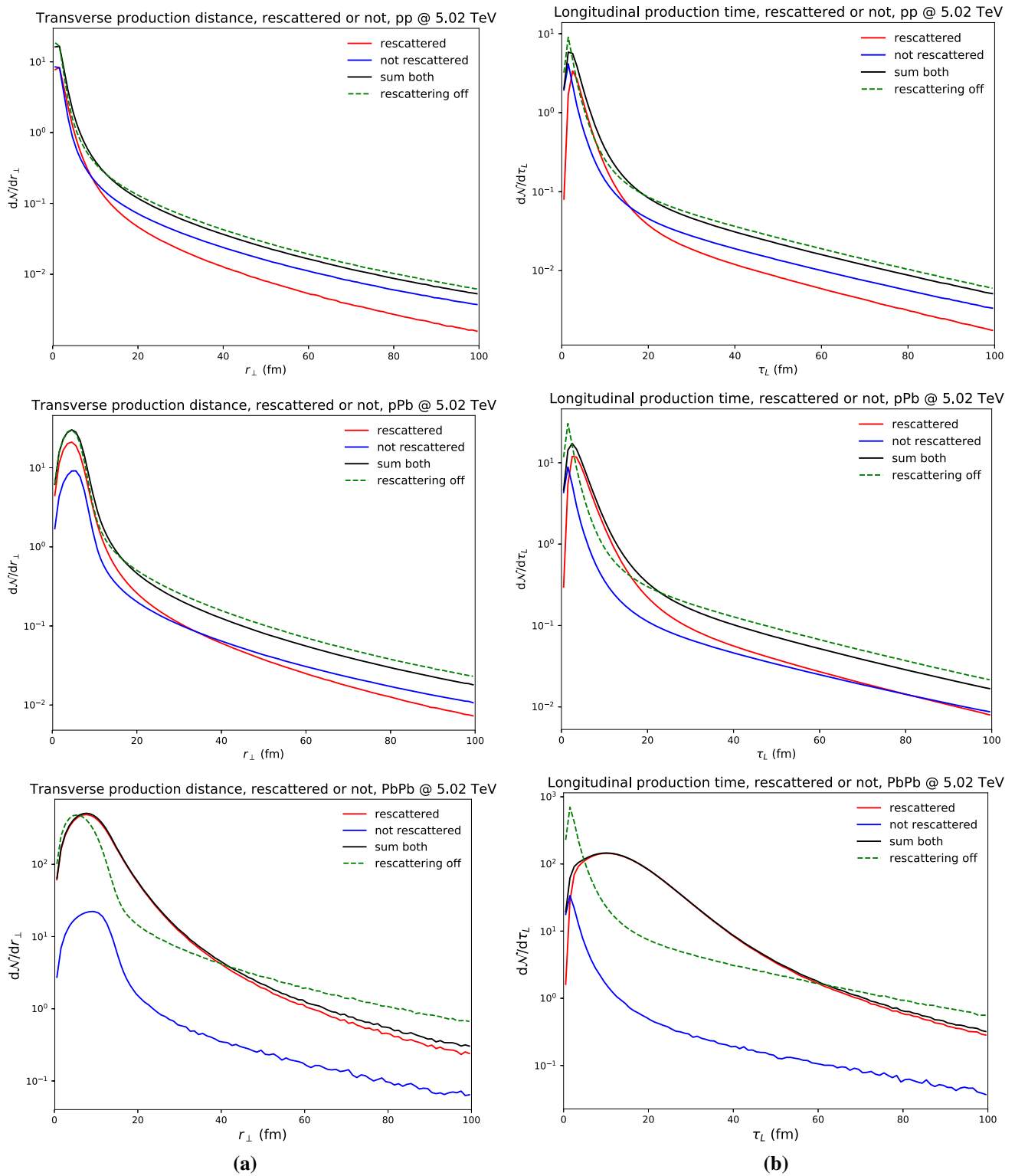
Fig. 4 **a** $r_\perp$ and **b** τ_L spectra. Note that rescattered also refers to hadrons produced in decays of rescattered particles, even though they themselves were not directly involved in rescattering
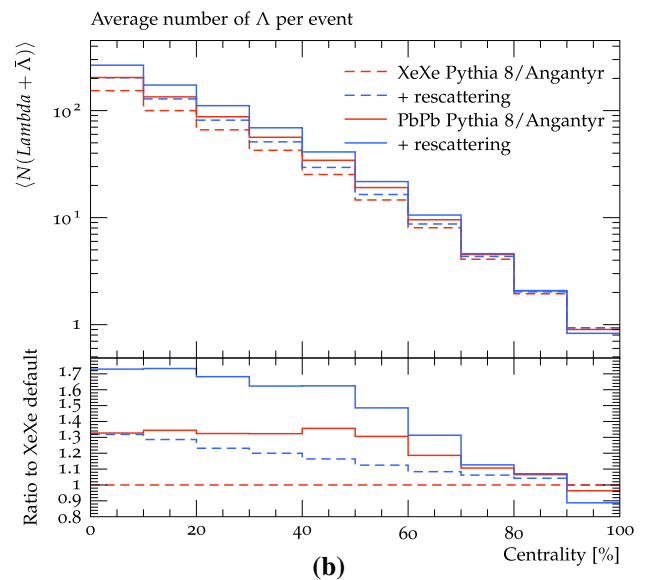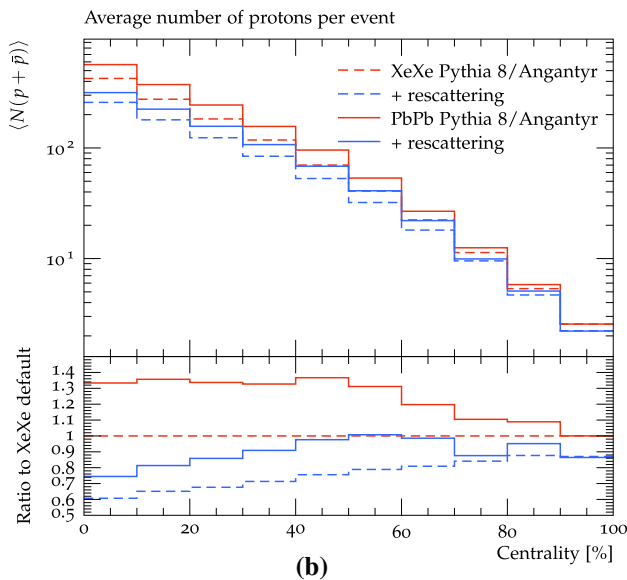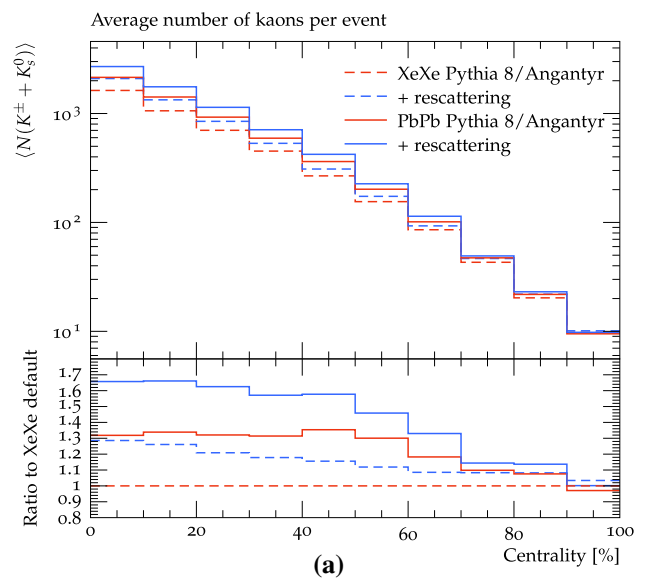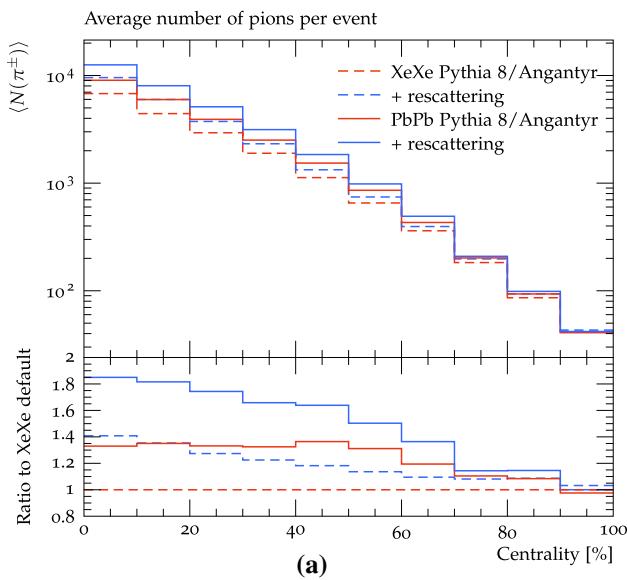
Fig. 5 Average per-event yields of **a** pions ($\pi^{\pm}$) and **b** protons (p, $\bar{\text{p}}$) in PbPb and XeXe collisions at $\sqrt{s_{NN}} = 2.74$ and 5.44 TeV respectively, as function of collision centrality
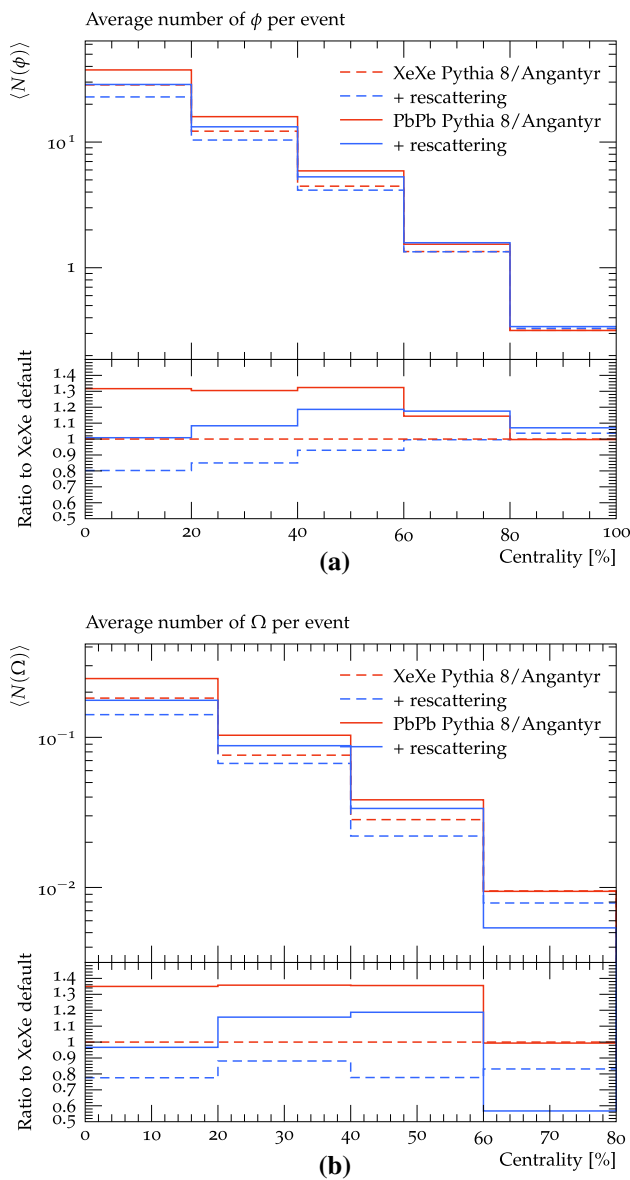
Fig. 6 Average per-event yields of **a** kaons ($K^{\pm}$, $K^0_{L,S}$) and **b** Λ (Λ, $\overline{\Lambda}$) in PbPb and XeXe collisions at $\sqrt{s_{NN}} = 2.74$ and 5.44 TeV respectively, as function of collision centrality

collisions at $\sqrt{s_{NN}} = 5.44$ TeV and PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Starting with light flavour mesons and baryons, we show the average multiplicity of (a) pions ($\pi^{\pm}$) and (b) protons (p, $\bar{\text{p}}$) per event in Fig. 5.

While the effect for pions is negligible in peripheral collisions, it grows to about 40% in central collisions. The effect on protons is also largest in central collisions, while in peripheral collisions it is still at a 10% level. This is particularly interesting in the context of recent years' introduction of microscopic models to explain the increase of strange baryon yields with increasing multiplicity, which overestimate the amount of protons [60].

In Fig. 6 we move to strange mesons and baryons, with the total kaon ($K^{\pm}$ and $K^0_{L,S}$) and Λ multiplicity, (a) and (b) respectively. While there is a large effect on the direct yields of both species, it is almost identical to the change in $\pi^{\pm}$ in Fig. 5a, leaving the K/π and Λ/π ratios unchanged.

We finish the investigation of the light-flavour sector by showing the total ϕ and Ω^- multiplicities in Fig. 7a, b respectively. The ϕ multiplicity decreases by about 20% in central events and is constant within the statistical errors in peripheral. The Ω multiplicity is decreased roughly the same amount. The decrease here, however, is rather constant in centrality in XeXe but increases for central events in PbPb.

Fig. 7 Average per-event yields of **a** ϕ and **b** Ω^- in PbPb and XeXe collisions at $\sqrt{s_{NN}} = 2.74$ and 5.44 TeV respectively, as function of collision centrality

Fig. 8 Average per-event yields of **a** J/ψ and **b** D mesons in PbPb and XeXe collisions at $\sqrt{s_{NN}} = 2.74$ and 5.44 TeV respectively, as function of collision centrality

The rescattering framework implemented in PYTHIA, includes cross sections for heavy flavour mesons and baryons. In Fig. 8 we show the effect on (a) J/ψ and (b) D mesons (D$^\pm$ and D^0). Starting with the J/ψ we see a significant effect in both collision systems in central events, less so in peripheral. While the initial J/ψ yield is roughly 10% larger in PbPb than in XeXe, the final value after rescattering saturates at a value at roughly 60% of the initial XeXe value, independent of the two collision systems.[1] Whether or not this is consistent with the measured nuclear modification factor [61] in

peripheral collisions (clearly not in central collisions, where an additional source of J/ψ production would be required) is left for future detailed comparisons to data.

In PYTHIA (rescattering or not) there is no mechanism for charm quarks to vanish from the event at early times. The constituents of the J/ψ would therefore have to end up in other charmed hadrons. In Fig. 8b we show the D meson yield, demonstrating that this is more than two orders of magnitude above the J/ψ one. It is then consistent to assume that the missing charm quarks can recombine into open charm without having observable consequences. Indeed there is no significant effect on the D meson yield from rescattering.

---

[1] This feature is clearly accidental. We have checked in smaller collision systems to confirm.

We note that while the most widely used flavours of UrQMD does not include cross sections for charmed hadrons, a modified version exists [62], implementing a scenario of J/ψ melting at high temperatures (thus not comparable to this study where no QGP is assumed), a fixed charmonium dissociation cross sections in the pre-hadronic regime at lower temperatures (also not completely comparable to this study, as no interactions between fully formed charmonia and string pieces are considered) and finally elastic scattering and dissociation of J/ψ, and recombination of D-mesons to charmonium states in the hadronic regime. The UrQMD study presented a suppression of J/ψ yields in high multiplicity pp collisions at LHC energies down to 70%. While we, in this present study, have not considered J/ψ production in pp collisions, we find it unlikely that a pure hadronic rescattering model, as implemented here, will give results of that magnitude in pp.

In the context of the PHSD model [63], effects from final state rescatterings of charm quarks have been studied in more detail [64,65]. Since this approach works both on the parton and hadron level (reducing the overall $c\bar{c}$ cross section), the results here are also not completely comparable to our study. We do note, however, that the results of the cited studies indicates a good agreement with data.

3.5.2 Elliptic flow

One of the most common ways to characterize heavy ion collisions is by the measurement of flow coefficients (v_n's), defined as the coefficients of a Fourier expansion of the single particle azimuthal yield, with respect to the event plane Ψ_n [66,67]:

$$E \frac{\mathrm{d}^3 N}{\mathrm{d}^3 p} = \frac{1}{2\pi} \frac{\mathrm{d}^2 N}{p_\perp \mathrm{d} p_\perp \mathrm{d} y} \left(1 + 2 \sum_{n=1}^{\infty} v_n \cos(n(\varphi - \Psi_n)) \right). \tag{5}$$

The azimuthal angle is denoted φ, and E, $p_\perp$ and y are the particle energy, transverse momentum and rapidity respectively. In experiments it is not possible to utilize this definition directly, as the event plane is unknown. Therefore one must resort to other methods. For the purpose of testing if a model behaves as expected, it is on the other hand preferable to measure how much or little particles will correlate with the true event plane (when we show comparisons to experimentally obtained values in Sect. 4.2, we will use the experimental definitions). In the following, we will therefore use an event plane obtained from the initial state model, defined as

$$\Psi_n = \frac{1}{n} \arctan\left(\frac{\langle r^2 \sin(n\varphi) \rangle}{\langle r^2 \cos(n\varphi) \rangle} \right) + \frac{\pi}{n}, \tag{6}$$

for all initial state nucleons participating in collisions contributing to the final state multiplicity (inelastic, non-diffractive sub-collisions). The origin is shifted to the center of the sampled distribution of nucleons, and r and φ are the usual polar coordinates. Flow coefficients can then simply be calculated as

$$v_n = \langle \cos(n(\varphi - \Psi_n)) \rangle. \tag{7}$$

As in the previous section we consider all particles in $|y| < 4$ and without any lower cut on transverse momentum.

In Fig. 9 we show (a) v_2 and (b) v_3 as functions of collision centrality for charged particles for XeXe collisions at $\sqrt{s_{NN}} = 5.44$ TeV and PbPb collisions at $\sqrt{s_{NN}} = 2.74$ TeV both with and without rescattering. It is seen that v_2 receives a sizeable contribution from rescattering. The contribution is larger for PbPb than for XeXe, which is not surprising, given the larger density. The v_2 arises because particles are pushed by rescatterings along the density gradient, which is larger along the event plane. Note that the curve without rescattering is zero, as the definition of v_n from (7) ensures that no non-flow contributions enter the results.

For v_3 (Fig. 9b) there is not much difference between PbPb and XeXe. Since v_3 is mainly generated by initial state shape fluctuations, this is a reasonable result.

Since different hadron species have different cross sections, hadronic rescattering will yield different flow coefficients for different hadron species. As an example, since the pp($\bar{p}p$) cross section is larger than the average hadron-hadron cross section (which is dominated mainly by pions), v_2 for protons will be higher. We note (without showing) that hadronic rescattering gives $v_2(p) > v_2(\pi) > v_2(K) \approx v_2(\Lambda) > v_2(\Omega) > v_2(\phi)$, with the latter reaching its maximum for v_2 about an order of magnitude less than for protons. While a dedicated analysis would likely reveal that the PYTHIA rescattering framework would thus reproduce constituent quark scaling of the elliptic flow [68,69], this has already been shown to be the case for other hadronic rescattering models [70], and little would be learned by unfolding the full exercise once more.

For heavy flavours, the results require more explanation, due to the differing production mechanisms. In PYTHIA, D mesons are produced in string fragmentation, requiring that one of the quark ends is a charm quark. The J/ψ, on the other hand, is predominantly produced early, either by direct onium production via colour-singlet and colour-octet mechanisms, or by an early "collapse" of a small $c\bar{c}$ string to a J/ψ. Onia are therefore excellent candidates for hadrons mainly affected by hadronic rescattering, and not any effects of strings interacting with each other before hadronization.

In Fig. 10, we show v_2 for (a) D mesons and (b) J/ψ. Starting with D mesons we see an appreciable v_2, numerically not too far from PbPb data [71]. A clear difference is
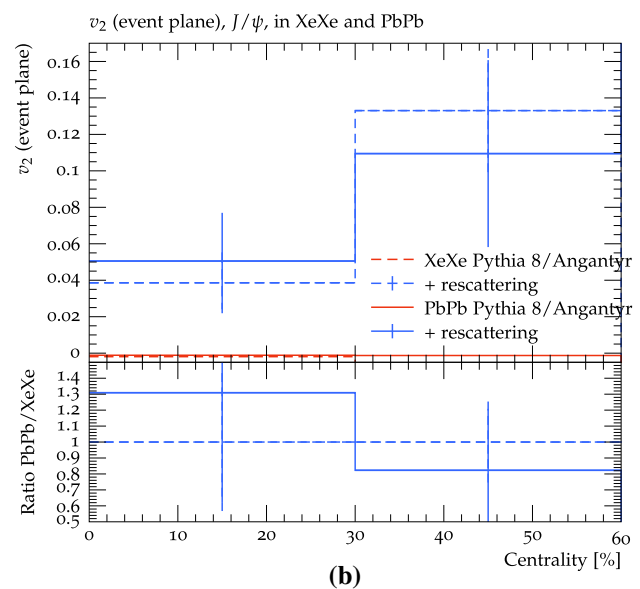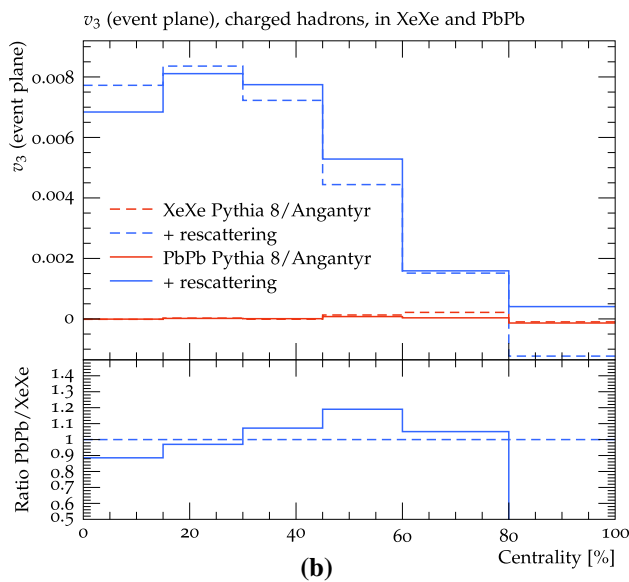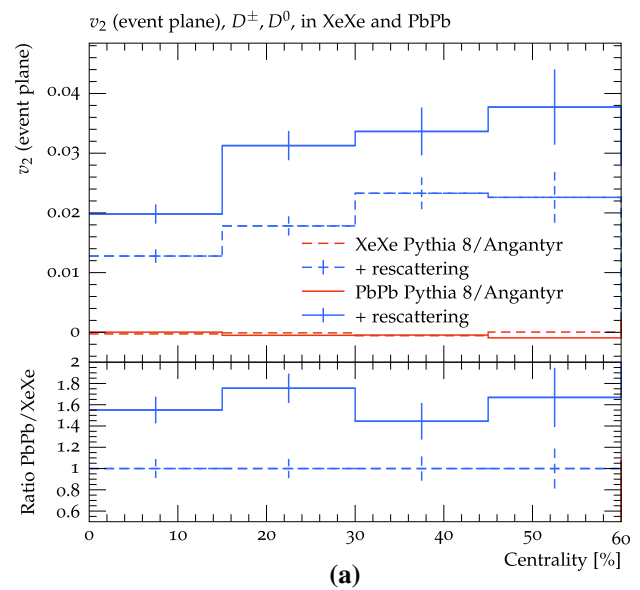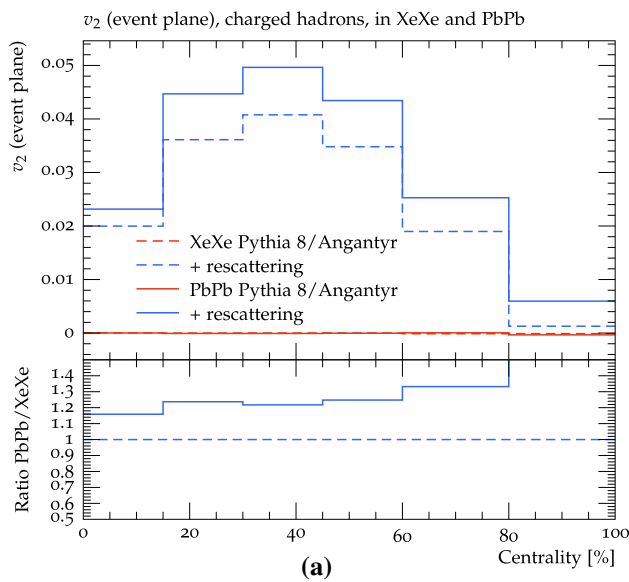
Fig. 9 Flow coefficients **a** v_2 and **b** v_3 in PbPb and XeXe collisions at $\sqrt{s_{NN}} = 2.74$ and 5.44 TeV respectively. Results shown with and without rescattering, and are calculated with respect to the event plane such that the sample without rescattering is zero by construction

Fig. 10 The v_2 flow coefficient for **a** D mesons ($D^{\pm}$, D^0) and **b** J/ψ as a function of centrality in PbPb and XeXe collisions at $\sqrt{s_{NN}} = 2.74$ and 5.44 TeV respectively. Error bars are statistical errors. Results shown with and without rescattering, and are calculated with respect to the event plane such that the sample without rescattering is zero by construction

observed between XeXe and PbPb. In the figure, statistical error bars are shown, as they are not negligible due long processing times for heavy flavour hadrons. For the J/ψ, shown in Fig. 10b, v_2 for PbPb and XeXe are compatible within the statistical error. More importantly, the result is also compatible with experimental data [72]. Together with the result from Fig. 8a, which suggests a sizeable nuclear modification to the J/ψ yield from rescattering, a detailed comparison with available experimental data should be performed. It should be noted that the treatment of charm in the PYTHIA hadronic rescattering model follows the additive quark model, as introduced earlier. Thus, no distinction is made between J/ψ and

other c$\bar{c}$ states. A foreseen improvement of this treatment would be to consider differences with input taken e.g. from lattice calculations. We note that the combined partonic and hadronic rescattering approach of PHSD, falls just short of reproducing elliptic flow of D-mesons [64] at RHIC energies. The initial state of the PYTHIA hadronic rescattering framework is, however, different, and a dedicated study could therefore still be worthwhile.

4 Comparison with data

In this section we go beyond the model performance plots shown in the previous section, and compare to relevant experimental data for XeXe and PbPb, in cases where Rivet [73] implementations of the experimental analysis procedure are available (though not in all cases validated by experiments). We focus on observables where the rescattering effects are large, and in some cases surprising.

In all cases centrality is defined according to (4), as it reduces computation time, and the difference between centrality defined by impact parameter and forward energy flow is not large in AA collisions.

4.1 Charged multiplicity

In Sect. 3.1 we described how rescattering increases the total multiplicities, owing especially to the current lack of $3 \rightarrow 2$ processses. In Fig. 11 ANGANTYR with and without rescattering is compared to experimental data [74,75].

In Fig. 11a, $dN_{ch}/d\eta|_{\eta=0}$ is shown as function of centrality. It is clear that the shift in multiplicity, caused by rescattering, is centrality dependent, with a larger effect seen in more central events. For centrality 0–5%, the agreement with data shifts from approximately 8% below data to 10% above. It is instructive to show the differential distributions as well, and in Fig. 11b, the η-distribution out to ± 5 is shown. It is seen that the shift is slightly larger at the edges of the plateau. This effect is most pronounced in the centrality bin shown here, and decreases for more peripheral events.

To further explore the change in charged multiplicity distributions, we show comparisons to invariant $p_\perp$ distributions in the same collision system, measured down to $p_\perp = 0.15$ GeV in $|\eta| < 0.8$ [76] in Fig. 12, with 0–20% centrality shown in Fig. 12a, and 40–60% in Fig. 12b. It is seen that particles at intermediate $p_\perp \approx 1$–6 GeV are pushed down to very low $p_\perp$, as expected since rescattering will generate more final state particles overall.

From this investigation of effects on basic single-particle observables from adding rescattering, it is clear that agreement with data is decreased. Since hadronic rescattering in heavy ion collisions is a physical effect that must be taken into account, this clearly points to the need of further model improvement.

4.2 Flow coefficients

As indicated in Sect. 3.5, rescattering has a non-trivial effect on flow observables, a staple measurement in heavy ion experiments. Anisotropic flow is generally understood as a clear indication of QGP formation, as it is well described by hydrodynamic response to the anisotropy of the initial geometry [77].

(a)

(b)
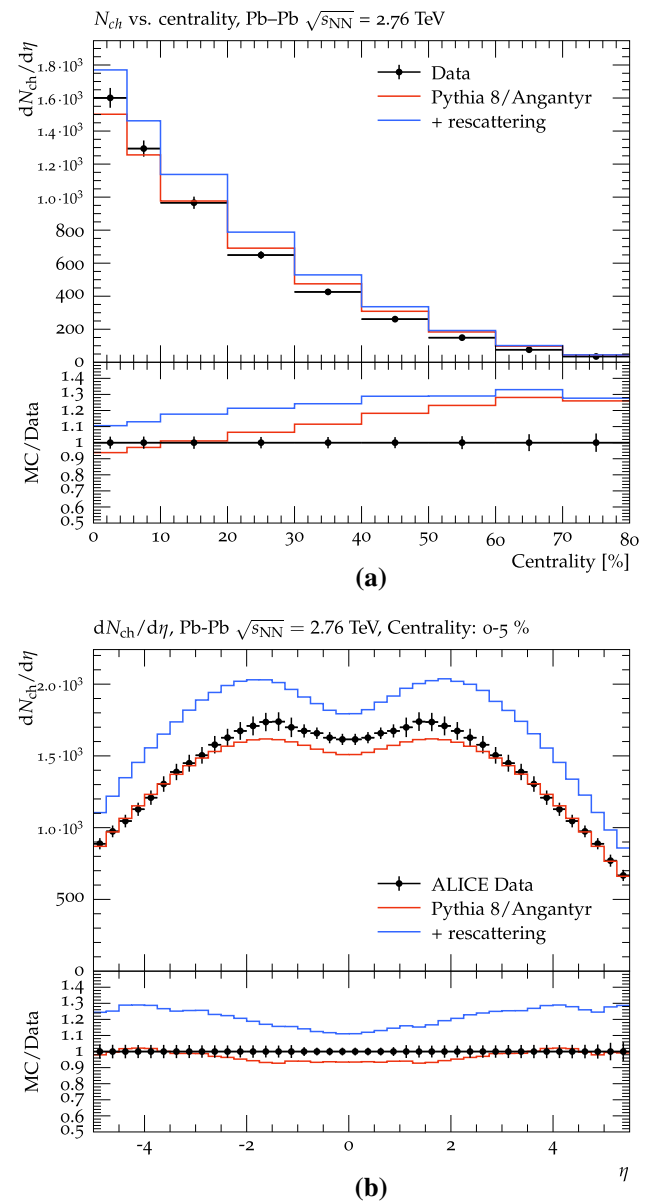
Fig. 11 Charged multiplicities in PbPb collisions $\sqrt{s_{NN}} = 2.76$ TeV. At mid-rapidity as **a** function of centrality, and **b** differential in η in centrality 0–5%. Data from ALICE [74,75]

The main difference between most previous investigations and this paper, of the effect of rescattering on flow, is the early onset of the hadronic phase. Recall that with a hadronization time of $\langle \tau^2 \rangle \approx 2$ fm^2, the initial hadronic state from string hadronization is much denser.
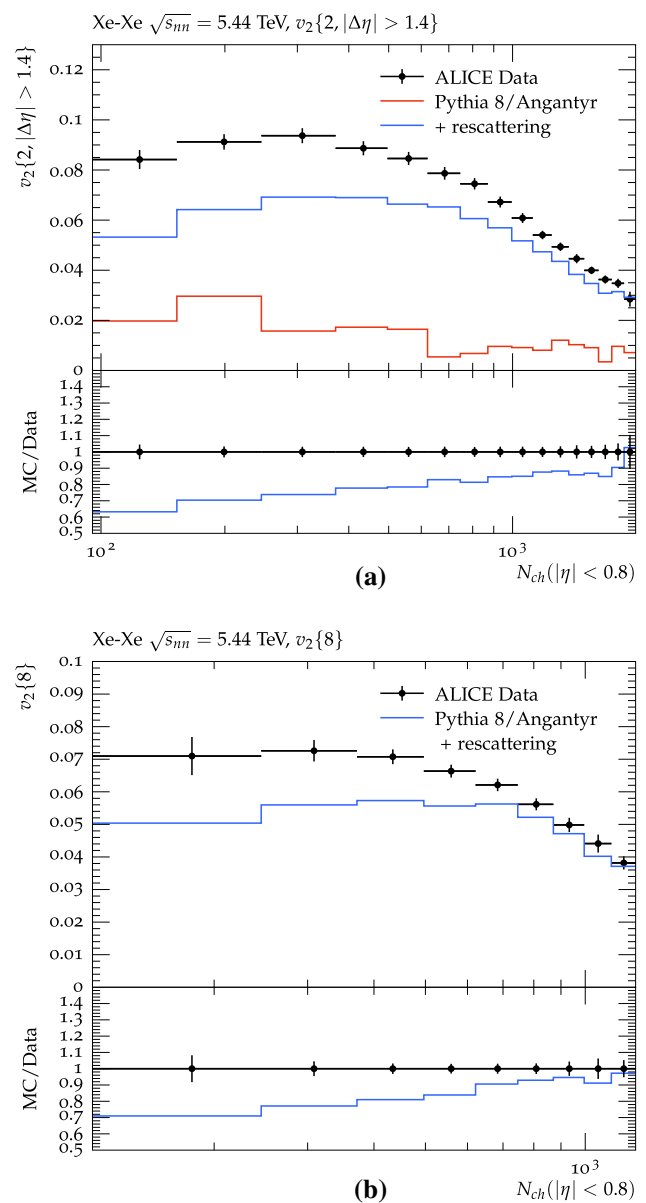
In this section we will compare to experimental data from XeXe and PbPb collisions obtained by the ALICE experiment [13]. When doing so, it is important to use the same definitions of flow coefficients as used by the experiment. Since the event plane is not measurable by experiment, Eqs. (5) and (7) cannot be applied directly. Instead the flow coefficients are calculated using two- and multi-particle azimuthal cor-
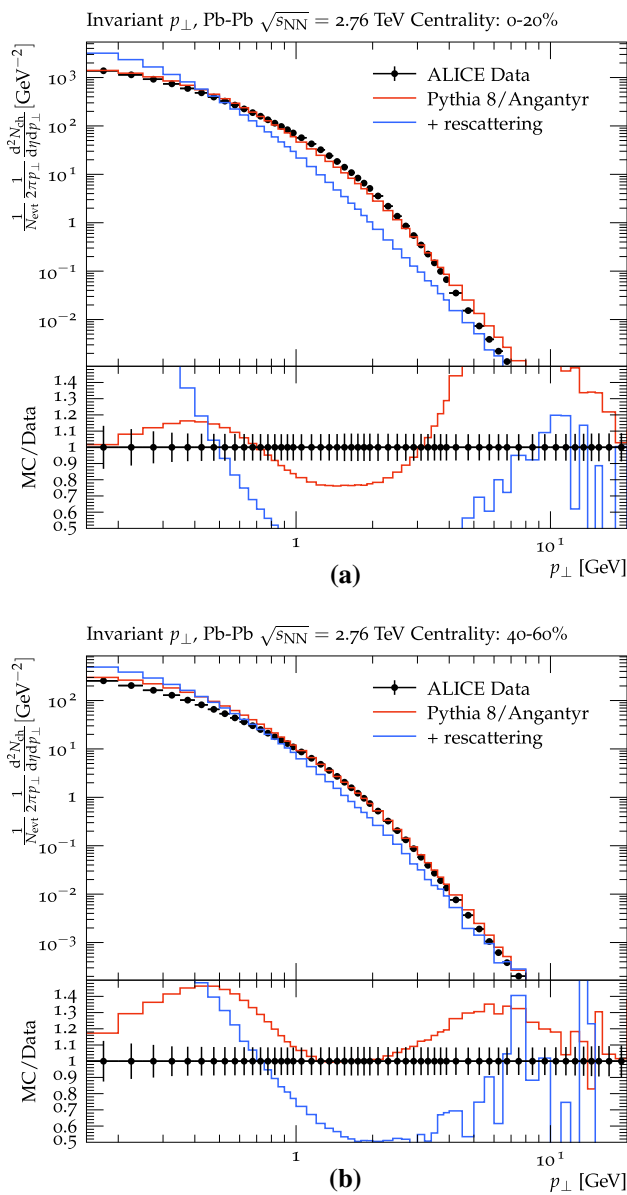
Fig. 12 Invariant $p_\perp$ spectra of charged particles in PbPb collisions $\sqrt{s_{NN}} = 2.76$ TeV, in $|\eta| < 0.8$. Shown for two different centrality intervals **a** 0–20% and **b** 40–60%. Data from ALICE [76]

Fig. 13 Elliptic flow in XeXe collisions at $\sqrt{s_{NN}} = 5.44$ TeV. **a** $v_2\{2\}$ with $|\Delta\eta| > 1.4$ and **b** $v_2\{8\}$. The v_2 calculated with 4- and 6-particle correlations show a similar trend, but are not shown in the figure. Data from ALICE [13]

relations using the so-called generic framework [78], implemented in the Rivet framework [79], including the use of sub-events [80].

In Fig. 13 we show elliptic flow v_2 in XeXe collisions at $\sqrt{s_{NN}} = 5.44$ TeV calculated with (a) two-particle correlations and $|\Delta\eta| > 1.4$, as well as (b) $v_2\{8\}$. In the former case we compare also to the no-rescattering option, which gives a measure of contributions from non-flow mechanisms such as (mini)jets and particle decays. In both cases the data is reproduced with good (within 10%) accuracy for very high multiplicities, but the calculation is up to 30–40% below data for more peripheral events. It is particularly interesting to note

that even in the case of using an 8-particle correlator, the calculation shows the same agreement as only two particles with a gap in η between them. This rules out the possibility that additional flow enters purely from a local increase in two-particle correlations. This should also already be clear from the treatment in Sect. 3.5, where it was clearly shown that the added v_2 by rescattering is in the correct direction with respect to the theoretical event plane.

In Fig. 14 we show the same observables, $v_2\{2, |\Delta\eta| > 1.4\}$ and $v_2\{8\}$ for PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. While the same overall picture is repeated, it is worth noticing
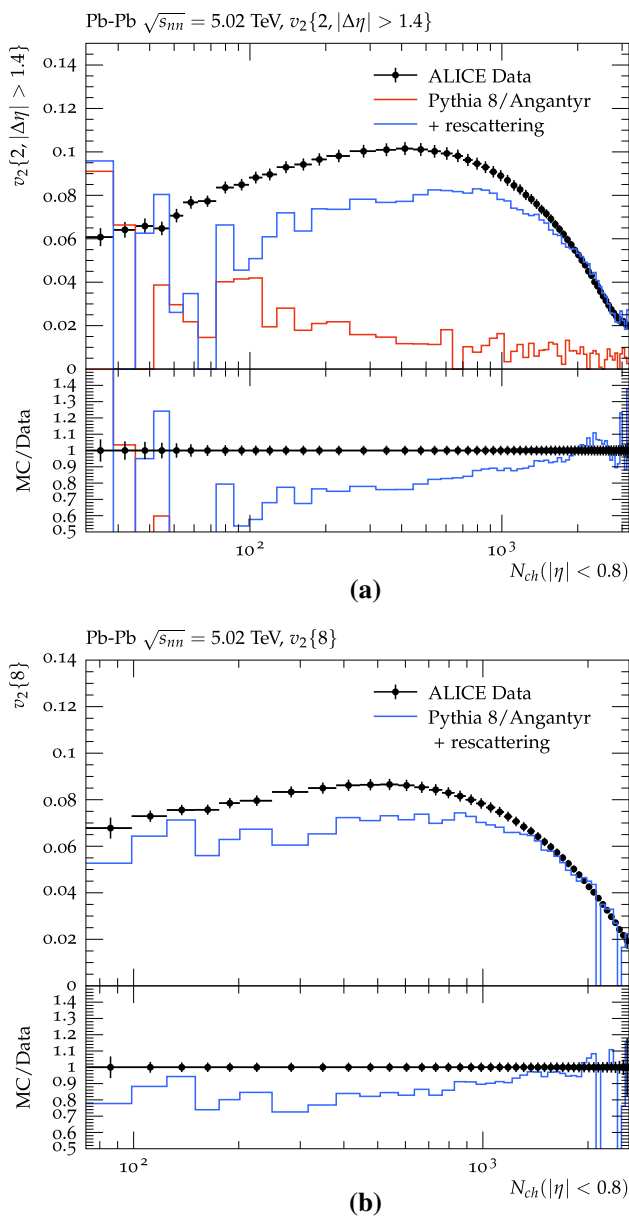
Fig. 14 Elliptic flow in PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. **a** $v_2\{2\}$ with $|\Delta\eta| > 1.4$ and **b** $v_2\{8\}$. The v_2 calculated with 4- and 6-particle correlations show a similar trend, but are not shown in the figure. Data from ALICE [13]

that the agreement at high multiplicities is slightly better. As it was also observed in Sect. 3.5, the effect of rescattering is in general larger in PbPb than in XeXe, due to the larger multiplicity of primaries.

We want here to emphasize that, while hadronic rescattering can obviously not describe data for elliptic flow completely, the results here suggest that hadronic rescattering with early hadronization has a larger effect than previously thought, for example in comparisons of UrQMD to RHIC data [70], where hadronic rescattering accounts for only about 50% of the elliptic flow. This is particularly interesting

seen in the connection with recent results, that interactions between strings before hadronization in the string shoving model [25] will also give a sizeable contribution to flow coefficients in heavy ion collisions, without fully describing data. The combination of the two frameworks, to test whether the combined effect is compatible with data, will be a topic for a future paper. It should be mentioned that the contributions from different models, acting one after the other, does not add linearly [33,81].

4.3 Jet modifications from rescattering

As shown, both in Figs. 3 and 12, hadronic rescattering has a significant effect on high-$p_\perp$ particle production. Studies of how the behaviour of hard particles changes from pp to AA collisions are usually aiming at characterising the interactions between initiator partons and the QGP. The observed phenomena are referred to as "jet quenching", and phenomenological studies usually ignore the presence of a hadronic phase. For a notable exception see Ref. [82] for a recent exploratory study using SMASH, as well as references therein.

In this final results section, we do not wish to go into a full study on the effect of rescattering on jet observables, but rather point to an interesting result which will be pursued further in a future study, as well as warn potential users of the PYTHIA rescattering implementation of a few pitfalls.

One of the early key observations of jet quenching effects was the disappearance of back-to-back high-$p_\perp$ hadron correlations in central AuAu collisions at RHIC [83]. Similar studies have since also been performed at the ALICE experiment, and we compare here to data from a study of azimuthal modifications in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV [84]. In this study, trigger particles of 8 GeV $< p_{\perp,\text{trig}} <$ 15 GeV are correlated in φ with associated particles of 4 GeV $< p_{\perp,\text{assoc}} < p_{\perp,\text{trig}}$. The PbPb/pp ratio of per-trigger yields is denoted I_{AA}, and it was noted in the study by ALICE that the PbPb per-trigger yield is suppressed to about 60% of pp on the away side ($\Delta\varphi$ of $\pi \pm 0.7$) and enhanced by about 20% on the near side ($\Delta\varphi$ of ± 0.7). In Fig. 15, I_{AA} in 0–5% centrality for PbPb collisions is shown on (a) the near-side and (b) the away side, compared to ALICE data [84].

It is seen that, by default, PYTHIA/ANGANTYR overestimates the away-side I_{AA} in the whole $p_{\perp,\text{assoc}}$ range, while the near-side is overestimated at low $p_\perp < 6$ GeV. Adding rescattering brings the simulation on par with data in all cases but the high-$p_\perp$ part of the away-side I_{AA}. No significant effect from rescattering was observed in peripheral events.

At first sight, this seems like a very significant result, but we wish to provide the reader with a word of caution. We remind that the current lack of $n \to 2$ processes and retuning causes a drastic shift in $p_\perp$ spectra as previously shown, incompatible with data. The depletion seen from rescatter-
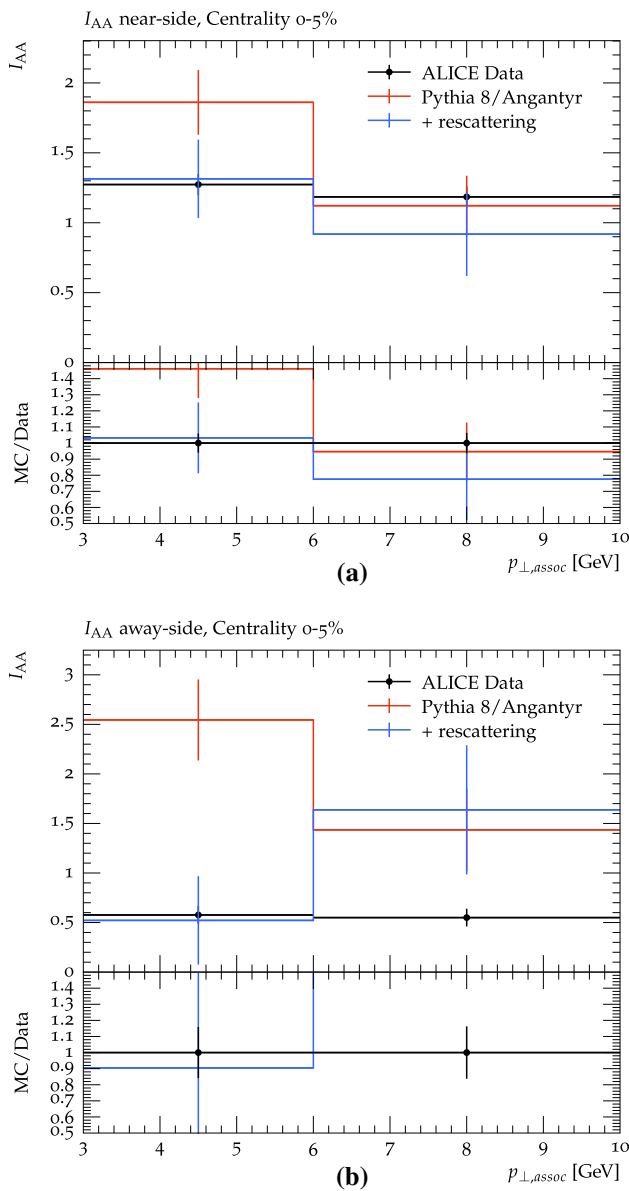
Fig. 15 Modification of high-$p_\perp$ azimuthal correlations, I_{AA}, in PbPb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV on **a** the near side ($\Delta\varphi$ of ± 0.7) and **b** the away side ($\Delta\varphi$ of $\pi \pm 0.7$), both in the 0–5% centrality bin. Error bars are statistical errors. Data from ALICE [84]

ing is exactly in the region where I_{AA} is now well reproduced. It can therefore very well be that the effect seen is mainly a token of current shortcomings. This is of course not a statement that hadronic rescattering has no impact on jet-quenching observables, but it goes to show that a potential user cannot run PYTHIA to explain this or similar observables without a deeper analysis.

Finally a technical remark. Running PYTHIA/ANGANTYR with rescattering to reproduce an observable requiring a high-$p_\perp$ trigger particle will require very long run times. The figures in this section are generated by first requiring that a

parton–parton interaction with $\hat{p}_\perp > 5$ GeV takes place at all, and secondly a veto is put in place ensuring that the time-consuming rescattering process is not performed if there are no trigger particles with the required $p_\perp$ present in the considered acceptance.

5 Summary and outlook

The PYTHIA rescattering framework was first introduced in [34], which focused on validating it in the context of pp collisions. In this paper, our main objective has been to study the model in the context of pA and AA collisions. This has led to a number of observations, both through studying model features and by comparing to data.

The most remarkable physics result is undoubtedly the observation of collective flow, which matches data particularly well at high multiplicities where there is much rescattering activity. Flow is also visible in D meson production, at a slightly lower rate than in the inclusive sample. The flow increases from XeXe to PbPb, i.e. when moving to larger systems. This should be contrasted with the pp results [34], where the rescattering flow effects were tiny and far below data. A rudimentary study shows that rescattering also modifies the PbPb/pp ratio of per-trigger yield I_{AA}, in particular giving a much better description for low $p_{\perp,\text{assoc}}$ on the away-side. This could be an indication of jet quenching, but could also be attributed to an overall reduction of mean $p_\perp$ per particle. Nevertheless, both these observations are usually seen as QGP-signatures, and the fact that they can be reproduced without a QGP is a strong incentive for further studies in a QGP-free paradigm.

In addition to these comparisons to data, we have studied other features of our model in detail without direct comparisons to experiment. In summary, the important conclusions are as follows:

– Since $2 \to n$ processes have been implemented, but not $n \to 2$, rescattering must increase multiplicity of events. This effect is about 20% in PbPb events. Implementing processes with multiple incoming particles would partly mitigate this problem, but an increase is still expected since the system is not in thermal equilibrium.
– Because of the increase in particle multiplicity, the average $p_\perp$ of particles is reduced, and the addition of rescattering leads to an overall worse description of data. However, the mean $p_\perp$ of certain particle species, in particular of nucleons, is increased due to pion wind.
– A few mechanisms modify the particle composition:

   – An increased number of particles means that lighter particles are favoured.

– The overall number of nucleons is reduced, due to baryon–antibaryon annihilation.

– The number of J/ψ is reduced, quite severely in central events. Inelastic events can break apart the $c\bar{c}$ to produce for example a $D\bar{D}$ pair, while the reverse process is very unlikely. Furthermore, charm is not produced in string fragmentation, nor do any other charm production mechanisms exist.

– Hyperon production rates drop, where data wants more such production [14].

– The production vertices of outgoing particles are shifted in space–time. Early produced particles are more likely to participate in rescattering, and thereby early final-state particle production is suppressed. Particles produced by decays at late times ($\tau \gtrsim 50$ fm) are also slightly suppressed, because rescattering enhances lighter particles that are more likely to be stable. A more detailed study of the space–time picture ought to consider resonance suppression and compare to data, but such a study is outside the scope of this article.

With these results in mind, in particular the observation of collective behaviours that are consistent with data and the shift in the $p_\perp$ spectrum away from data, an important follow-up project would combine all the features that have been introduced on top of the basic PYTHIA/ANGANTYR model. In addition to the rescattering, two such features are shoving [23], where an improved modelling [25] will soon be part of the standard code, and rope formation. With all these effects in place, it would also be pertinent to attempt an overall tuning. It is not possible to tell where results will land at the end, since effects tend to add nonlinearly. One can remain optimistic that many features of the data will be described qualitatively, if not quantitatively.

The rescattering framework also lends itself to other applications. One example is as a production mechanism for specific particles. Deuteron production has been modelled in the past using coalescence of particles close in momentum space, on the assumption that such particles have also been produced near each other in space–time [85]. This assumption breaks down in AA where particles are created over a much larger volume. Since rescattering uses the space–time picture directly, it offers a more realistic model for deuteron production. Other particles for which rescattering can be a significant production mechanism are $f_2(1270)$ resonances [86] and exotic hadrons. One potential issue in such studies is the fact that particles that form resonances already are correlated, and hence the appearance of a particle in the event record does not necessarily translate directly to an observable signal. We also see other future applications of space–time information, notably for Bose–Einstein studies.

The rescattering model is made freely available, starting with PYTHIA 8.303, with a few tiny corrections in 8.304 to allow the extension from pp to pA and AA. In the past we have seen how new PYTHIA capabilities have led to follow-up studies by the particle physics community at large, both foreseen and unforeseen ones, and we hope that this will be the case here as well, although admittedly the long run times is a hurdle.

Appendix: Algorithmic complexity

The rescattering algorithm needs to compare each hadron pair. This has an asymptotic complexity of $\mathcal{O}(n_{record}^2)$, where n_{record} is the total number of particles in the event record, including those that are not final-state particles. In practice this asymptotic bound is never reached, since a large number of the comparisons are trivial, e.g. if one of the compared par-

Table 1 The average generation time per event. Events were generated on an Intel(R) Core(TM) i7-6700K CPU at 4.00 GHz

Case	Resc. off	Resc. on	Ratio
pp	2.24 ms	4.02 ms	1.79
pPb	6.40 ms	25.6 ms	4.00
PbPb	0.594 s	150.4 s	253

ticles has already decayed or rescattered. Instead, profiling shows that the bottlenecks are calculating the total cross sections and rescattering vertices for pairs that can potentially rescatter. These are calculated for a much smaller number of pairs, and give a complexity that is less than quadratic in practice.

The average generation time is shown in Table 1. We see that it is in the order of milliseconds for pp and pPb, both with and without rescattering. The rescattering accounts for about 45% of the total runtime for pp, and about 75% for pPb. The situation is radically different for PbPb, where the rescattering takes more than 99.5% of the time, making the average generation time go from less than a second to more than two minutes per event. Thus more careful planning is needed for PbPb rescattering studies, since a rerun will cost.

In Fig. 16 the average generation time per event is shown as a function of the number of primary hadrons. We do not tune the $p_{\perp 0}$ parameter for this study, so that the primary hadron distribution is the same with and without rescattering. The runtime as a function of the primary multiplicity is essentially unchanged by this, however. In all three processes the rescattering overhead is modest for small multiplicities. At the tail towards larger multiplicities the slowdown is about a factor ~ 2 for pp, ~ 20 for pPb and ~ 1000 for PbPb.

In PbPb studies focused on peripheral or mid-centrality events, unnecessarily generating high-multiplicity events can incur a significant slowdown. This can be mitigated by writing an impact-parameter generator tailored to the specific needs, and passing it to PYTHIA via a user hook.

(a)

(b)

(c)
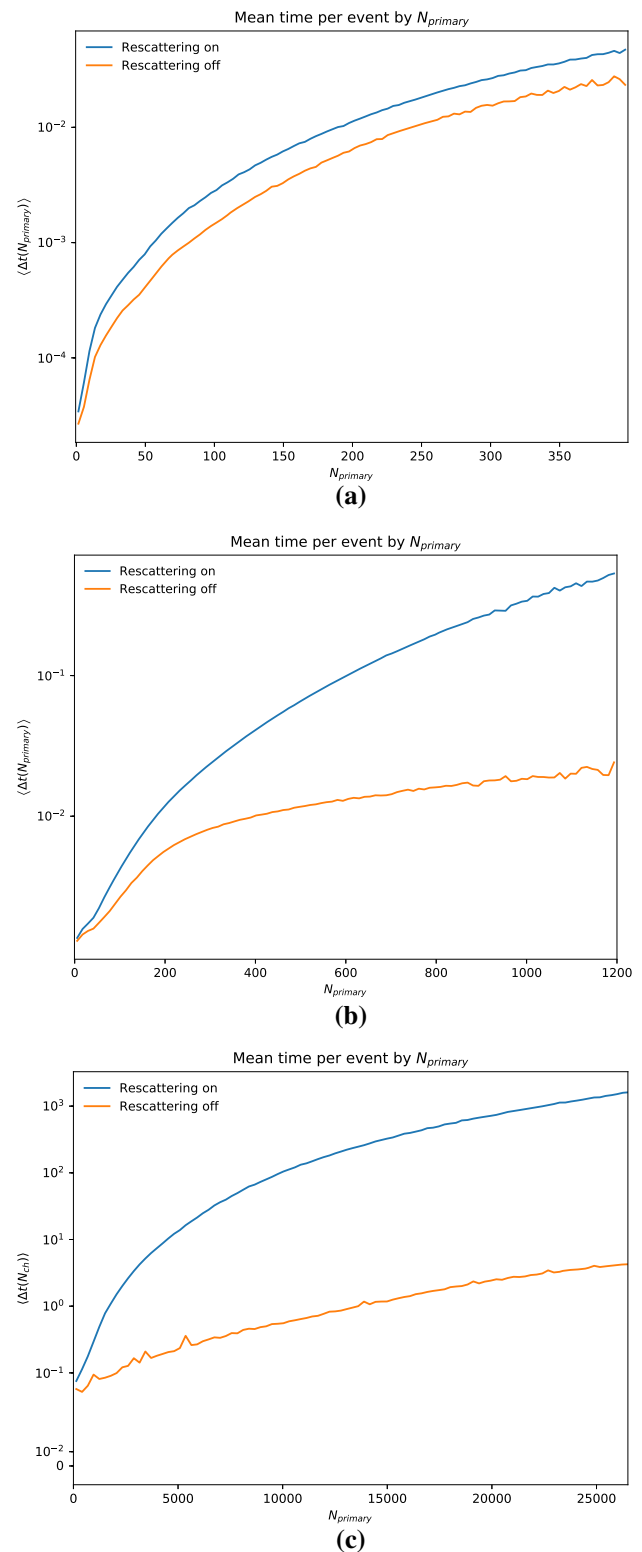
Fig. 16 The average generation time of each event with a specified primary hadron multiplicity, for **a** pp, **b** pPb and **c** PbPb, with $\sqrt{s_{NN}} = 5.02$ TeV

References

1. M.L. Miller, K. Reygers, S.J. Sanders, P. Steinberg, Glauber modeling in high energy nuclear collisions. Ann. Rev. Nucl. Part. Sci. **57**, 205–243 (2007). arXiv:nucl-ex/0701025
2. P. Bożek, W. Broniowski, M. Rybczyński, Wounded quarks in A+A, p+A, and p+p collisions. Phys. Rev. C **94**(1), 014902 (2016). arXiv:1604.07697
3. F. Gelis, E. Iancu, J. Jalilian-Marian, R. Venugopalan, The color glass condensate. Ann. Rev. Nucl. Part. Sci. **60**, 463–489 (2010). arXiv:1002.0333
4. B. Schenke, P. Tribedy, R. Venugopalan, Fluctuating glasma initial conditions and flow in heavy ion collisions. Phys. Rev. Lett. **108**, 252301 (2012). arXiv:1202.6646
5. A. Kurkela, A. Mazeliauskas, J.-F. Paquet, S. Schlichting, D. Teaney, Matching the nonequilibrium initial stage of heavy ion collisions to hydrodynamics with QCD kinetic theory. Phys. Rev. Lett. **122**(12), 122302 (2019). arXiv:1805.01604
6. U. Heinz, R. Snellings, Collective flow and viscosity in relativistic heavy-ion collisions. Ann. Rev. Nucl. Part. Sci. **63**, 123–151 (2013). arXiv:1301.2826
7. M. Luzum, H. Petersen, Initial state fluctuations and final state correlations in relativistic heavy-ion collisions. J. Phys. G **41**, 063102 (2014). arXiv:1312.5503
8. C. Gale, S. Jeon, B. Schenke, Hydrodynamic modeling of heavy-ion collisions. Int. J. Mod. Phys. A **28**, 1340011 (2013). arXiv:1301.5893
9. R.D. Weller, P. Romatschke, One fluid to rule them all: viscous hydrodynamic description of event-by-event central p+p, p+Pb and Pb+Pb collisions at $\sqrt{s}$ = 5.02 TeV. Phys. Lett. B **774**, 351–356 (2017). arXiv:1701.07145
10. CMS Collaboration, V. Khachatryan et al., Observation of long-range near-side angular correlations in proton-proton collisions at the LHC. JHEP **09**, 091 (2010). arXiv:1009.4122
11. ATLAS Collaboration, G. Aad et al., Observation of long-range elliptic azimuthal anisotropies in $\sqrt{s}$ =13 and 2.76 TeV pp collisions with the ATLAS detector. Phys. Rev. Lett. **116**(17), 172301 (2016). arXiv:1509.04776
12. C.M.S. Collaboration, V. Khachatryan et al., Evidence for collectivity in pp collisions at the LHC. Phys. Lett. B **765**, 193–220 (2017). arXiv:1606.06198
13. A.L.I.C.E. Collaboration, S. Acharya et al., Investigations of anisotropic flow using multiparticle azimuthal correlations in pp, p-Pb, Xe-Xe, and Pb-Pb collisions at the LHC. Phys. Rev. Lett. **123**(14), 142301 (2019). arXiv:1903.01790
14. A.L.I.C.E. Collaboration, J. Adam et al., Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions. Nat. Phys. **13**, 535–539 (2017). arXiv:1606.07424
15. K. Werner, Core-corona separation in ultra-relativistic heavy ion collisions. Phys. Rev. Lett. **98**, 152301 (2007). arXiv:0704.1270
16. T. Pierog, I. Karpenko, J. 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
17. 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
18. C. Bierlich, G. Gustafson, L. Lönnblad, H. Shah, The Angantyr model for heavy-ion collisions in PYTHIA8. JHEP **10**, 134 (2018). arXiv:1806.10820
19. T. Sjostrand, S. Mrenna, P.Z. Skands, PYTHIA 6.4 physics and manual. JHEP **05**, 026 (2006). arXiv:hep-ph/0603175
20. 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
21. 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)
22. B. Andersson, G. Gustafson, H. Pi, The FRITIOF model for very high-energy hadronic collisions. Z. Phys. C **57**, 485–494 (1993)
23. C. Bierlich, G. Gustafson, L. Lönnblad, A shoving model for collectivity in hadronic collisions. arXiv:1612.05132
24. C. Bierlich, G. Gustafson, L. Lönnblad, Collectivity without plasma in hadronic collisions. Phys. Lett. B **779**, 58–63 (2018). arXiv:1710.09725
25. C. Bierlich, S. Chakraborty, G. Gustafson, L. Lönnblad, Setting the string shoving picture in a new frame. arXiv:2010.07595
26. C. Bierlich, G. Gustafson, L. Lönnblad, A. Tarasov, Effects of overlapping strings in pp collisions. JHEP **03**, 148 (2015). arXiv:1412.6259
27. S. Bass et al., Microscopic models for ultrarelativistic heavy ion collisions. Prog. Part. Nucl. Phys. **41**, 255–369 (1998). arXiv:nucl-th/9803035
28. J. Weil et al., Particle production and equilibrium properties within a new hadron transport approach for heavy-ion collisions. Phys. Rev. C **94**(5), 054905 (2016). arXiv:1606.06642
29. W. Cassing, E.L. Bratkovskaya, Hadronic and electromagnetic probes of hot and dense nuclear matter. Phys. Rep. **308**, 65–233 (1999)
30. Z.-W. Lin, C.M. Ko, B.-A. Li, B. Zhang, S. Pal, A multi-phase transport model for relativistic heavy ion collisions. Phys. Rev. C **72**, 064901 (2005). arXiv:nucl-th/0411110
31. Y.-X. Zhang et al., Comparison of heavy-ion transport simulations: collision integral in a box. Phys. Rev. C **97**(3), 034625 (2018). arXiv:1711.05950
32. S. Ferreres-Solé, T. Sjöstrand, The space-time structure of hadronization in the Lund model. Eur. Phys. J. C **78**(11), 983 (2018). arXiv:1808.04619
33. A.V. da Silva, W.M. Serenone, D. Dobrigkeit Chinellato, J. Takahashi, C. Bierlich, Studies of heavy-ion collisions using PYTHIA Angantyr and UrQMD. arxiv:2002.10236
34. T. Sjöstrand, M. Utheim, A framework for hadronic rescattering in pp collisions. Eur. Phys. J. C **80**(10), 907 (2020). arXiv:2005.05658
35. W. Broniowski, M. Rybczynski, P. Bozek, GLISSANDO: Glauber initial-state simulation and more. Comput. Phys. Commun. **180**, 69–83 (2009). arXiv:0710.5731
36. M. Rybczynski, G. Stefanek, W. Broniowski, P. Bozek, GLISSANDO 2: GLauber initial-state simulation AND mOre..., ver. 2. Comput. Phys. Commun. **185**, 1759–1772 (2014). arXiv:1310.5475
37. G.A. Schuler, T. Sjostrand, Hadronic diffractive cross-sections and the rise of the total cross-section. Phys. Rev. D **49**, 2257–2267 (1994)
38. X.-N. Wang, M. Gyulassy, HIJING: a Monte Carlo model for multiple jet production in p p, p A and A A collisions. Phys. Rev. D **44**, 3501–3516 (1991)
39. T. Sjostrand, M. van Zijl, A multiple interaction model for the event structure in hadron collisions. Phys. Rev. D **36**, 2019 (1987)
40. T. Sjöstrand, The development of MPI modeling in Pythia. Adv. Ser. Direct. High Energy Phys. **29**, 191–225 (2018). arXiv:1706.02166
41. C. Bierlich, C.O. Rasmussen, Dipole evolution: perspectives for collectivity and γ^*A collisions. JHEP **10**, 026 (2019). arXiv:1907.12871
42. T. Sjostrand, P.Z. Skands, Multiple interactions and the structure of beam remnants. JHEP **03**, 053 (2004). arXiv:hep-ph/0402078
43. B. Andersson, G. Gustafson, G. Ingelman, T. Sjostrand, Parton fragmentation and string dynamics. Phys. Rep. **97**, 31–145 (1983)

44. G. 't Hooft, A planar diagram theory for strong interactions. Nucl. Phys. B **72**, 461 (1974)

45. T. Sjostrand, Jet fragmentation of nearby partons. Nucl. Phys. B **248**, 469–502 (1984)

46. Z. Xu, C. Greiner, Thermalization of gluons in ultrarelativistic heavy ion collisions by including three-body interactions in a parton cascade. Phys. Rev. C **71**, 064901 (2005). arXiv:hep-ph/0406278

47. R. Garcia-Martin, R. Kaminski, J. Pelaez, J. Ruiz de Elvira, F. Yndurain, The Pion-pion scattering amplitude. IV: improved analysis with once subtracted Roy-like equations up to 1100 MeV. Phys. Rev. D **83**, 074004 (2011). arXiv:1102.2183

48. J. Pelaez, A. Rodas, J. Ruiz De Elvira, Global parameterization of $\pi\pi$ scattering up to 2 GeV. Eur. Phys. J. C **79**(12), 1008 (2019). arXiv:1907.13162

49. J. Pelaez, A. Rodas, Pion-kaon scattering amplitude constrained with forward dispersion relations up to 1.6 GeV. Phys. Rev. D **93**(7), 074025 (2016). arXiv:1602.08404

50. Particle Data Group Collaboration, M. Tanabashi et. al., Review of particle physics. Phys. Rev. D **98**(3), 030001 (2018)

51. Particle Data Group Collaboration, L. Montanet et. al., Review of particle properties. Particle Data Group. Phys. Rev. D **50**, 1173–1823 (1994)

52. P. Koch, C. Dover, $K^{\pm}$, $\bar{p}$ and Ω^- production in relativistic heavy ion collisions. Phys. Rev. C **40**, 145–155 (1989)

53. G.A. Schuler, T. Sjostrand, A scenario for high-energy gamma gamma interactions. Z. Phys. C **73**, 677–688 (1997). arXiv:hep-ph/9605240

54. E. Levin, L. Frankfurt, The quark hypothesis and relations between cross-sections at high-energies. JETP Lett. **2**, 65–70 (1965)

55. H.J. Lipkin, Quarks for pedestrians. Phys. Rep. **8**, 173–268 (1973)

56. ALICE Collaboration, J. Adam et al., Measurement of pion, kaon and proton production in proton-proton collisions at $\sqrt{s} = 7$ TeV. Eur. Phys. J. C **75**(5), 226 (2015). arXiv:1504.00024

57. N. Fischer, T. Sjöstrand, Thermodynamical string fragmentation. JHEP **01**, 140 (2017). arXiv:1610.09818

58. ALICE Collaboration, S. Acharya et al., Pion-kaon femtoscopy and the lifetime of the hadronic phase in Pb−Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. B **813**, 136030 (2021). arXiv:2007.08315

59. ALICE Collaboration, S. Acharya et al., Evidence of rescattering effect in Pb-Pb collisions at the LHC through production of $K^*(892)^0$ and $\phi(1020)$ mesons. Phys. Lett. B **802**, 135225 (2020). arXiv:1910.14419

60. C. Bierlich, J.R. Christiansen, Effects of color reconnection on hadron flavor observables. Phys. Rev. D **92**(9), 094010 (2015). arXiv:1507.02091

61. ALICE Collaboration, S. Acharya et al., Centrality and transverse momentum dependence of inclusive J/ψ production at midrapidity in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. Phys. Lett. B **805**, 135434 (2020). arXiv:1910.14404

62. T. Lang, M. Bleicher, Possibility for J/ψ suppression in high-multiplicity proton-proton collisions at $\sqrt{s_{NN}} = 7$ TeV. Phys. Rev. C **87**(2), 024907 (2013). arXiv:1302.0655

63. W. Cassing, E.L. Bratkovskaya, Parton-hadron-string dynamics: an off-shell transport approach for relativistic energies. Nucl. Phys. A **831**, 215–242 (2009). arXiv:0907.5331

64. T. Song, H. Berrehrah, D. Cabrera, J.M. Torres-Rincon, L. Tolos, W. Cassing, E. Bratkovskaya, Tomography of the quark-gluon-plasma by charm quarks. Phys. Rev. C **92**(1), 014910 (2015). arXiv:1503.03039

65. T. Song, H. Berrehrah, D. Cabrera, W. Cassing, E. Bratkovskaya, Charm production in Pb + Pb collisions at energies available at the CERN Large Hadron Collider. Phys. Rev. C **93**(3), 034906 (2016). arXiv:1512.00891

66. S. Voloshin, Y. Zhang, Flow study in relativistic nuclear collisions by Fourier expansion of Azimuthal particle distributions. Z. Phys. C **70**, 665–672 (1996). arXiv:hep-ph/9407282

67. A.M. Poskanzer, S.A. Voloshin, Methods for analyzing anisotropic flow in relativistic nuclear collisions. Phys. Rev. C **58**, 1671–1678 (1998). arXiv:nucl-ex/9805001

68. S.T.A.R. Collaboration, J. Adams et al., Particle type dependence of azimuthal anisotropy and nuclear modification of particle production in Au + Au collisions at s(NN)**(1/2) = 200-GeV. Phys. Rev. Lett. **92**, 052302 (2004). arXiv:nucl-ex/0306007

69. PHENIX Collaboration, S.S. Adler et al., Elliptic flow of identified hadrons in Au+Au collisions at s(NN)**(1/2) = 200-GeV. Phys. Rev. Lett. **91**, 182301 (2003). arXiv:nucl-ex/0305013

70. Y. Lu, M. Bleicher, F. Liu, Z. Liu, H. Petersen, P. Sorensen, H. Stoecker, N. Xu, X. Zhu, Anisotropic flow at RHIC: how unique is the number-of-constituent-quark scaling? J. Phys. G **32**, 1121–1130 (2006). arXiv:nucl-th/0602009

71. ALICE Collaboration, B.B. Abelev et al., Azimuthal anisotropy of D meson production in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Rev. C **90**(3), 034904 (2014). arXiv:1405.2001

72. ALICE Collaboration, E. Abbas et al., J/Psi elliptic flow in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Rev. Lett. **111**, 162301 (2013). arXiv:1303.5880

73. C. Bierlich et al., Robust independent validation of experiment and theory: Rivet version 3. SciPost Phys. **8**, 026 (2020). arXiv:1912.05451

74. ALICE Collaboration, K. Aamodt et al., Centrality dependence of the charged-particle multiplicity density at mid-rapidity in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Rev. Lett. **106**, 032301 (2011). arXiv:1012.1657

75. ALICE Collaboration, E. Abbas et al., Centrality dependence of the pseudorapidity density distribution for charged particles in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. B **726**, 610–622 (2013). arXiv:1304.0347

76. ALICE Collaboration, B. Abelev et al., Centrality dependence of charged particle production at large transverse momentum in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. B **720**, 52–62 (2013). arXiv:1208.2711

77. D. Teaney, L. Yan, Triangularity and dipole asymmetry in heavy ion collisions. Phys. Rev. C **83**, 064904 (2011). arXiv:1010.1876

78. A. Bilandzic, C.H. Christensen, K. Gulbrandsen, A. Hansen, Y. Zhou, Generic framework for anisotropic flow analyses with multiparticle azimuthal correlations. Phys. Rev. C **89**(6), 064904 (2014). arXiv:1312.3572

79. C. Bierlich et al., Confronting experimental data with heavy-ion models: RIVET for heavy ions. Eur. Phys. J. C **80**(5), 485 (2020). arXiv:2001.10737

80. P. Huo, K. Gajdošová, J. Jia, Y. Zhou, Importance of non-flow in mixed-harmonic multi-particle correlations in small collision systems. Phys. Lett. B **777**, 201–206 (2018). arXiv:1710.07567

81. J. Auvinen, H. Petersen, Evolution of elliptic and triangular flow as a function of $\sqrt{s_{NN}}$ in a hybrid model. Phys. Rev. C **88**(6), 064908 (2013). arXiv:1310.1764

82. P. Dorau, J.-B. Rose, D. Pablos, H. Elfner, Jet quenching in the hadron gas: an exploratory study. Phys. Rev. C **101**(3), 035208 (2020). arXiv:1910.07027

83. S.T.A.R. Collaboration, C. Adler et al., Disappearance of back-to-back high p_T hadron correlations in central Au+Au collisions at $\sqrt{s_{NN}} = 200$-GeV. Phys. Rev. Lett. **90**, 082302 (2003). arXiv:nucl-ex/0210033

84. ALICE Collaboration, K. Aamodt et al., Particle-yield modification in jet-like azimuthal di-hadron correlations in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Rev. Lett. **108**, 092301 (2012). arXiv:1110.0121

85. L.A. Dal, A.R. Raklev, Alternative formation model for antideuterons from dark matter. Phys. Rev. D **91**(12), 123536 (2015). arXiv:1504.07242. [Erratum: Phys. Rev. D **92**, 069903 (2015). Erratum: Phys. Rev. D **92**, 089901 (2015)]

86. P. Lebiedowicz, A. Szczurek, Production of $f_2(1270)$ meson in pp collisions at the LHC via gluon-gluon fusion in the k_t-factorization approach. Phys. Lett. B **810**, 135816 (2020). arXiv:2007.12485