Hadronic Interactions at High and Low Energies

Utheim, Marius

2021

Document Version:
Publisher's PDF, also known as Version of record

Hadronic Interactions at High and Low Energies

Hadronic Interactions at High and Low Energies

by Marius Utheim

LUND
UNIVERSITY

Thesis for the degree of Doctor of Philosophy
Thesis advisor: Prof. Torbjörn Sjöstrand
Faculty opponent: Prof. Dr. Hannah Elfner

To be presented, with the permission of the Faculty of Science of Lund University,
for public criticism in Lundmarksalen at the Department of Astronomy and Theoretical Physics
on Friday the 17th of September 2021 at 10:00

Organization
LUND UNIVERSITY

Department of Astronomy and Theoretical Physics
Sölvegatan 14A
SE-223 62 Lund
Sweden

Document name
DOCTORAL DISSERTATION

Date of disputation
2021-09-17

Sponsoring organization

Author(s)
Marius Utheim

Title and subtitle
Hadronic Interactions at High and Low Energies

Abstract
In the past, development of the Pythia event generator has mainly concentrated on the physics of proton–proton collisions at high energies, as this is the most important case to modern high energy physics experiments. However, there are situations where more generic beam configurations are relevant. This thesis presents implementations and applications of such processes in Pythia. The two applications discussed in this thesis are hadronic rescattering, and hadronic cascades in media. These are developed in Papers I and IV, while Papers II and III present physics studies of rescattering.

Paper I: A framework for hadronic rescattering is implemented. This includes both the rescattering algorithm itself, as well as modelling low energy interactions. Model tests are presented for proton–proton collisions.

Paper II: The effects of rescattering in heavy ion collisions are studied, using the Angantyr framework to simulate the collisions. Of note, results indicate that rescattering can produce a significant amount of collective flow.

Paper III: This work studies rescattering as a production mechanism for exotic hadrons, specifically the $\chi_{c1}(3872)$ tetraquark and the $P_c^+(4312)$, $P_c^+(4440)$ and $P_c^+(4457)$ pentaquarks.

Paper IV: Modelling of high energy collisions in Pythia is extended to general hadron–nucleon interactions. The major parts of this extension are to define cross sections for these interactions and parton distribution functions for the hadrons involved. The paper studies the features of this model in the context of hadronic cascades in media, primarily cascades initiated by cosmic rays in the atmosphere, using a toy model.

Key words
QCD, Phenomenology, Hadronic rescattering, Heavy ions, Exotic hadrons, Cosmic rays

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title

ISBN
978-91-7895-973-0 (print)
978-91-7895-974-7 (pdf)

Recipient's notes

Number of pages
256

Price

Security classification

Hadronic Interactions at High and Low Energies

by Marius Utheim

LUND
UNIVERSITY

A doctoral thesis at a university in Sweden takes either the form of a single, cohesive research study (monograph) or a summary of research papers (compilation thesis), which the doctoral student has written alone or together with one or several other author(s).

In the latter case the thesis consists of two parts. An introductory text puts the research work into context and summarizes the main points of the papers. Then, the research publications themselves are reproduced, together with a description of the individual contributions of the authors. The research papers may either have been already published or are manuscripts at various stages (in press, submitted, or in draft).

Cover illustration: E6 (Credits: Hans Kristian Utheim).

Funding information: This thesis was supported primarily by the MCnetITN3 H2020 Marie Curie Innovative Training Network, Grant agreement 722104. It has also received funding from the Swedish Research Council, Contract number 2016-05996, and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant agreement no. 668679.

Faculty of Science, Department of Astronomy and Theoretical Physics

ISBN: 978-91-7895-973-0 (print)
ISBN: 978-91-7895-974-7 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2021

Til mor og bestefar

Contents

List of publications

This thesis is based on the following publications, referred to by their Roman numerals:

I **A Framework for Hadronic Rescattering in pp Collisions**
T. Sjöstrand, M. Utheim
Eur. Phys. J. C **80**, 907 (2020)
e-Print: `arXiv:2005.05658` [hep-ph]
MCnet–20-11, LU TP 20–12

II **Hadronic Rescattering in pA and AA Collisions**
C. Bierlich, T. Sjöstrand, M. Utheim
Eur. Phys. J. A **57**, 227 (2021)
e-Print: `arXiv:2103.09665` [hep-ph]
MCnet–21–03, LU TP 21–08

III **Forming Molecular States with Hadronic Rescattering**
P. J. Ilten, M. Utheim
To be submitted to *Eur. Phys. J. A*,
e-Print: `arXiv:2108.03479` [hep-ph]
MCnet–21–13, LU TP 21–31

IV **Hadron Interactions for Arbitrary Energies and Species, with Applications to Cosmic Rays**
T. Sjöstrand, M. Utheim
To be submitted to *Eur. Phys. J. C*,
e-Print: `arXiv:2108.03481` [hep-ph]
MCnet–21–14, LU TP 21–32

Popular summary in English

Physics research is about understanding how the universe works on a fundamental level. It tries to address questions such as how do things move, how does electricity work, and why do stars burn.

One deep physics question is what the fundamental building blocks of matter are. If you take a rock and smash it into pieces, then grind the pieces down to specks of dust, and continue breaking the grains of dust into smaller and smaller pieces, will you eventually reach a point where it is no longer possible to break it into smaller bits? Or, given the right tools and technology, will it always possible to break it into smaller pieces? If there are some smallest possible pieces, do they behave similarly to the large things we observe in our everyday lives, or are they very different?

The philosophical idea that perhaps the laws of our universe are such that there is some smallest fundamental pieces of matter is called *atomism*. Over the course of the last couple centuries, humans have discovered that the answer seems to be that atomism is correct, and that there are at least 17 types of fundamental building blocks in our universe. These are called *elementary particles*, and everything we experience in our everyday lives – rocks, water, air, light, plants, animals, stars – is made from combinations of elementary particles. *Particle physics* is about studying the properties and behaviours of these elementary particles. The best description of particle physics that we have today is a theory called the *Standard Model*.

One of the most important questions in science is *how* do we know what we know? Our theories must predict something about reality, and it must be possible to perform experiments to test whether reality really behaves the way the theory predicts. Otherwise, it is not science, but science fiction. In particle physics, we make predictions by using computer simulations of our theoretical models. These predictions can then be compared to actual data, like the data from experiments at the Large Hadron Collider (LHC) at CERN. One example of such a simulation program is Pythia, whose development began in Lund about 40 years ago, and which this thesis revolves around.

There are many unanswered questions in particle physics. One question that a lot of particle physicists are working on these days is what happens when certain particles are put under extreme pressure and heated up to temperatures a 100,000 times warmer than the core of the sun. With what we know today, the most likely answer seems to be that these particles would "melt" into a new state of matter called a *quark-gluon plasma* (QGP), but we still don't really understand how this happens, what the properties of this state of matter are, or if it can be used for anything interesting. The desire to learn the answers to questions like these is profound for many people – so much that some are willing to dedicate lifetimes of work and invest billions of euros on developing technology like the LHC that can help us

find the answers.

Now, the QGP is so extreme that it evaporates within a fraction of a billionth of a nanosecond, which makes it impossible to detect it directly. Instead, many scientists believe in the QGP hypothesis because it makes certain predictions that turn out to fit really well with data. However, one of the ongoing research programmes in Lund is challenging this hypothesis, and asks whether there are other models that are unrelated to the QGP, but that still can make the same predictions. Building such a model requires a lot of additions to the Pythia simulation. The work I have done in Lund has been to make such improvements to Pythia, both as a contribution to this bigger project to challenge the QGP paradigm, and as part of other projects. This thesis is the culmination of that work.

Populærvitenskaplig sammendrag på norsk

Fysikk handler om å forstå hvordan universet fungerer. Typiske fysikkspørsmål er for eksempel hvordan ting beveger seg, hvordan elektrisitet fungerer, eller hvorfor stjerner lyser.

Et dypt fysikkspørsmål er om det finnes noen fundamentale byggesteiner som all materie består av. Hvis du tar en stein og knuser den i småbiter, maler småbitene til støv, og fortsetter å bryte støvkornene i mindre og mindre deler, vil du da til slutt komme til et punkt der det ikke lengre er mulig å dele dem i mindre deler? Eller kan man med rett verktøy og teknologi alltid bryte noe ned til mindre deler? Om det finnes noen minste mulige byggesteiner, har disse byggesteinene samme egenskaper som de tingene vi er vant til å møte i hverdagen, eller er de veldig forskjellige?

Idéen om at universets lover kanskje er slik at det finnes noen minste fundamentale bestanddeler som ikke kan deles, uansett hvor hardt man forsøker, kalles *atomisme*, og er et verdenssyn som går tilbake til de gamle grekerne. Basert på det vitenskapen har oppdaget de siste par århundrene, så ser det ut til at atomismen stemmer. Så vidt vi vet finnes det minst 17 typer fundamentale byggesteiner i universet. Disse kalles *elementærpartikler*, og alt vi møter i hverdagen – fra steiner, vann, luft og lys, til planter, dyr og stjerner – er laget av forskjellige kombinasjoner av elementærpartiklene. Målet med *partikkelfysikk* er å forstå egenskapene til disse elementærpartiklene, og hvordan de oppfører seg på et fundamentalt nivå. Den beste beskrivelsen vi har i dag er en teori som heter *Standardmodellen*.

Ett av de viktigste spørsmålene i forskning er hvordan vi vet det vi vet? Vitenskaplige teorier må påstå noe om virkeligheten, og det må være mulig å gjøre eksperimenter som tester om virkeligheten faktisk er slik teoriene forutsier. I partikkelfysikk gjøres forutsigelser ved hjelp av dataprogrammer som simulerer de teoretiske modellene. Disse forutsigelsene kan testes gjennom å gjøre eksperimenter som for eksempel de som gjøres ved LHC (Large Hadron Collider) på CERN i Sveits, og på den måten kan man sammenlikne teori med faktiske data. Et eksempel på et dataprogram som gjør slike simuleringer er Pythia, som startet utviklingen i Lund for omlag 40 år siden. Denne avhandlingen dreier seg i stor grad om Pythia.

Det er flere ubesvarte spørsmål i partikkelfysikk. Et spørsmål som mange partikkelfysikere er opptatte av for tiden er hva som skjer med visse partikler når man utsetter dem for ekstremt trykk og varmer dem opp til 100 000 ganger varmere enn solas kjerne. Basert på det vi vet i dag er den regjerende hypotesen at partiklene "smelter" og går over til en ny fasetilstand som kalles *kvark-gluon-plasma* (QGP, fra engelsk *quark–gluon plasma*). Men det er ingen som helt vet hvordan dette skjer, hva som er egenskapene til denne fasen, eller om det kan brukes til noe spennende. Mange mennesker mener at slike spørsmål er så viktige for vår sivilisasjon og kultur at de er villige til å dedikere livene sine til å forske på det, og å investere flere titalls milliarder kroner på å bygge verktøy som LHC for å finne svaret.

QGP er så ekstremt at det fordamper etter bare brøkdelen av et milliarddels nanosekund, noe som gjør det umulig å studere direkte. I stedet er grunnlaget for QGP-modellen at den gjør visse prediksjoner som viser seg å stemme godt overens med data. Et pågående forkningsprosjekt i Lund stiller spørsmålet om det da også finnes andre modeller som gir de samme prediksjonene, for hvis det gjør det betyr det at QGP-modellen ikke nødvendigvis er forklaringen. For å simulere en alternativ modell kreves det mye videreutvikling av Pythia-programmet. Det jeg har jobbet med i Lund har handlet om å gjøre slike videreutviklinger, både for å utfordre QGP-modellen, og i forbindelse med andre prosjekter. Denne avhandlingen er resultatet av arbeidet jeg har gjort.

Introduction

This thesis is a collection of the work I have accomplished during my time in Lund. It is divided into two parts: an introduction part where I outline the theories and concepts that underlie my research, and the main body that consists of the four papers I have written. Before we get to the physics content, I would like to say a few words about what this thesis means to me.

There are many reasons to do science, both pragmatic and idealistic. For me personally, one of the most important motivations is that I enjoy learning things about the world. I am fascinated by documentaries about life in the deepest part of the oceans, about supermassive black holes, and about the discovery of the Higgs boson. I will never witness these parts of nature with my own eyes, so in a sense they are just stories to me. But I do find these stories particularly fascinating because they are *true*. At this very moment, there really are animals living their lives at the bottom of the Mariana trench, and with the right technology, humans can travel there to look at them. Similarly, black holes actually are out there, and one day in a far distant future, perhaps humans will develop the technology to travel close enough to observe them firsthand.

This thesis then represents the story that I have discovered and want to tell. I was given the opportunity to spend this time doing research thanks to grants that down the line came from tax money. Society is paying me to figure out these things, so that the knowledge can later be shared with others – I take this as a sign that I'm not the only one who's interested in these stories. Therefore, an important part of doing science is to disseminate the results and share them with people in a way that does not demand years of study to understand. In Section 1 of this thesis, I try to describe my work in a way that is accessible to non-experts.

Specifically, in Section 1.1, I start by giving a basic introduction to particle physics. In Section 1.2, I give a short overview of how modern particle physics experiments are performed, and where my work fits into the scientific landscape. Section 1.3 will introduce some open questions in particle physics. These are questions that are too big to be answered over the

course of only a four-year PhD, but my work represents a few steps towards deeper understanding. Then in Section 1.4, I will get slightly more technical and discuss some basic ideas behind one of the fundamental forces in the universe, namely the *strong force*, which this thesis is focused on.

Starting from Section 2, I will shift my focus towards describing more technical details of the theories that underlie my research, and the tone of writing becomes significantly more advanced. The common theme of this thesis is hadronic interactions, and all my work has resulted in extensions to the Pythia software. In Section 2, we discuss theoretical concepts behind such interactions, and see how they are implemented in Pythia. An important theoretical concept is that of cross sections, and this is discussed in Section 3. Two concepts that are particularly relevant to my work are parton distribution functions and hadronization, which will be discussed in Sections 4 and 5, respectively. Section 6 discusses the applications of the new interactions that I have implemented. The applications addressed in this thesis are hadronic rescattering and hadronic propagation through media. Hadronic rescattering in particular plays a large role in this thesis, with Papers II and III being focused on physics that involve rescattering. Finally, some concluding remarks and outlooks are presented in Section 7.

1 Overview of particle physics

Nothing exists except atoms and empty space, everything else is opinion.

– Democritus, *c.* 500 BC

1.1 Introduction

Particle physics is about trying to describe the most fundamental building blocks in the universe. The philosophical idea that there should be some fundamental building blocks is called *atomism* and goes all the way back to the ancient Greeks. In fact, the name "atom" comes from Greek, meaning "indivisible". What we call atoms today are composed of smaller particles, so the name "atom" is a bit of a misnomer. Nevertheless, the idea that there are *some* smallest indivisible building blocks seems to be correct. Particle physics is about describing the properties and behaviours of these building blocks.

Our current theory for particle physics is called the Standard Model (SM). In this theory, there are 17 types of *elementary particles* that make up everything we experience in our everyday lives. Twelve of these are the *fermions*, which make up all matter, and can be further classifed as six *quarks* and six *leptons*. Four of the elementary particles are called *gauge bosons* and are responsible for the three fundamental forces in the Standard Model[1]: the *photon* is associated with electromagnetism, the *gluon* with the strong nuclear force, and the W and Z bosons with the weak nuclear force. The last SM particle is the Higgs boson, which provides an elegant and consistent explanation for how fermions and W/Z bosons can have mass. For each particle type, there is also a corresponding antiparticle, which can be the particle itself.

While leptons can exist in a free form, quarks are always bound together with other quarks by the strong force in specific combinations called *hadrons*. In atomic physics, we encounter electrons, protons and neutrons. Of these, the electron is a lepton that is an elementary particle, while protons and neutrons are hadrons – they are composed of quarks and are not elementary. There are many other hadrons that are important when studying particle physics, but they decay after only a hundred millionth of a second or less, so they usually don't appear outside the field. This thesis is about interactions between such hadrons.

[1] A fourth fundamental force is gravity, but to date nobody has been able to describe it in terms of particle physics. There may also be other forces that we haven't discovered yet.

1.2 Phenomenology

The scientific method usually follows a set structure: a hypothesis is formulated, predictions are drawn based on this hypothesis, and experiments are performed to check whether the predictions are accurate. Consider for instance a collision between two electrons that scatter elastically at an angle θ. Since quantum mechanics is fundamentally random, it is not possible to predict the value of this angle in any single collision, but a good theory for particle physics must be able to predict the probability distribution function for this variable. This prediction can then be tested by repeatedly colliding electrons and measuring the actual distribution in the outgoing angles.

The most important experiments today involve colliding particles at very high energies. The most famous particle collider is undoubtedly the Large Hadron Collider (LHC) at CERN, which focuses on proton–proton collisions and collisions involving heavy ions. In principle, one can think of pp-collisions as having a fixed set of possible final states, but unlike the simple example of elastic electron–electron scattering, these final states can contain hundreds or even thousands of particles. Because the final state space is so huge, it is impossible to determine the probability distribution function, both through experiment and on a theoretical level. This makes it difficult to test the predictions of the theory.

Phenomenology is a branch of modern particle physics research whose purpose is to bridge the gap between theory and experiment. An important tool in phenomenology is software called *event generators*, that is used to simulate particle collisions. One example of such an event generator is Pythia, a program whose roots go back to the Jetset program developed in Lund in the late 70's, and which lies at the heart of this thesis [1, 2][2]. In a nutshell, Pythia attempts to simulate the processes that occur when two particles collide, in order to replicate the structure of events seen at for example the LHC. Since quantum mechanics is random, these programs also make random choices at several points during the simulation. Algorithms that rely on randomness are called *Monte Carlo* (MC) algorithms, and for this reason, event generators are often called *Monte Carlo generators* in the particle physics community.

With Pythia, we are able to simulate particle physics processes similar to the ones that occur at the LHC. But the events are still so complicated that it is impossible to directly compare the results of the simulation to the experimental results. Instead, we look at aggregate values called *observables*, such as the number of charged particles in an event, how energy is distributed between these particles, or the fraction of events that produce Higgs bosons. Pythia can be used to predict the distribution functions for these observables, and these distributions can then be measured experimentally and compared to predictions.

[2] Two other general-purpose event generators that are used today are Herwig [3] and Sherpa [4]

1.3 Open questions in particle physics

Even though the SM is well formulated on a theoretical level, there are still many open questions concerning the implications of the theory[3]. One particular question that is central to this thesis is what happens to quarks under extreme pressures and temperatures up to a 100,000 times hotter than the center of the sun. The current reigning paradigm is that a new state of matter called a *Quark–Gluon Plasma* (QGP) is formed, where hadrons dissolve and form a continuous medium of quarks and gluons[4]. It is believed that matter in the universe was in a QGP state during the first moments after the big bang, and that it may occur naturally in the core of neutron stars.

Some high energy physics experiments try to create a QGP in order to study its properties. The one with the highest energy to date is the ALICE experiment at the LHC. This experiment tries to create a QGP by colliding *heavy ions*, that is, the nuclei of heavy atoms such as gold or lead. If a QGP is formed, it is under such violent conditions that it expands and cools down and turns back into hadrons within a fraction of a billionth of a nanosecond. This makes it impossible to study directly, and the QGP hypothesis is instead based on a number of indirect phenomena, referred to as *QGP signatures*. Two phenomena that come up in Paper II of this thesis are *collective flow* and *jet quenching*.

Collective flow means that particles tend to move in the same direction as other particles, instead of in a random and uncorrelated fashion. In a liquid where particles are bound together, one would expect a lot of collective behaviour, while particles in a gas would have much less collectivity (imagine the difference between a wobbling droplet of water, compared to a dissipating droplet of vapor). When QGP was first theorised, the hypothesis was that it would behave more similar to a gas. It was therefore surprising when experiments observed collective flow, as it indicated the QGP was was actually more similar to a liquid.

Jets are energetic sprays of particles that are produced in high energy collisions. They are commonly produced in pairs that, because of momentum conservation, have similar but opposite momenta. If one of the jets travels through a dense medium such as a QGP, it may lose energy to the medium and have its shape altered – it becomes *quenched* – so that the two jets no longer have similar momenta. The first observations of jet quenching were at the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory in New York in 2002 and 2003 [5, 6].

Any scientific hypothesis must be falsifiable, and must be subjected to thorough scrutiny before we accept it to be accurate. One of the problems with the QGP hypothesis is that a QGP requires sufficiently high temperatures and densities to form. However, the signature

[3]Similar to how Maxwell's equations for electromagnetism were formulated in the mid 1800's, but their potential application to electronics were unimaginable at the time.

[4]Like ice cubes melting into water, except a trillion times more extreme.

QGP phenomena have been observed not only in collisions between two heavy ions, but also between a proton and a heavy ion, and even between two protons, where the conditions are not extreme enough. There is ongoing work to try to explain these observations within the QGP paradigm. Current research in Lund is taking a different approach, namely looking for alternative explanations for these observations that do not require a QGP. That is not to say that we necessarily do this because we don't believe that QGP exists, but rather that we wish to challenge the paradigm. Our objective is to determine which observations can have alternative explanations and which, if any, truly demand a QGP. This will then give us hints as to which experiments we should perform in the future to determine which model is correct.

The way we approach this is by using the Angantyr framework [7, 8], a part of Pythia that handles heavy ion collisions. In its basic form, Angantyr cannot produce the signature QGP observations, and the ongoing research therefore revolves around extending the framework with new physics effects that can produce these signatures. One example of such a physics effect is *hadronic rescattering*, which refers to secondary collisions between hadrons after they are produced in the primary event. Hadronic rescattering in Pythia is the most prominent application discussed in this thesis. It is developed in Paper I, while Paper II studies the physics implications of hadronic rescattering in Angantyr. We found that it gives rise to a rather substantial amount of collective flow, which in my opinion might be the most significant result of this thesis.

Moving on from QGP, another open question in particle physics concerns the nature of a class of hadrons called *exotic hadrons*. These are hadrons consisting of four or more quarks. While research on exotic hadrons is not as prominent as on QGP, studying their properties is still important for our overall understanding of the strong force. Hadronic rescattering is a possible mechanism for producing these exotic hadrons, and this is explored in Paper III.

A third topic that is currently being researched is cosmic rays. These are particles coming from space that hit the Earth's atmosphere. When this happens, it causes cascades of particles, which can be detected for example by detectors on the surface or on balloons. Some of these particles are extremely energetic, with the most famous experimental observation being the Oh-My-God particle detected in 1991 [9], which had an energy of about 3×10^{11} GeV. This is 50 million times more than the energy of a particle accelerated at the LHC, and is equivalent to a baseball traveling at nearly 100 km/h – in a *single particle*. One of several open questions concerning cosmic rays is how such particles can be produced.

As with high energy physics experiments, simulations of particle cascades are crucial for comparing our theoretical models to experimental observations. Such simulations need to describe how cascades propagate throughout the atmosphere and how particles in the cascade interact with particles in the air. One example of such a program is CORSIKA [10], which simulates the cascades, but relies on external models to describe the colliding

particles. An important feature of these cascades is that they can contain many different hadron types, and since PYTHIA has traditionally focused only on proton–proton (and a handful other) types of collisions, it has not been applicable to such scenarios. The goal of Paper IV is to extend PYTHIA to handle collisions involving other hadrons species, so that it can be used by cascade simulations like CORSIKA.

1.4 Quantum chromodynamics

Out of the three fundamental SM forces (the electromagnetic force, the strong force, and the weak force), all the work in this thesis concerns the strong force. The theory of the strong force is called Quantum Chromodynamics (QCD), and in this subsection I will introduce some of the main features of QCD. To do this, I will compare it to *Quantum Electrodynamics* (QED), which is the theory of electromagnetic interactions in the Standard Model.

Electromagnetic interactions occur between electrically charged particles. The SM particles that are electrically charged include the quarks, three of the leptons, and the W boson. The electromagnetic potential is proportional to $V_{\mathrm{EM}} \sim q_1 q_2 / r$, where q_1 and q_2 are the electric charges of the particles, and r is their separation distance. If the charges have opposite signs the force is attractive, otherwise it is repulsive. The $1/r$ factor in the potential means that the potential vanishes as the charges are infinitely far separated; in other words, only a finite amount of energy is required to move the particles arbitrarily far apart. Matter particles do not interact with each other directly, but rather indirectly via *force carriers*. For QED, the force carrier particle is called the *photon*. The properties of the force carriers are key to the behaviour of the force itself. For example, the photon is massless, and this is why the electromagnetic force can have infinite range.

The strong nuclear force is mediated by the *gluon*, and is experienced by quarks and the gluon itself. Like the photon, the gluon is massless, but it has other properties that make the strong force very different from the electromagnetic one. Similar to how electromagnetic interactions occur between particles with *electric* charge, the strong interactions are between particles with *color* charge. This is not to be confused with the visible color of the particles, but the reason behind this name is indeed due to an analogy with color theory. In electrodynamics, there is only one type of charge (which can be positive or negative), but in QCD there are three types, which we label r, g and b. If you combine two objects with opposite charges, say one object with $+1$ unit of r charge and another with -1 unit, the overall combination becomes neutral. However, if you combine the three types in equal amounts, the result also becomes neutral[5]. This is analogous to how in color theory combining red, green and blue light gives white, and hence the labels indicate the "primary

[5]Technically, the quantum states that give color singlets are $\delta_{ij} |i\rangle |\bar{j}\rangle$ for mesons and $\epsilon_{ijk} |i\rangle |j\rangle |k\rangle$ for baryons

colors" red, blue and green.

As mentioned, the gluon itself experiences the strong force. This is because the gluon carries color charge, in contrast with the photon, which carries no electric charge and therefore does not itself experience electromagnetic forces. This crucial fact makes QCD very different from QED. When two color charges are separated from each other and a color field is formed, the self-interactions of this color field therefore produce a much higher potential than the EM potential in QED. Specifically, the QCD potential contains a term that is proportional to r that grows larger the further apart the particles are. If two color charges are pulled sufficiently far apart, the energy in the color field is eventually converted to mass by creating new particles, in accordance with Einstein's $E = mc^2$. Hence, you can never pull a color charge away from other charges and study it individually. This phenomenon is called *confinement*.

Quarks carry $+1$ unit color, antiquarks carry -1 unit, and gluons carry a combination of color and anticolor. The color type carried by a quark is not fixed by the quark species, e.g. a u quark can carry either red, blue, or green charge, or a superposition of the three[6]. Confinement implies that quarks cannot appear freely, but only in color-neutral combinations called *hadrons*. For quarks, there are two obvious ways to achieve this: either combine a quark and antiquark with opposite colors, or combine three quarks with different colors. These combinations are referred to as *mesons* and *baryons*, respectively. Hadrons are in a sense "quark molecules", with the most common ones being the proton (which consists of two up-quarks and one down-quark) and the neutron (which is one u and two d's). Other combinations of quarks are also possible, such as two quarks and two antiquarks (resulting in a *tetraquark*), or four quarks and one antiquark (a *pentaquark*). Such combinations are called *exotic hadrons* and are far less common than mesons and baryons. Exotic hadrons are studied in Paper III.

2 QCD processes in Pythia

This thesis revolves around QCD interactions, and the product of the work is in the form of new code in the Pythia event generator[7]. In this section, we discuss the basic concepts of QCD interactions that are relevant to this thesis, emphasizing how they are implemented in Pythia.

A standard Pythia event is generated in three steps: process level, parton level and hadron

[6]Thus in a sense, there are three variations of each quark, e.g. a red u, a blue u and a green u. Whether these should be thought of as separate particle species or simply as different states of the same particle is a matter of definition.

[7]At the time of writing, the code developed in Paper III and Paper IV has not yet been made publicly available.

level. The process level involves selecting partons from the incoming hadrons based on their parton distribution functions (PDFs) and simulating the initial process. In PYTHIA, processes are categorized as either "soft QCD" or "hard QCD". While "hardness" normally refers to the scale of the momentum transfers involved, in this context these terms refer only to the way the event is modelled. It is possible for a soft process to involve a higher momentum transfer than a hard one. To add to the confusion, the initial process is often referred to as "the hard process", as it is usually the hardest part of the event, regardless of how it is modelled.

The parton level gives corrections to the hard process, and includes multiparton interactions, beam remnants, initial- and final state radiation, and color reconnection. Finally, the hadron level deals with hadronization, as well as post-hadronization effects such as decays and hadronic rescattering.

2.1 Scattering processes

From a very general perspective, a quantum mechanical process can be described in terms of an initial state $|i\rangle$ at $t \to -\infty$ and an evolution operator S (the S-matrix) which brings the system to a final state $S|i\rangle$. The elements of S are given by $S_{fi} = \langle f|S|i\rangle$, where $\langle f|$ are the possible final states at $t \to \infty$. In the kind of scattering processes considered in this thesis, $|i\rangle$ corresponds to a two-particle state, while $\langle f|$ can be a more general multiparticle state.

It is common to write $S = 1 + iT$, where the 1 corresponds to processes where there is no interaction. Then define the amplitude $\mathcal{M}$ by

$$\langle f|iT|i\rangle = i\mathcal{M}(i \to f)(2\pi)^4\delta^{(4)}\left(p_A + p_B - \sum k_j\right), \tag{1}$$

where p_A and p_B are the momenta of the incoming particles, and k_j are the momenta for each outgoing particle. The differential (with respect to the outgoing momenta) cross section of the process can be written as

$$d\sigma = \frac{1}{2E_A 2E_B|v_A - v_B|}\prod_j\left(\frac{d^3k_j}{(2\pi)^3(2E_j)}\right)(2\pi)^4\delta^{(4)}\left(p_A + p_B - \sum k_j\right)|\mathcal{M}|^2, \tag{2}$$

where E and v denote the energy and velocity of each particle, respectively (the details of this derivation are given in [11], pp. 99-106). In this expression, $\mathcal{M}$ captures the dynamical behaviour, i.e. the dependency on the details of the interaction Hamiltonian. Thus, calculating the cross section becomes a matter of finding $\mathcal{M}$ using the framework of quantum field theory, which is usually expressed in terms of Feynman diagrams like the ones shown in Figure 1.

Figure 1: Feynman diagrams for some QCD processes.

The first complication we encounter when we apply this to QCD is that due to confinement, the initial and final states have to consist of hadrons, while the process itself should be modelled in terms of the constituent partons (quarks and gluons). We therefore need a procedure to convert a hadron into partons and then back into hadrons after the interaction. *Parton Distribution Functions* (PDFs) are used to describe the parton contents of hadrons. A PDF $f_q^h(x, Q^2)$ is interpreted as the rate of finding a particular parton q in a hadron h. It depends on two parameters: x, the fraction of the hadron's total momentum carried by the parton; and Q^2, the virtuality of the interaction (these will be explained in more detail in Section 4). For two hadron species A and B, the lowest order total cross section can be written in terms of the PDFs as

$$\frac{d\sigma_{AB}(s)}{d\hat{t}} = \sum_{i,j} \int_0^1 dx \int_0^1 dx' \, f_i^A(x, Q^2) f_j^B(x', Q^2) \, \frac{d\hat{\sigma}_{ij}(\hat{s})}{d\hat{t}}, \tag{3}$$

where $\hat{\sigma}_{ij}$ is the cross section for the hard process on the parton level, and the sum is over all parton species combinations i and j, including antiquarks and gluons. Here, s refers to the Mandelstam variable for the full process, while the hatted $\hat{s} = xx's$ is for the parton-level process. The virtuality can be defined in different ways, but in this context a common choice is $Q^2 = -\hat{t}$. The specific process to simulate is chosen with probabilities proportional to the cross section for each process. In addition to selecting the process, the other important variable that must be sampled is the momentum transfer $\hat{t}$.

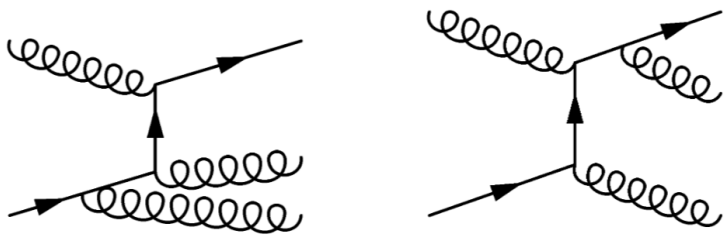
After partons have been produced, the process of converting them into hadrons is called *hadronization*, which we will get back to when we discuss the Lund string model in Section 5. For now, note that while (3) shows that the $AB \to$ partons cross section is directly dependent on PDFs, the hadronization process is largely independent of how the partons were produced. Therefore, we can comfortably talk about final partonic states here without worrying about how the eventual hadronization will happen.

2.2 Hard QCD processes

Hard QCD processes in PYTHIA are directly expressed in terms of Feynman diagrams like the ones shown in Figure 1. When simulating hard processes, a minimum momentum transfer $p_{\perp,\min}$ needs to be specified. This must not be too low, as hard processes use divergent perturbative QCD cross sections, and a lower $p_\perp$ bound serves as a way to regularize these.

The primary use case for hard processes is jet production. Jets are highly energetic sprays of particles, and are therefore often correlated with high energy phenomena such as those involving the Higgs boson or top quarks (with masses around 125 and 173 GeV, respectively). For this reason, jets are often the interesting part of the event, while the rest of the event consisting of lower energy particles is called the *underlying event*, and is treated as noise in some contexts. In order to make predictions for the shape of jets, PYTHIA must generate a large number of such events. Thus in practice, $p_{\perp,\min}$ can be used to make this generation more efficient by generating only jets with large enough $p_\perp$ for the study in question.

There may be additional gluons produced, such as in the diagrams



This is called *initial-* and *final state radiation* (ISR/FSR), depending on whether it happened before or after the hard interaction, and is handled on the parton level in PYTHIA. FSR in particular has a large influence on the substructure of jets. One issue encountered is that the rate of emitting an additional gluon with a momentum fraction x rises to infinity in the limit $x \to 0$. That is, if you consider sufficiently soft gluons, there will always be another one emitted, so the number of gluons emitted is technically infinite. In practice, whether we can resolve these gluons depends on the resolution of our experiment. Furthermore, very soft gluons will be absorbed by the hadronization process, and will have little overall effect on the observable event.

2.3 Soft QCD processes

In other studies, we may be interested in more general events without demanding trigger conditions such as a sufficiently hard jet. Such events are called *minimum bias* events.

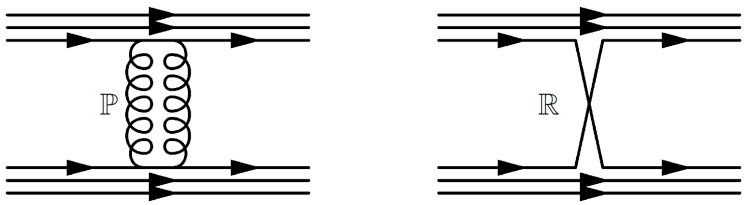


Figure 2: Pomeron and reggeon mediators

Now, since we wish to generate events with no $p_{\perp,\min}$ condition, approaching the process in terms of diagrams like the ones in Figure 1 is unfeasible, and PYTHIA instead takes an approach based on Regge theory. In PYTHIA, processes modelled this way are called soft QCD processes.

Regge theory is an effective field theory where the force carriers are pomerons $\mathbb{P}$ and reggeons $\mathbb{R}$. These are effectively treated as mediator particles, where pomerons are analogous to glueballs while reggeons can be thought of as analogous to mesonic propagators. For example, diagrams for elastic scattering processes with a pomeron $\mathbb{P}$ and reggeon $\mathbb{R}$ in the t-channel can be drawn as shown in Figure 2[8]. But in the effective theory we think of these mediators as particles in their own right. The cross sections for these processes, both total and partial, are calculated in terms of this effective theory, and we will discuss the details of this in the next section. For simplicity, going forward we will refer to the mediator particle only as a "pomeron", even though similar ideas apply to the reggeon.

Possible diagrams using pomerons are shown in Figure 3, but instead of calculating the amplitudes of diagrams like we did above, these diagrams correspond to square amplitudes $|\mathcal{M}|^2$. The amplitude-level processes are then described in terms of *cut rules*, where the diagram is cut along the middle (indicated by dashed lines in Figure 3). If a pomeron is cut, this corresponds to a gluon color octet on the amplitude level. In other words, an uncut pomeron interacting with a hadron corresponds to an elastic momentum transfer (Figure 3a), while a cut pomeron corresponds to formation of diffractive systems, which in PYTHIA are handled using the Lund string model.

In the inelastic case, the process is classified according to the string topology it produces. If the string is stretched between the two hadrons (Figure 3b), it is called non-diffractive. Otherwise it is called diffractive, and can be further categorized as single diffractive (SD), double diffractive (DD), or central diffractive (CD), as shown in Figure 3c, d and e, respectively. More complicated diffractive topologies are possible, but their contribution is negligible in most situations. The gaps left by uncut pomerons in diffractive processes are

[8]To be more technical, these particles are called *Regge trajectories*, and are basically sums of meson propagators. For example, the reggeon in Figure 2 is more accurately thought of as a sum of mesons $\rho^0 + \mathrm{f}_2(1270) + \rho_3(1700) + \cdots$.

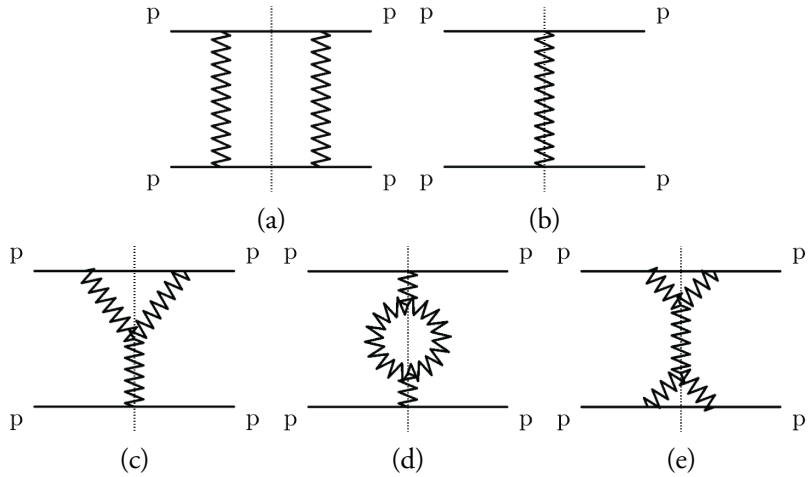


Figure 3: Diffractive diagrams for a) elastic, b) non-diffractive, c) single diffractive, d) double diffractive, and e) central diffractive processes. The dotted lines indicate cuts.

usually visible experimentally in the rapidity spectrum of the produced particles. In practice, it is possible that other phenomena such as color reconnection can modify the color topology, so that processes generated in a particular Pythia process may be interpreted differently on an experimental level.

There are several ways to improve the accuracy of this model, the most straightforward being including higher order diagrams. Another approach due to Ingelman and Schlein [12] is to take a more detailed view of the internal structure of the pomeron, endowing it with a parton distribution fuction. This then views the formation of a diffractive system as being initiated by a $\mathbb{P}$p collision/absorption. A third option is to include other propagator particles such as the *odderon*, which is similar to a pomeron, but with odd charge parity. For a more complete discussion of the various models in Pythia, see [13].

2.4 Multiparton interactions

Parton distribution functions are crucial both to hard QCD processes, where they determine which partons interact, and to soft QCD processes, where they determine the contents of the diffractive and nondiffractive systems. In both cases, the total content of a hadron might include several further partons, and these can interact with each other in additional subprocesses called *multiparton interactions* (MPIs). Such interactions are modelled similarly to soft QCD processes. In hard QCD processes, it is also possible to specify a second hard process, demanding that the first MPI must also have a $p_{\perp,\min}$ and modelling the interaction using hard QCD.

In Pythia, the MPI framework defines an *overlap function* that determines the rate of interactions as a function of the impact parameter b. Placing two particles A and B at $x = \pm b/2$ and moving in the z direction, this overlap is

$$\mathcal{O}(x, y) \propto \int dz_A \, P_A(x - b/2, y, z_A) \int dz_B \, P_B(x + b/2, y, z_B), \tag{4}$$

where $P_{A,B}$ are the probability distributions for the location of partons within each hadron. If these distributions are narrow, one expects to get many events with a few MPIs and some events with very many, while if they are wider, the number of MPIs has less variance. With the development of the spacetime picture in [14], these overlap functions are also used to define the location of parton interactions. This is explained in more detail in Section 2.2 of Paper I.

The parts of the hadrons that do not interact in MPIs are called *beam remnants*. Even though these do not interact directly and are produced at high rapidities so that they are usually not detected in experiment, they have some overall effect on the event as they are color-connected to the MPIs.

2.5 Heavy ion collisions

Pythia includes an integrated framework for heavy ions called Angantyr [8]. This framework is inspired by the old Fritiof program [15], and is based on the wounded nucleon model [16]. Collisions in Angantyr are handled using Pythia's framework for multiparton interactions, which allows for sophisticated diffractive topologies and color connection between individual wounded nucleons. These collisions result in strings being formed that hadronize using the Lund string model, which is radically different from the hydrodynamic picture of quark–gluon plasma.

With only its basic features, Angantyr is not able to produce QGP–like signals. Even though Angantyr produces anisotropic overlap regions in space, these regions will not produce collective flow in momentum unless there are further interactions between the strings or hadrons. Hadronic rescattering implements such interactions, and the idea that rescattering can give rise to such effects was perhaps the largest incentive for developing the framework in the first place. Indeed, as we will see in Paper II, rescattering does produce a significant amount of flow, but not enough to fully explain data. This is expected, as other phenomena such as ropes and string shoving (see Section 5.3) should also contribute.

3 Cross sections

In this section, we discuss the total and partial cross sections for the minimum bias processes discussed in the previous section. Unlike hard QCD processes, which specify a hard cutoff $p_{\perp,\min}$ so that the total cross section can be found simply by integration, minimum bias processes need to provide integrated cross sections that are independent of such a cutoff.

The total cross section for a particular beam combination at a specified CM energy $\sqrt{s}$ is related to the event rate at particle accelerators. In default Pythia, each call to the generator always produces one event, and so the total cross section has no impact on this part of generation. As such, in this context, the total cross section is only useful for predicting the rate of events in experiments[9]. However, this rate of events has implications for both hadronic rescattering and for cosmic rays, which we will come back to when we discuss applications in Section 6. Partial cross sections are still used directly in order to pick which process to simulate when there are several possible.

The way an event is classified in event generators can sometimes different from how it gets classified experimentally. For example, in our low-energy framework, we implement formation of resonances, which can then decay, e.g. $\pi^+\pi^- \to \rho^0 \to \pi^+\pi^-$. In the context of Pythia, this process is different from an elastic scattering based on t-channel exchange, and the outgoing particles will have a different t distribution. However, in experiment, only the final state particles are observable, and both these processes would be classified as elastic. We will refer to processes such as these as *pseudo-elastic*. In some contexts this is taken even further, considering processes like $\pi^+\pi^- \to \mathrm{f}_0(500) \to \pi^0\pi^0$ to also be elastic. Similarly, phenomena such as color reconnection can leave mass gaps in a non-diffractive event, leading it to be interpreted as a diffractive event (a "pseudo-diffractive" event, if you will). These technicalities are important to be aware of, as the parameters that determine partial cross sections are fitted to data. If the rate of, say, directly elastic interactions are fitted to exactly match data, then the contribution from pseudo-elastic processes may lead to an incorrect elastic cross section in practice. As phenomenologists it is our job to understand these issues and bridge those gaps.

3.1 Total cross sections

Total cross sections for minimum bias processes in Pythia at high energies are to first order described by the Donnachie-Landshoff (DL) model [17], which writes the total cross section as the sum of two diagrams, one corresponding to pomerons and one to reggeons. This gives a cross section on the form

$$\sigma_{AB} = X^{AB} s^{\epsilon} + Y^{AB} s^{-\eta}, \tag{5}$$

[9] Though it is used indirectly to calculate non-diffractive cross sections.

where $\epsilon \approx 0.0808$ and $\eta \approx 0.4525$. The s^{ϵ} term corresponds to pomerons and grows with energy, while $s^{-\eta}$ is the effective reggeon term and diminishes at higher energies. The X and Y coefficients depend on the incoming hadron species. Since the pomeron couples the same way to a hadron and its antihadron, X is symmetric when replacing either A or B with its antiparticle. In Paper IV we aim to describe interactions where general hadron species interact with the atmosphere, and such a description needs to specify X and Y for all hadron–proton combinations.

At very low energies, total cross sections usually deviate from the high energy parameterizations, and may have complicated shapes that are difficult to parameterize with simple expressions, especially due to the presence of resonance particles (described below). The DL model is accurate down to about 6 GeV, but in our framework we consider 10 GeV to be the boundary between "low" and "high" energy. There are numerous low energy cases that need a unique treatment, but due to time constraints, we have in our framework focused only on the most common particle interactions and cases where experimental data is available, at the expense of using simplified and generic models in less common cases. Here, we present a few examples to illustrate the kind of ideas that go into this modelling, while a full description is given in Paper I.

For pp collisions, we fit the total and elastic cross sections to data, as shown in Figure 4a). There are a few insights to be gained from this plot. First, below $\sqrt{s} = 2m_{\mathrm{p}} + m_{\pi} \approx 2.1$ GeV, there is not enough energy to produce new hadrons, and hence the elastic cross section is equal to the total one below this threshold[10]. Second, the cross section diverges as the kinetic energy approaches zero. This is due to electromagnetic interactions that are negligible at higher energies where QCD dominates. In principle one could argue that the cross section should grow to infinity as $\sqrt{s} \to 2m_{\mathrm{p}}$, since electromagnetic interactions have infinite range. However, if the particles are very far away, the interaction between them would result in only a very small momentum transfer, so in practice those interactions do occur but are negligible[11]. Above 5 GeV, the cross section is parameterized according to the HPR_1R_2 parameterization [18],

$$\sigma_{\mathrm{total}} = P + H\log^2\left(\frac{s}{s_0}\right) + R_1\left(\frac{s}{s_0}\right)^{-\eta_1} + R_2\left(\frac{s}{s_0}\right)^{-\eta_2} . \tag{6}$$

where P, R_1 and R_2 depend on the specific particle species, H, η_1 and η_2 are constants, and $s_0 = (m_A + m_B + M)^2$ where M also is a constant. The P and H terms are associated with pomerons, and the R terms with reggeons. Specifically, R_1 and R_2 correspond to reggeons with even and odd charge parity, respectively, so replacing one hadron with its antiparticle will result in the same value for R_1, while the sign is flipped on R_2. We parameterize other common cross section using HPR_1R_2 when available.

[10] Not accounting for QED effects such as bremsstrahlung

[11] In Pythia, we handle interactions as independent of the impact parameter, and we therefore put a cutoff on the interaction range in order to suppress these negligible interactions.

In processes involving mesons, resonance production is possible (details described below) and are the dominant processes at very low energies, leading to complicated behaviour for the total cross sections, as shown in Figure 4b. When explicit resonances have been implemented, the total cross section is found by summing the individual contributions from each resonance, possibly also adding a purely elastic cross section. Two other cases that are very common and therefore require an even more precise handling are $\pi\pi$ and $\mathrm{K}\pi$. In this case, we rely on parameterizations by Pelaez et al. [19–21].

For most processes involving less common hadrons, in particular those involving heavy flavors, there is no data and no well-established theory for low energy cross sections. In these cases, the additive quark model (AQM) is the only available option [22, 23]. AQM is a very generic model for total cross sections, which depends only on the quark content of the hadrons, and is given by

$$\sigma_{\mathrm{AQM}} = (40\ \mathrm{mb})\frac{n_{\mathrm{eff},A}}{3}\frac{n_{\mathrm{eff},B}}{3}, \tag{7}$$

where n_{eff} is the effective number of quarks in each hadron, by default given by

$$n_{\mathrm{eff}} = n_{\mathrm{u}} + n_{\mathrm{d}} + 0.6n_{\mathrm{s}} + 0.2n_{\mathrm{c}} + 0.07n_{\mathrm{b}}, \tag{8}$$

and n_{q} is the number of each quark type in the hadron. In its original form, AQM did not consider charm or bottom hadrons, and the n_{c} and n_{b} terms were added by us in Paper I. In addition to directly giving us some low energy cross sections, AQM also plays a more important role in generalizing cross sections by rescaling parameters. For example, the total cross section for $\mathrm{p\bar{p}}$ is given by a parameterization by UrQMD [24]. To generalize this to $A\bar{B}$ collisions, we first evaluate the $\mathrm{p\bar{p}}$ cross section at the effective energy that corresponds to the energy of a $\mathrm{p\bar{p}}$ system with the same CM momentum as the $A\bar{B}$ system,

$$s_{\mathrm{eff}}^{A\bar{B}}(s) = 4m_{\mathrm{p}}^2 + \frac{1}{s}\left(s - (m_A + m_B)^2\right)\left(s - (m_A - m_B)^2\right), \tag{9}$$

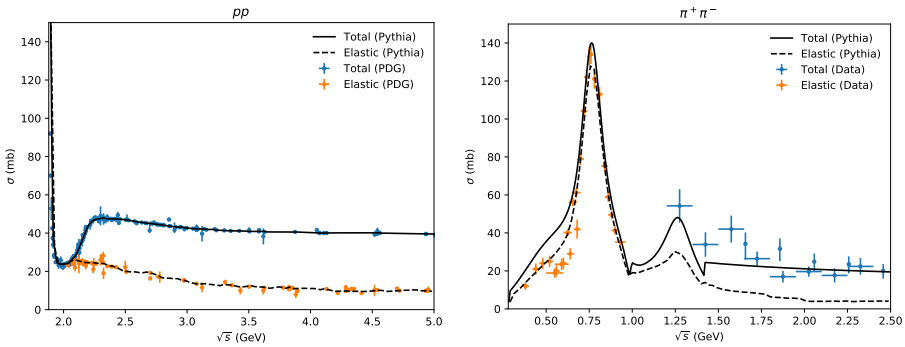
then rescale by an overall AQM factor, i.e. the resulting cross section is

$$\sigma_{\mathrm{ann}}^{A\bar{B}}(s) = \frac{\sigma_{\mathrm{AQM}}^{A\bar{B}}}{\sigma_{\mathrm{AQM}}^{\mathrm{p\bar{p}}}}\sigma_{\mathrm{ann}}^{\mathrm{pp}}(s_{\mathrm{eff}}^{A\bar{B}}). \tag{10}$$

The AQM is used similarly in many other contexts, as we will see below.

3.2 Elastic and diffractive cross sections

In elastic interactions, the only parameter that must be sampled is the momentum transfer t. In diffractive processes, also the invariant masses M of each diffractive system must be



(a) (b)

Figure 4: Elastic and total cross sections for a) pp collisions, with data from PDG [18], and b) $\pi^+\pi^-$ collisions, with data from [25, 26] (elastic) and [27, 28] (total). Note that since we are comparing to data, the elastic cross section includes pseudo-elastic scattering. In (b), the elastic cross section is almost completely dominated by pseudo-elastic scattering through resonances

determined. The default elastic and diffractive cross sections in Pythia are given by the Schuler-Sjöstrand (SaS) parameterization [29],

$$\frac{d\sigma_{\text{el}}}{dt} = (1+\rho^2)\frac{\sigma^2_{\text{tot}}(s)}{16\pi}e^{B_{\text{el}}(s)t}, \tag{11}$$

$$\frac{d\sigma_{XB}(s)}{dtdM_X^2} = \frac{g_{3\mathbb{P}}}{16\pi}\frac{\beta_{A\mathbb{P}}(s)\beta^2_{B\mathbb{P}}(s)}{M_X^2}e^{B_{XB}(s)t}F_{\text{SD}}(M_X^2,t), \tag{12}$$

$$\frac{d\sigma_{XY}(s)}{dtdM_X^2dM_Y^2} = \frac{g^2_{3\mathbb{P}}}{16\pi}\frac{\beta_{A\mathbb{P}}(s)\beta_{B\mathbb{P}}(s)}{M_X^2M_Y^2}e^{B_{XY}(s)t}F_{\text{DD}}(M_X^2,M_Y^2,t), \tag{13}$$

where $g_{3\mathbb{P}}$ corresponds to the coupling strength of the triple pomeron vertex, each β is a hadron–pomeron coupling strength. X and Y indicate diffractive systems, not to be confused with X^{AB} and Y^{AB} in (5). The non-diffractive cross section is given by what is left over when the partial cross sections for all other processes have been subtracted from the total one. The slope parameters are

$$\begin{aligned} B_{\text{el}}(s) &= 2b_A + 2b_B + 4s^\epsilon - 4.2, \\ B_{XB}(s) &= 2b_B + 2\alpha'_{\mathbb{P}}\log\frac{s}{M_X^2}, \\ B_{XY}(s) &= 2\alpha'_{\mathbb{P}}\log\left(e^4 + \frac{s/\alpha'_{\mathbb{P}}}{M_X^2M_Y^2}\right), \end{aligned} \tag{14}$$

where $b = 2.3$ for light (i.e. containing no charm or bottom quarks) baryons and 1.4 for light mesons, $\alpha'_{\mathbb{P}} = 0.25$ GeV^{-2} is the slope of the Regge trajectory, and the e^4 term is

introduced to stop B_{XY} from blowing up at large $M_X^2 M_Y^2$ (e denotes Euler's number). The b terms lead to a narrower t distribution for elastic and SD processes. The interpretation is that they are a measure of how easy the hadron is to excite to a diffractive system, and are associated with suppression of elastic processes with large t, as larger momentum transfers are more likely to excite the hadrons. In the low energy cases where the total cross section is given by AQM, the elastic cross section is

$$\sigma_{\mathrm{AQM,el}} = 0.039\sigma_{\mathrm{AQM}}^{3/2}, \tag{15}$$

based on the optical theorem. The differential distribution has the same t-dependency as the high energy case.

For heavy flavour hadrons, the β and b coefficients are rescaled by AQM factors, but $g_{3\mathbb{P}}$ and $\alpha'_{\mathbb{P}}$ are not, as they depend only on properties of the pomeron. The F factors in (12) and (13) are fudge factors related to low energy resonances, and are given by

$$\begin{aligned} F_{\mathrm{SD}}(M_X^2, s) &= \left(1 - \frac{M_X^2}{s}\right)\left(1 + \frac{c_{\mathrm{res}} M_{\mathrm{res}}^2}{M_{\mathrm{res}}^2 + M_X^2}\right), \\ F_{\mathrm{DD}}(M_X^2, M_Y^2, s) &= \left(1 - \frac{(M_X + M_Y)^2}{s}\right)\left(1 + \frac{s m_{\mathrm{p}}^2}{s m_{\mathrm{p}}^2 + M_X^2 M_Y^2}\right) \\ &\quad \times \left(1 + \frac{c_{\mathrm{res}} M_{\mathrm{res}}^2}{M_{\mathrm{res}}^2 + M_X^2}\right)\left(1 + \frac{c_{\mathrm{res}} M_{\mathrm{res}}^2}{M_{\mathrm{res}}^2 + M_Y^2}\right), \end{aligned} \tag{16}$$

where $c_{\mathrm{res}} = 2$ and $M_{\mathrm{res}} = 2$ for pp and $\mathrm{p\bar{p}}$ collisions.

At low energies, below $E_{\mathrm{CM,min}} = 10$ GeV, the integrated diffractive cross sections use a different parameterization. The formation of a diffractive system $A \to X$ is forbidden below a threshold $M_{\mathrm{min}} = m_A + 2m_\pi$, and a smooth transition from a vanishing cross section at threshold to the high energy cross section is constructed by the parameterizations

$$\sigma_{XB}(E_{\mathrm{CM}}) = \sigma_{XB}(E_{\mathrm{CM,min}})\left(\frac{E_{\mathrm{CM}} - M_{X,\mathrm{min}} - m_B}{E_{\mathrm{CM,min}} - M_{X,\mathrm{min}} - m_B}\right)^{0.6} \tag{17}$$

for single diffractive and

$$\sigma_{XY}(E_{\mathrm{CM}}) = \sigma_{XY}(E_{\mathrm{CM,min}})\left(\frac{E_{\mathrm{CM}} - M_{X,\mathrm{min}} - M_{Y,\mathrm{min}}}{E_{\mathrm{CM,min}} - M_{X,\mathrm{min}} - M_{Y,\mathrm{min}}}\right)^{1.5} \tag{18}$$

for double diffractive.

3.3 Resonance formation

One of the most salient features of low energy processes that is not found at high energy is the formation of resonance particles. As we just discussed, high energy collisions can

lead to hadrons becoming excited to form diffractive systems. In a low energy regime, it is also possible that these excitations result in resonance particles such as a proton becoming excited to a Δ^+ or $\mathrm{p}(1520)$. In Pythia, such diffractive resonances are implemented explicitly for nucleon–nucleon collisions, and the partial cross section for these processes are based on parameterizations by UrQMD [24]. The fudge factors in (12) and (13) are included to compensate for the lack of such explicit resonances for other processes. This means that they are not included in the low energy nucleon–nucleon cross sections.

Collisions can also result in a single resonance particle if the CM energy of the collision is near the mass of that particle. For example, a $\mathrm{p}\pi^+$ scattering near 1.2 GeV can form a Δ^{++} resonance. The cross section for the process $AB \to R$ is given by a non-relativistic Breit–Wigner,

$$\sigma(AB \to R) = |\mathcal{M}|^2 \Phi \sim \frac{\Gamma_R \Gamma_{R\to AB}}{(m_R - \sqrt{s})^2 + \frac{1}{4}\Gamma_R^2}, \tag{19}$$

where Φ is a phase space factor, m_R is the resonance mass of R, and Γ_R and $\Gamma_{R\to AB}$ are the total and partial widths of R. These widths are mass dependent, as described in Paper I. One exceptional resonance is the $f_0(500)$, which can form in $\pi\pi$ collisions. This particle has a width that is larger than its mass, and (19) does not describe its formation cross section well. Instead, we use a parameterization from Pelaez et al.

In addition to forming hadron resonances, particles can theoretically also combine to other bound states, for example the formation of a deuteron in the process $\mathrm{pn} \to {}^2\mathrm{H}$. It is assumed that the cross section for the formation of such particles also follows the Breit–Wigner form of (19). In Paper III, we explore this as a mechanism for producing exotic hadrons.

An important point to keep in mind is that resonances formed in rescattering may not be easily detectable experimentally. Consider for example a ρ^+, whose decays are dominated by the channel $\rho^+ \to \pi^+\pi^0$. The ρ^+ appears experimentally as a bump around 0.77 GeV in the $\pi^+\pi^0$ invariant mass spectrum. If a $\pi^+\pi^0$ pair collides during hadronic rescattering, they can form a ρ^+ resonance, but only if they are already correlated. When that ρ^+ then decays, the outgoing $\pi^+\pi^0$ will have the same invariant mass as the original pair. However, if a particle has several decay channels, rescattering between two particles can change the mass spectrum of other particle species. As with pseudo-elastic scattering, this is another example of a situation where the theoretical viewpoint cannot be easily translated to experiment.

3.4 Annihilation processes

Another type of process that occurs at low energy, but has no significant high energy equivalent, is *baryon annihilation*. By this we refer to baryon–antibaryon interactions where

the baryon number is annihilated. Specifically, one or two quark–antiquark pairs are annihilated, creating string systems between the remaining pairs. As such, it can only happen between baryons that share at least one quark type in common, e.g. it is not possible for Δ^{++} and $\bar{\Sigma}^{-}$, which have quark content uuu and $\bar{\mathrm{s}}\bar{\mathrm{d}}\bar{\mathrm{d}}$, respectively. While baryon annihilation processes are relatively rare compared to other low energy processes during rescattering, they can have a significant effect on the composition of an event, in particular by giving rise to a substantial amount of nucleon depletion.

The cross section for annihilation is given by a parameterization by Koch and Dover [30],

$$\sigma_{\mathrm{ann}} = (120\ \mathrm{mb})\frac{s_0}{s}\left(\frac{A^2 s_0}{(s-s_0)^2 + A^2 s_0} + 0.6\right), \tag{20}$$

where $s_0 = 4m_{\mathrm{p}}^2$ and $A = 0.05$ GeV. For $\mathrm{p\bar{p}}$ below $2m_{\mathrm{p}} + m_\pi$, the whole inelastic cross section corresponds to annihilation, and is instead given by $\sigma_{\mathrm{ann}} = \sigma_{\mathrm{total}} - \sigma_{\mathrm{elastic}}$. This cross section is generalized to other beam combination using AQM factors in the same way as for the total baryon–antibaryon cross section.

In our model, we simulate baryon annihilation as follows. We start by selecting one quark–antiquark pair from all possible pairs, with uniform probability. If it is possible to pick more than one pair, we have a fixed probability (0.2 by default) of doing so. In principle one could imagine a process where all three pairs are annihilated, but this is unlikely to happen and it is not clear how to model it; if the modelling involves recreating a quark–antiquark pair to get a string, it is not very different from annihilating two pairs, anyway. If two pairs remain, they always form two strings; they are not allowed to form a single string between a diquark–antidiquark pair, since this would fragment into a baryon–antibaryon pair and it would no longer be an annihilation process.

4 Parton distribution functions

At perturbative energies, interactions involving hadrons must be modelled in terms of quark interactions. It is then necessary to describe the specific quark content of a hadron. While hadrons have some net quark content called the *valence quarks* (e.g. 2u + 1d for the proton), they can also contain other partons due to quantum fluctuations. Thus, the picture of a proton consisting of two ups and one down quark is a simplification.

The parton content of a hadron is given by *parton distribution functions* (PDFs). A PDF $f_q^h(x, Q^2)$ is interpreted as the rate of finding a parton q in a particular hadron h (e.g. $f_{\mathrm{u}}^{\mathrm{p}}$ denotes the u-quark content of the proton). It depends on two parameters: first, the momentum fraction x is the longitudinal momentum of the quark divided by the hadron's

total momentum[12]. The second parameter Q^2 is called the *virtuality*. It is a measure of the energy scale used to probe the hadron. A higher virtuality implies a higher "resolution" of the hadron contents, analogously to how objects cannot be clearly resolved by light with a wavelength that is larger than the size of the object. We will see a more technical definition shortly.

There are two important relations involving PDFs. The first constrains the valence content of a hadron,

$$\int_0^1 dx \left(f_q - f_{\bar{q}}\right) = n_{\mathrm{q}}, \tag{21}$$

where n_{q} is the number of valence quarks of that particular type in the hadron. Note that while the summed *valence* content is fixed for a particular particle species, the *total* quark content is not constrained the same way, so there is no straightforward constraint on $\int dx f_q$. It is thus convenient to define the valence distribution $v_q = f_q - f_{\bar{q}}$ for each quark species. We also define the sea distribution to be the difference between the total and the valence contents $s_q = f_q - v_q$. The sea must equal the $\bar{\mathrm{q}}$ content, since all quarks are produced as quark–antiquark pairs in QCD.

A second constraint is that the summed momentum fractions must equal to one,

$$\int_0^1 dx\, x \sum_q f_q(x, Q^2) = 1, \tag{22}$$

the sum is over all parton species, including antiquarks and gluons. For each valence quark, we also define the mean momentum fraction

$$\langle x \rangle_q = \frac{\int_0^1 dx\, x f_q(x, Q^2)}{\int_0^1 dx\, f_q(x, Q^2)}. \tag{23}$$

Valence quarks must have roughly the same velocity, since otherwise the hadron would not stay together over time. Therefore, heavier quarks tend to have higher mean momentum fractions.

4.1 Deep inelastic scattering

We have now discussed the basics of PDFs and their interpretation as probability distributions for parton content, but we have not yet given a formal definition in terms of meas-

[12]Here, it is assumed the quark has no transverse momentum. There is currently research being done on *generalized* PDFs that also consider features such as transverse momentum, but such higher order effects are outside the scope of this work.

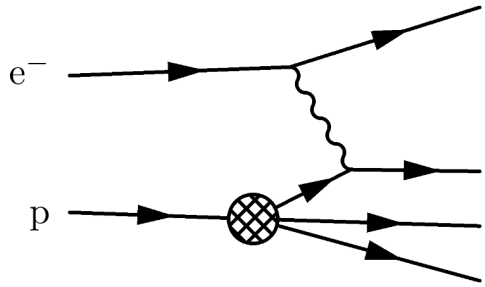


Figure 5: Deep inelastic scattering

urable physics. To see how they appear in practice, we consider the $\mathrm{e}^-\mathrm{p}$ scattering process shown in Figure 5.

Here, the incoming proton has momentum p, and the incoming/outgoing electrons have momenta k/k'. The electron interacts with a quark in the proton via a photon with momentum $k - k'$, which is spacelike. The virtuality is defined by $Q^2 = -(k - k')^2$. This is the energy scale of the photon, and is naturally related to the notion of virtuality being the energy scale at which the hadron is probed. The process $\mathrm{e}^-(k)\mathrm{p}(p) \to \mathrm{e}^-(k') + X$, where X represents some hadronic state, is called *deep inelastic scattering* (DIS) and its cross section is given by

$$\sigma = \sum_{\mathrm{q}} \int_0^1 dx\, f_q^{\mathrm{p}}(x, Q^2)\, \hat{\sigma}\left(\mathrm{e}^-(k)\mathrm{q}(xp) \to \mathrm{e}^-(k')\mathrm{q}(xp + k - k')\right), \tag{24}$$

at leading order. The electron–parton differential cross section to lowest order is given by

$$\frac{d\hat{\sigma}}{d\hat{t}} = \frac{2\pi\alpha^2 e_{\mathrm{q}}^2}{\hat{s}^2}\left(\frac{\hat{s}^2 + \hat{u}^2}{\hat{t}^2}\right), \tag{25}$$

where e_{q} is the charge of the quark q. The hats on the Mandelstam variables denote that they are the variables for the parton level process, as opposed to those for the full $\mathrm{e}^-\mathrm{p}$ process. In DIS, we use $Q^2 = -\hat{t}$ and $\hat{s} = xs$. Then the differential cross section for DIS is given by

$$\frac{d\sigma}{dQ^2} = \sum_{\mathrm{q}} \int_0^1 dx\, f_q(x, Q^2) \frac{2\pi\alpha^2 e_{\mathrm{q}}^2}{Q^4}\left[1 + \left(1 - \frac{Q^2}{xs}\right)^2\right]\Theta(xs - Q^2), \tag{26}$$

where the last Heaviside step function imposes the kinematic constraint $\hat{s} \geq |\hat{t}|$. This way, the parton distribution function is given in terms of the experimentally measurable DIS cross section. Note that since the gluons have no electric charge, $e_{\mathrm{g}} = 0$, they are not visible in DIS, but they contribute in analogous strong interactions.

4.2 Parameterizing PDFs

A common approach to describing PDFs in practice is to start by assuming they follow some parameterization at a particular initial scale Q_0, with parameters fitted to data. One example of such a parameterization is based on work by Glück and Reya [31–34]. In this work, pion PDFs are parameterized on the form

$$f(x, Q_0^2) = N x^a (1-x)^b \left(1 + A\sqrt{x} + Bx\right), \tag{27}$$

where f is a valence quark, sea quark, or gluon distribution, and a, b, A, B, N are parameters depending on the hadron species in question that must be fitted to data. For protons, there is a plethora of data, and the default dataset used by Pythia is from the NNPDF collaboration [35]. There is also a decent amount of data for pions [36–39]. Beyond these, only a small set exists for the u quark content of kaons [40], but there is no data at all for any other hadrons.

A central part of Paper IV is to extend parameterizations like these to other hadron species. In short, we do this by taking the form of (27) with $A = B = 0$ as an ansatz. For valence quarks, we choose b and $\langle x \rangle$ based only on qualified guesses, as there is no available data. We then derive a to give the correct $\langle x \rangle$ in (23), and derive N from (21). The details of this selection process are given in Section 2.4 of Paper IV, including fixing parameters for sea quarks and gluons.

In some cases, explicit parmeterizations can also be given as a function Q^2. However, it is more common to start with PDFs on the initial scale and evolve them to higher scales using the DGLAP equations [41–43]. These equations are

$$\frac{df_{\mathrm{g}}(x, Q^2)}{d\log Q} = \frac{\alpha_S(Q^2)}{\pi} \int_x^1 \frac{dz}{z} \left(P_{\mathrm{g}\leftarrow\mathrm{g}}(z) f_{\mathrm{g}}(x/z) + P_{\mathrm{g}\leftarrow\mathrm{q}}(z) \sum_q f_q(x/z) \right), \tag{28}$$

$$\frac{df_q(x, Q^2)}{d\log Q} = \frac{\alpha_S(Q^2)}{\pi} \int_x^1 \frac{dz}{z} \left(P_{\mathrm{q}\leftarrow\mathrm{q}}(z) f_{\mathrm{q}}(x/z) + P_{\mathrm{q}\leftarrow\mathrm{g}}(z) f_{\mathrm{g}}(x/z) \right). \tag{29}$$

The P functions are called *splitting functions*, and represent the rate of virtual fluctuations such as represented by the following diagrams shown in Figure 6. Intuitively, one can imagine that virtual fluctuations like these occur all the time on arbitrarily small length scales, but that the small scale fluctuations are only visible at sufficiently high resolutions. From this point of view splitting functions describe the rate at which the apparent structure of the parton changes as the resolution increases.

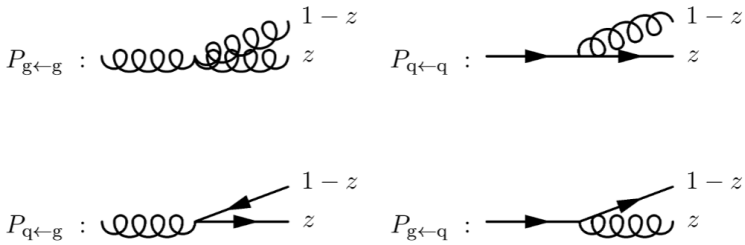


Figure 6: Splitting functions

The splitting functions are

$$\begin{aligned} P_{q\leftarrow q}(z) &= \frac{4}{3}\left(\frac{1-z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z)\right), \\ P_{g\leftarrow q}(z) &= \frac{4}{3}\left(\frac{1+(1-z)^2}{z}\right), \\ P_{q\leftarrow g}(z) &= \frac{1}{2}\left(z^2+(1-z)^2\right), \end{aligned} \tag{30}$$

and

$$P_{g\leftarrow g}(z) = 6\left(\frac{1-z}{z} + \frac{z}{(1-z)_+} + z(1-z) + \left(\frac{11}{12} - \frac{n_f}{18}\right)\delta(1-z)\right),$$

where $n_f$is the number of light quark flavors. In the $1/(1-z)_+$ denominators, the subscript $+$ indicates that this is a distribution[13] given by

$$\int_0^1 dx \frac{f(x)}{(1-x)_+} = \int_0^1 dx \frac{f(x)-f(1)}{(1-x)}, \tag{31}$$

which circumvents the singularity at $x = 1$.

Several frameworks exist for solving DGLAP equation, such as APFEL [44], HERAFitter/xFitter [45], and QCDNUM [46], with the latter used in Paper IV. Solving the equations is generally fast, even when evolving from an initial scale $Q_0 < 1$ GeV to a scale of up to 10^8 GeV. In practice, it can still be too slow for Pythia, and instead the PDFs are pre-calculated and tabulated in the LHAPDF format [47]. Several such new PDFs will be included in a future release of Pythia, based on our work in Paper IV.

[13]From a mathematical point of view, I like to think of distributions as functions that might not be possible to evaluate everywhere, but that make sense when integrated. For example, $\delta(x)$ cannot be evaluated at $x = 0$, but integrating it is still well-defined. This way, the property $\int dx\, \delta(x)f(x) = f(0)$ can be taken to be its definition. The analogous defining property of $1/(1-x)_+$ is given by (31).

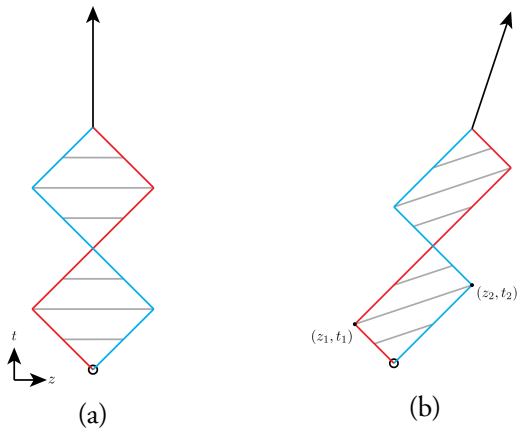
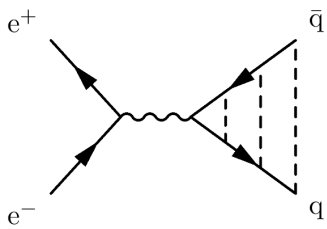


Figure 7: Yo-yo motion in the string model, a) in the CM frame for the system, and b) for a system with $p_z > 0$.

5 The Lund string model

5.1 Overview

The Lund string model [48, 49] is a model for *hadronization*. That is, it describes how the partons produced are converted into final state hadrons. The simplest illustration of the string model is through the process $e^+e^- \to \gamma/Z \to q\bar{q}$,



As the outgoing quarks move away from each other, a color field is drawn between them, indicated by dashed lines in the figure. This field is what is considered a "string" in the Lund string model. To give a more quantitative description, we make the ansatz that the potential of a color dipole at separation r is given by

$$V(r) = \kappa r. \tag{32}$$

The parameter κ is referred to as the *string tension*, and is 1 GeV/fm by default in Pythia. As the particles move apart, they expand the color field, converting their kinetic energy to potential energy in the string. Eventually, the quarks will reach a point where they have no more kinetic energy and start moving towards each other. As they do, they gain energy

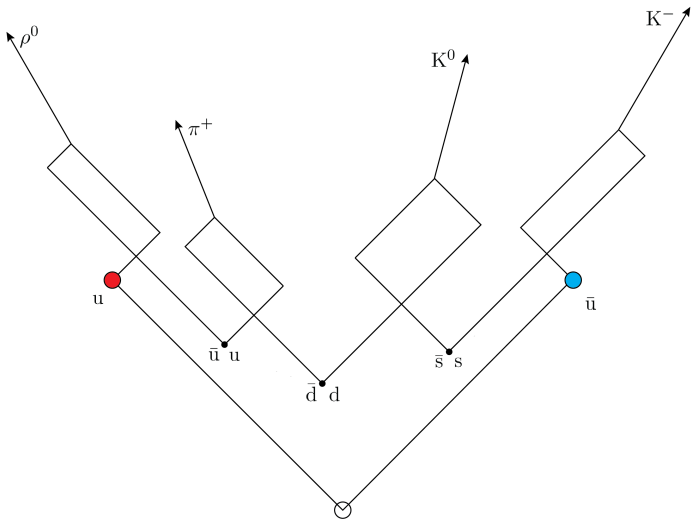


Figure 8: A possible result of fragmentation starting from a $u\bar{u}$ pair.

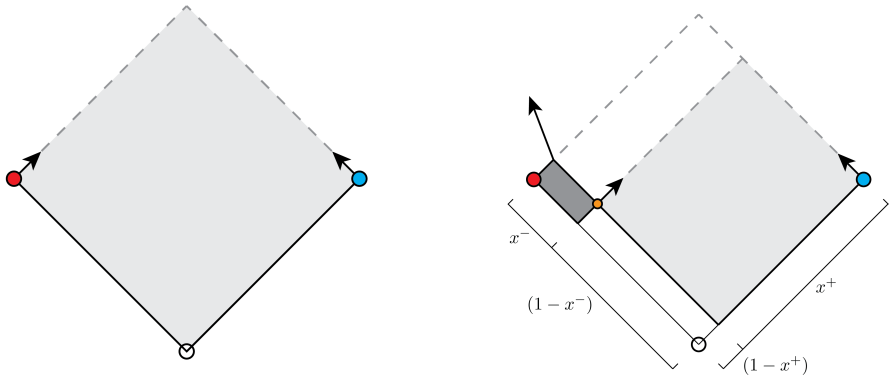
from the string and thus increase their momentum. If left to their own devices, this leads to an oscillatory "yo-yo" motion as shown in Figure 7a. Quarks that move in this fashion are interpreted as hadrons.

In other reference frames, the quark–antiquark pair may have a net momentum. In this case, the oscillations will appear as in Figure 7b. The overall energy and momentum of the system can be written as

$$E = \kappa(z_2 - z_1), \qquad p_z = \kappa(t_2 - t_1), \tag{33}$$

where z and t are the spacetime coordinates of the points where each quark has zero momentum (marked in Figure 7b). That is, this will be the energy and momentum of the produced hadron.

During this process, the string can fragment, producing new $q\bar{q}$ pairs, as shown in Figure 8. The quarks are assumed to have zero momentum when they are created, and gain momentum from the pull of the string. Eventually, adjacent quark–antiquark pairs will end up in a yo-yo motion and become hadrons. To implement this process in Pythia, we need to describe the procedure to pick the resulting hadron species and their momenta and spacetime production coordinates. The algorithm for this is iterative, starting from one endpoint and picking a breakup point to form a hadron with that endpoint. The breakup point is now a new endpoint for the leftover string, and the algorithm can be iterated to fragment this new string. This is illustrated in Figure 9.



(a) (b)

Figure 9: An illustration of a step in the string fragmentation algorithm. a) The initial setup at the beginning of the algorithm. The gray area indicates the region that would be spanned by the string if no fragmentation happens. b) After the first break has occurred, marked in orange, a hadron is produced on the left, indicated by the dark gray region, while a new system that is analogous to the initial setup is formed on the right.

In the most basic version of the algorithm, the first step is to pick a candidate hadron. To do this, we pick a flavor for the breakup point, and combine it with the flavor of the endpoint to form a hadron. The quarks are assumed to be produced at the same point; this means that essentially they are produced at zero energy, since they have not absorbed any string length at the time of their creation. If the particles are massive, this means they must tunnel through an energetically forbidden region. In principle this gives a suppression factor $\sim e^{-\pi m^2/\kappa}$, where m is the mass of the quark. In practice, however, it is not clear which quark masses to use, and the ratio for production of the different species is simply tuned to LEP data [50]. The ratios are $\mathrm{d} : \mathrm{u} : \mathrm{s} = 1 : 1 : 0.217$ by default, while the suppression of charm and bottom quarks is so large that they are not produced at all. The quark is also assigned a $p_\perp$, which also follows a Gaussian distribution for the same reason. After choosing a flavor, it can then combine with the existing endpoint to form a hadron. The ratios of pseudoscalar meson to vector meson production rates are given by input parameters. The rates are limited by spin-counting rules, but vector-pseudoscalar mass splitting gives different rates in practice.

The location of the breakup points are sampled in terms of fractions x of the lightcone momenta, as indicated in Figure 9. The probability of a particular breakup structure with n final hadrons is given by

$$d\mathcal{P} \propto \left(N d^2 p_i \, \delta(p_i^2 - m^2)\right) \, \delta\left(\sum_{i=1}^{n} p_i - P\right) \, e^{-bA}, \tag{34}$$

where P is the overall momentum of the string, A is the area covered by the string before fragmentation, and b is a parameter. In the iterative procedure used here, this is realized by sampling the fraction of lightcone momentum taken by the resulting hadron according to

$$f(x) = N\frac{(1-x)^a}{x}e^{-bm_\perp^2/x}, \tag{35}$$

where a is related to N and b by normalization and $x = x^-$ (x^+) if fragmenting from the left (right) side. The relation

$$m_h^2 = p^2 = x^+x^-E \tag{36}$$

can then be used to find the other lightcone coordinate.

The spacetime coordinates of hadrons are calculated after they have all been assigned their momenta. This step is optional since these vertices do not affect other part of a default Pythia run. However, they are vitally important to set up the initial geometry when considering for example hadronic rescattering. In the simplest case, the spacetime coordinates can be found in terms of energy and momentum according to (33). In a more complex string topologies, for example involving an additional hard gluon produced in the process $\mathrm{e^+e^-} \to \mathrm{q\bar{q}g}$, selecting breakup vertices becomes more complicated. This is true both for the energy–momentum and for the space–time picture, and the two are no longer as simply related to each other. The details of the space–time picture are worked out in [14].

5.2 Baryon production

For simplicity, we have so far considered only meson production, but the Lund string model can be extended in several ways to also provide mechanisms for baryon production.

The most straightforward way of doing this is to allow strings to fragment directly into diquark–antidiquark pairs (where the diquark carries anticolor). Diquark production is suppressed according to an $e^{-\pi m^2/\kappa}$ factor, as described above. A consequence of this approach is that baryons produced this way can only be created next to each other on the string, and they will always share at least two flavors. One mechanism to produce more general configurations is the *popcorn model* [51, 52]. In this model, quark–antiquark pairs can be produced as vacuum fluctuations on a string without breaking it. This is shown in Figure 10. When this happens, there is a segment of the string that basically carries two colors (red and green in Figure 10), which is equivalent to an antiblue string segment with the color flow is in the other direction. This segment can then fragment at several points, producing particles in between the two baryons.

Another way to form baryons is through junction topologies [53]. A *junction* is a point where three strings of different colors meet, all pointing either inwards or outwards. These strings can then fragment, and in the end the leftover junction will be treated as a baryon.

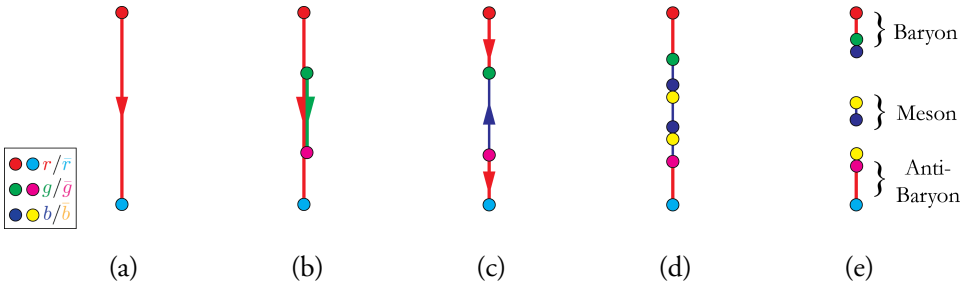


Figure 10: Illustration of the popcorn model. Red/green/blue indicate colors, while cyan/magenta/yellow indicate anticolors. a) Start with a red string. b) Vacuum fluctuations produce a green–antigreen pair. c) The red+green segment is equivalent to a blue segment with color flowing in the opposite direction. d) The blue string fragments at two locations. d) This produces a baryon, a meson and an antibaryon.

In Pythia, junctions can form for example if all three valence quarks are kicked out of an incoming proton, or through color reconnection processes as illustrated in Figure 11.

5.3 String interactions

Since QCD fields carry color charge, they can interact with each other. One of the powers of the Lund string model is that it invites a picture of strings as physical objects, which easily lends itself to extensions where we consider interactions between these strings.

One example of such an interaction is rope formation [54]. When two color fields lie close to each other they can combine, essentially forming a string with a higher string tension κ. Such a string is called a *rope*. An important consequence of this increased κ is that it mitigates the exponential suppression of diquarks and strange quarks. This phenomenon can give a significant contribution to the observed strangeness enhancement in high multiplicity events. Another example is *shoving* [55, 56], which refers to forces between strings giving corrections to their momenta. These corrections are of particular importance because they can give rise to collective flow. For two parallel strings at a transverse separation $d_\perp$, the interaction force per unit length is given by

$$f(d_\perp) = \frac{g\kappa d_\perp}{R^2} e^{-d_\perp^2/4R^2}, \tag{37}$$

where g is of the order 1 and is a tunable parameter . This force can be either attractive or repulsive, depending on whether the color flow in the two strings is in the same or opposite directions, and whether the combined color is in a singlet or octet state. The repulsive cases are more common, and in Pythia it is in fact assumed that all interactions

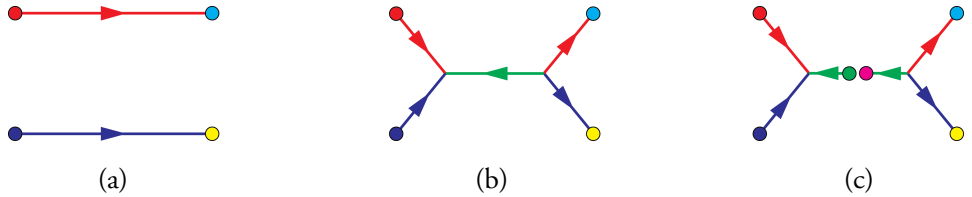


Figure 11: Formation of string junctions. a) Two strings, one red and one blue, are produced close to each other. b) The strings combine to create one green string with color flow in the opposite direction. c) The green string can fragment, producing a baryon–antibaryon pair.

are repulsive, which is an acceptable approximation when compared with other assumptions that go into the description. At the time of writing, ropes and shoving are only handled for parallel strings, and the interaction strength is amplified to compensate for this. This gives a reasonable description of data in pp collisions, but breaks down for heavy ions where the number of non-parallel strings is much greater. An extension to Angantyr for handling general string distributions is currently under development in Lund. This framework is called Gleipnir, and is expected to be published in the near future.

A third interaction between strings is *color reconnection* [53]. A simplification often used in MC generators is that there are infinitely many colors in QCD, $N_c \to \infty$, instead of the actual $N_c = 3$. This makes things simpler because each new gluon emission creates a new color, and the colors do not interfere with each other. In reality where $N_c = 3$, one would expect more complicated rules for how color topologies should be set up. Color reconnection refers to modifications to the color structure, and affects amongst other things the average length of strings.

5.4 Ministrings

Some string systems may be produced at very low energies, for example towards the end of a parton shower or during rescattering. One problem with such systems is that there might not be enough energy for a valid hadronization. Consider for instance an $s\bar{s}$ pair produced at 0.9 GeV. This system cannot fragment normally, as this would produce at least a $K\bar{K}$ pair and there is not enough energy for this. Nor can the string collapse into a yo-yo motion, since there is no particle with the right mass and compatible flavour content – the closest candidates is η', but it would be produced too far off shell. Another problem is that even when valid outgoing configurations do exist, the valid phase space might be so small that it is not possible to sample efficiently with the default approach. A third problem that can occur, especially in the context of rescattering, is that if two strings are produced and they each collapse to a single hadron, the process is reduced to a pseudo-elastic process

$AB \to XY \to AB$, where X and Y represent string systems.

Low energy string systems are referred to as *ministrings* [57]. These are given special treatment in Pythia that is reminiscent of cluster models for hadronization [58]. Several attempts are made to hadronize, and if they fail, additional measures are taken to increase the chance of success. These measures include collapsing one string to a single hadron and imparting momentum to a parton in the other string, choosing only the lightest available hadron when picking the hadron types, and collapsing a system to a single π^0 plus the lightest possible hadron given its endpoint flavours. This last measure would solve the aforementioned example, since an $s\bar{s}$ pair at 0.9 GeV could result in an outgoing $\eta\pi^0$ system.

Since ministrings are particularly important for rescattering, some small additions to Pythia's ministring framework were developed Paper I, and a more technical discussion of the situation near the threshold is given there (Section 3.6).

6 Applications

The theme of this thesis has been to generalize hadron–hadron interactions in Pythia. In isolation, such simulations are mere curiosities, which are not useful until we look at applications where the results of these simulations can be linked to experiment.

This work focuses on two applications, namely hadronic rescattering and medium propagation. In fact, while the theme of this thesis is to generalize hadronic interactions, a large part of it actually revolves around rescattering. In this section, we outline the basic concepts of hadronic rescattering, while a more detailed and technical discussion is found in Paper I. We also discuss the basics of exotic hadrons and medium propagation, which are the topics of Paper III and Paper IV, respectively.

6.1 Hadronic rescattering

Hadronic rescattering refers to secondary collisions between hadrons that were produced in the primary high-energy collision. To make an analogy with billiards, one can think of the primary collision as the break shot, and billiard balls subsequently bouncing off each other as the secondary collisions, as illustrated in Figure 12. Now, if we play on an infinitely large pool table, it seems reasonable that these secondary collisions should not significantly affect the final outgoing distribution of the pool balls. Analogously, rescattering has not been thought to have a significant effect on the kind of studies that Pythia were used for in the past, and hence implementing it has not had high priority. However, in recent times, observables such as collective flow have become relevant. Rescattering is a mechanism that

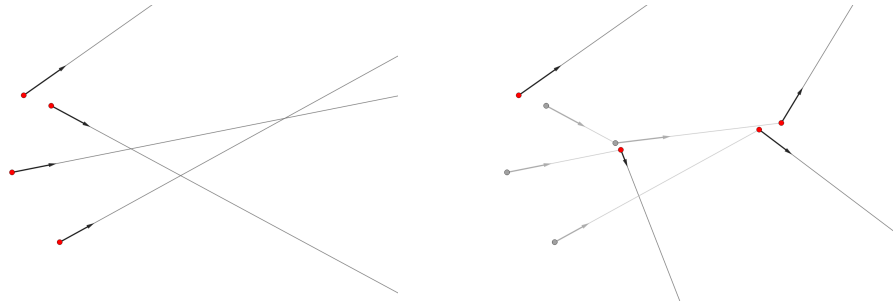


Figure 12: Illustration of rescattering. a) The initial state, showing the location and trajectories of particles. b) A possible final state after rescattering.

can give rise to these particular phenomena, and this has incentivised implementing it in Pythia.

While rescattering has never been part of a general-purpose event generator, other frameworks exist that are specifically designed to do rescattering. One example is UrQMD [24], whose approaches inspired a lot of the work we did in Paper I, and which also discusses the theoretical foundations underlying hadronic transport. A more modern example that is still under active development is SMASH [59].

One study has been performed that used UrQMD as an afterburner to perform rescattering on hadronic states generated by Angantyr [60]. This study found signs of collective flow, which is consistent with our results in Paper II. However, creating and maintaining an interface between Angantyr and UrQMD is a non-trivial task, which requires converting the Pythia event format into a form that can be read by UrQMD. This conversion also erases some particle history, which makes it impossible to trace the evolution of the hadronic cascade. One of the main appeals of rescattering in Pythia is therefore its simplicity: it can be enabled by a single line in the configuration file, and having it in a native format makes it easier to study the output. Furthermore, being built on the Pythia infrastructure, it lends itself more easily to further extensions by other Pythia authors in the future.

Secondary collisions happen predominantly at low energies, which have been discussed in Section 3. What remains in order to model hadronic rescattering is to define which particles will interact with each other when viewing the event as a whole, and when and where these interactions occur. To this end, we consider each hadron pair in the rest frame of that pair and define the interaction time t as the time at which they pass each other (if they are moving away from each other at the time they are produced, they cannot rescatter). Whether the particles actually do interact is probabilistic, and the probability P depends on the particle species, their CM energy, and the impact parameter b. We assume that the dependency on particle species and CM energy is only through the total cross section σ, as

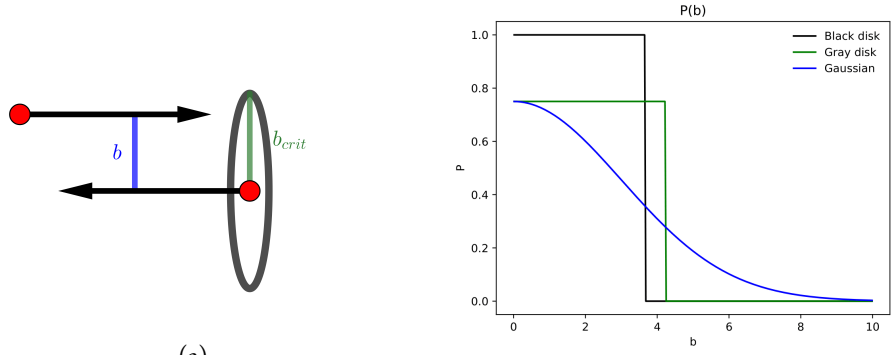


Figure 13: a) Illustration of two particles colliding at an impact parameter b. b) Different models for the probability of an interaction as a function of the impact parameter.

defined in Section 3.1, and define the interaction range $b_{\text{crit}} = \sqrt{\sigma/\pi}$. This views the cross section geometrically as a "target size", as shown in Figure 13a. The exact dependency of P on b is not well understood, so we offer two models. A *black disk* model refers to one where the particles will interact with certainty if $b < b_{\text{crit}}$ and not interact otherwise. In Pythia we generalize this to a "gray disk", where the probability is $P_0 < 1$ below a threshold and in return, the threshold is wider. An alternative shape is a Gaussian dropoff with b. These models are shown in Figure 13b.

6.2 Exotic hadrons

The term exotic hadron refer to hadrons that are not mesons or baryons, such as tetraquarks, which consist of two quarks and two antiquarks, or pentaquarks, consisting of four quarks and one antiquark. Examples are the tetraquark state $\chi_{c1}(3872)$ (also called X(3872)), which was first discovered by the Belle experiment in 2003 [61], and the pentaquark states $P_c^+(4312)$, $P_c^+(4440)$ and $P_c^+(4457)$, which are based on data reported by LHCb in 2015 [62] and 2019 [63].

One of the reasons exotic hadrons are interesting is that we do not have a detailed understanding of their structure. The $P_c^+(4312)$ pentaquark, for instance, is believed to have a $uudc\bar{c}$ quark content, but the question is how these are arranged. One option is to view it as five quarks bound together without any particular substructure[14], while another is to see it as a *molecular state* consisting of a Σ_c^+ and a $\bar{D}^0$ meson. Evidence for the latter include the fact that its mass lies only a few MeV above the $\Sigma_c^+\bar{D}^0$ mass threshold, and that the

[14]Except for substructures expected in regular heavy hadrons, e.g. light quarks orbiting the heavy flavored ones.

decay $P_c^+(4312) \to \Sigma_c^+ \bar{D}^0$ is the dominant channel (the "fall-apart" channel). One model suggests that the $P_c^+(4440)$ and $P_c^+(4457)$ correspond to $\Sigma_c^+ \bar{D}^{0*}$ molecular states, with total spin 1/2 and 3/2, respectively.

The picture of exotic hadrons as molecular states suggests that they should also be able to form in collisions between the constituent hadrons. In Paper III, we study this by implementing pentaquark production using the rescattering framework.

There are several theoretical models that describe exotic hadrons (see [64] for a review), but they have never been studied in the context of general-purpose event generators. By taking this first step, we hope others may use PYTHIA in the future to compare models and formulate more concrete experimental predictions.

6.3 Medium propagation

The main application that motivated us to extend PYTHIA to include high energy collisions between generic hadron types is hadronic cascades in media. When a particle enters a medium, such as the Earth's atmosphere or a particle detector, it causes a cascade of hadrons. These cascades includes any hadron species that can be produced in a normal high energy particle collision, and if they are sufficiently long-lived, they can cause further interactions with the medium.

In a uniform medium, the distance a particle would propagate is an exponential distribution with average value $l_0 = 1/\sigma n$, where σ is the total cross section for a collision between the propagating particle and a medium particle and n is the particle number density in the medium. One way to interpret this is to view the particle as a black disk with area σ that after moving a distance l has covered a volume σl in space. An interaction occurs if a medium particle lies within this volume. l_0 is often referred to as the *mean free path*. In some media such as the atmosphere, the density n varies with height, but the probability of an interaction as the particle at height z moves a short distance ds is still given by $\sigma n(z)\, ds$ as for an exponential distribution. This allows us to sample the distance a particle travels, either by an analytic expression (which we are able to do in Paper IV), or using the veto algorithm [1] if such an expression cannot be obtained.

In most realistic media, the medium particles are nuclei. Ideally we would like to model such collisions using ANGANTYR, but there are currently a few issues with this approach. Unlike in default PYTHIA, it is not (yet) possible to easily change the CM energy in ANGANTYR after initialization. To simulate secondary collisions, a new energy must be chosen, which in ANGANTYR would require significant reinitialization, and this is too computationally inefficient in practice. A second issue is that ANGANTYR makes certain assumptions that are not valid at low energies, for example assuming ions are flat in the z-direction due

to Lorentz contraction. Furthermore, the MPI machinery used for modelling collisions in Angantyr is only applicable to high energies, and the framework does not yet use the low energy processes introduced in this thesis. Finally, Angantyr has no detailed handling of the nucleons that do not participate in any subcollisions, but rather lumps them all together to one single nuclear remnant for each beam. The nuclear remnants would surely have significant effects on the hadronic cascade, and as such the current solution is inadequate. For these reasons, we were not able to use Angantyr in Paper iv. Instead, we relied on a simplified description where the collision is first modelled as a hadron–nucleon collision (either p or n, chosen according to the nucleon content of the nucleus), then the outgoing hadrons have a chance to immediately interact with further nucleons. The details of this procedure are outlined in Section 3.3 of Paper iv.

There are a few existing frameworks for medium interactions. For cosmic rays, two well-known examples are CORSIKA [10] and AIRES [65, 66]. Examples that focus on detector simulations are Fluka [67] and GEANT [68–71], with the latter having more recently been used also for atmospheric cascades [72, 73]. Our intention is not to implement a framework for medium interactions within Pythia, as we would not be able to get close to the level of detail provided by these state-of-the-art frameworks. Instead, our goal is to provide a framework for only the collisions themselves, in the hopes that Pythia can be used as a plugin by these existing models. Nevertheless, we developed a toy model for cascades in Paper iv as a way to test our model and to provide an example for how to use Pythia in this fashion.

7 Outlook

The main theme of this thesis has been to extend Pythia's repertoire of allowed beam configurations. In the past, it supported only pp- at energies above 10 GeV. With this work, hadron–hadron collisions have been made available for all species, both above and below 10 GeV. The main applications presented here are hadronic rescattering, where mostly low-energy interactions are important, and cosmic ray physics, where both low and high energies become relevant.

In this thesis, I have made two studies where hadronic rescattering was central. The first is its effect in heavy ion collisions. One of the most interesting results in Paper ii is that rescattering gives rise to a non-negligible collective flow, especially at large multiplicities where there are more rescatterings. The big question is if other observations that are commonly associated with a QGP can be described within the Angantyr model. Hadronic rescattering is only one piece of this puzzle, and to answer the question, further ingredients such as string shoving, ropes and color reconnection are needed. It is my hope that the ongoing research in Lund will lead to important insights on this topic within the next

decade.

The second application of rescattering, explored in Paper III, is as a mechanism for the production of exotic hadrons. While exotic hadrons are not as hot a topic as QGP, their properties are still important for our understanding of QCD. Paper III serves as a first step towards a phenomenological approach to studying exotic hadrons. In the paper, we see that rescattering gives a relatively small contribution to exotic hadron production, but other model choices may give larger contributions. As this is only a first step, it is my hope that Pythia may now be used in the future as a way to study the details of various theoretical models for exotic hadrons, providing a framework for comparing models to each other and enabling them models to make experimental predictions. The general framework developed in this paper can also be applied to the production of other kinds of particles, such as deuterons.

In Paper IV, our focus was on particles propagating in mediums. The paper focuses on cosmic rays, but the basic ideas are easily generalizable. Such a propagation involves numerous hadronic interactions with the medium, with many different hadron species and at both high and low energies. To test our implementation, we used a toy model for hadronic cascades and found that the behaviour was roughly as one would expect. But the main goal for the paper was to model the hadronic interactions themselves, so that other programs that do detailed modelling of propagation (such as CORSIKA or Geant4) can utilize Pythia as a way to simulate the collisions.

There are a few venues for future development within Pythia. One of the possible future improvements to the hadronic rescattering model is the inclusion of processes involving three or more incoming hadrons. Hadronic rescattering tends to increase multiplicity, especially in heavy ion collisions, and the inclusion of processes that can also reduce the multiplicity may improve the precision of the model.

In the high energy framework developed in Paper IV, there are some technical issues that can be improved on. First, a technical detail is that our simulation required one Pythia instance initialized for each class of particles (i.e. particles which are excitations of each other). Each such particle class is initialized with an MPI configuration, but the infrastructure to switch between these after initialization was not implemented due to time constraints. Doing this in the future may improve performance such as reducing memory usage and initialization time, and make the framework more user friendly.

In a similar vein, there are some technical limitations in Angantyr. While Pythia provides a mechanism for changing the energy of hadron–hadron collisions on an event-by-event basis, there is as of today no such mechanism for Angantyr collisions. In other words, varying the energy will require a complete reinitialization, which would have an unacceptable performance impact in any cosmic ray simulation. This led us to use an ad hoc nucleon model in Paper IV, but it would be highly desirable to solve this issue before applying An-

GANTYR in state-of-the-art cosmic ray simulations. Furthermore, ANGANTYR is not designed to handle collisions at low energies (in this case meaning $\lesssim$ 100 GeV). Thus there are still several outstanding issues in ANGANTYR.

PYTHIA is a vital tool in modern particle physics research, and there is an active community that is working on it. As such, several of these aforementioned issues will surely be addressed in the future, by me or others. For now, I am proud to have been part of this ongoing development, and look forwards to seeing how my work will be used by others in the future.

References

[1] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP* **05** (2006) 026, hep-ph/0603175.

[2] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.* **191** (2015) 159–177, 1410.3012.

[3] J. Bellm *et. al.*, "Herwig 7.0/Herwig++ 3.0 release note," *Eur. Phys. J. C* **76** (2016), no. 4 196, 1512.01178.

[4] T. Gleisberg, S. Hoeche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *JHEP* **02** (2009) 007, 0811.4622.

[5] **PHENIX** Collaboration, K. Adcox *et. al.*, "Suppression of hadrons with large transverse momentum in central Au+Au collisions at $\sqrt{s_{NN}}$ = 130-GeV," *Phys. Rev. Lett.* **88** (2002) 022301, nucl-ex/0109003.

[6] **STAR** 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** (2003) 082302, nucl-ex/0210033.

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

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

[9] D. J. Bird *et. al.*, "Detection of a cosmic ray with measured energy well beyond the expected spectral cutoff due to cosmic microwave radiation," *Astrophys. J.* **441** (1995) 144–150, astro-ph/9410067.

[10] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw, "CORSIKA: A Monte Carlo code to simulate extensive air showers,".

[11] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[12] G. Ingelman and P. E. Schlein, "Jet Structure in High Mass Diffractive Scattering," *Phys. Lett. B* **152** (1985) 256–260.

[13] C. O. Rasmussen and T. Sjöstrand, "Models for total, elastic and diffractive cross sections," *Eur. Phys. J. C* **78** (2018), no. 6 461, 1804.10373.

[14] S. Ferreres-Solé and T. Sjöstrand, "The space–time structure of hadronization in the Lund model," *Eur. Phys. J. C* **78** (2018), no. 11 983, 1808.04619.

[15] B. Andersson, G. Gustafson, and B. Nilsson-Almqvist, "A Model for Low p(t) Hadronic Reactions, with Generalizations to Hadron - Nucleus and Nucleus-Nucleus Collisions," *Nucl. Phys. B* **281** (1987) 289–309.

[16] A. Bialas, M. Bleszynski, and W. Czyz, "Multiplicity Distributions in Nucleus-Nucleus Collisions at High-Energies," *Nucl. Phys. B* **111** (1976) 461–476.

[17] A. Donnachie and P. V. Landshoff, "Total cross-sections," *Phys. Lett. B* **296** (1992) 227–232, hep-ph/9209205.

[18] **Particle Data Group** Collaboration, M. Tanabashi *et. al.*, "Review of Particle Physics," *Phys. Rev. D* **98** (2018), no. 3 030001.

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

[20] J. R. Pelaez, A. Rodas, and J. Ruiz De Elvira, "Global parameterization of $\pi\pi$ scattering up to 2 GeV," *Eur. Phys. J. C* **79** (2019), no. 12 1008, 1907.13162.

[21] J. R. Pelaez and A. Rodas, "Pion-kaon scattering amplitude constrained with forward dispersion relations up to 1.6 GeV," *Phys. Rev. D* **93** (2016), no. 7 074025, 1602.08404.

[22] E. M. Levin and L. L. Frankfurt, "The Quark hypothesis and relations between cross-sections at high-energies," *JETP Lett.* **2** (1965) 65–70.

[23] H. J. Lipkin, "Quarks for pedestrians," *Phys. Rept.* **8** (1973) 173–268.

[24] S. A. Bass *et. al.*, "Microscopic models for ultrarelativistic heavy ion collisions," *Prog. Part. Nucl. Phys.* **41** (1998) 255–369, nucl-th/9803035.

[25] V. Srinivasan *et. al.*, "Pi- pi+ –> pi- pi+ Interactions Below 0.7-GeV from pi- p –> pi- pi+ n Data at 5-GeV/c," *Phys. Rev. D* **12** (1975) 681.

[26] S. D. Protopopescu, M. Alston-Garnjost, A. Barbaro-Galtieri, S. M. Flatte, J. H. Friedman, T. A. Lasinski, G. R. Lynch, M. S. Rabin, and F. T. Solmitz, "Pi pi Partial Wave Analysis from Reactions pi+ p —> pi+ pi- Delta++ and pi+ p —> K+ K- Delta++ at 7.1-GeV/c," *Phys. Rev. D* **7** (1973) 1279.

[27] N. N. Biswas, N. M. Cason, I. Derado, V. P. Kenney, J. A. Poirier, and W. D. Shephard, "Total Pion-Pion Cross Sections for the 2-GeV Di-Pion Mass Region," *Phys. Rev. Lett.* **18** (1967), no. 7 273–276.

[28] W. J. Robertson, W. D. Walker, and J. L. Davis, "High-energy pi pi collisions," *Phys. Rev. D* **7** (1973) 2554–2564.

[29] G. A. Schuler and T. Sjöstrand, "Hadronic diffractive cross-sections and the rise of the total cross-section," *Phys. Rev. D* **49** (1994) 2257–2267.

[30] P. Koch and C. B. Dover, "$K^{\pm}$, $\bar{p}$ and Ω^{-} Production in Relativistic Heavy Ion Collisions," *Phys. Rev. C* **40** (1989) 145–155.

[31] M. Glück, E. Reya, and A. Vogt, "Pionic parton distributions," *Z. Phys. C* **53** (1992) 651–656.

[32] M. Glück, E. Reya, and I. Schienbein, "Pionic parton distributions revisited," *Eur. Phys. J. C* **10** (1999) 313–317, hep-ph/9903288.

[33] M. Glück, E. Reya, and M. Stratmann, "Mesonic parton densities derived from constituent quark model constraints," *Eur. Phys. J. C* **2** (1998) 159–163, hep-ph/9711369.

[34] M. Glück, P. Jimenez-Delgado, and E. Reya, "Dynamical parton distributions of the nucleon and very small-x physics," *Eur. Phys. J. C* **53** (2008) 355–366, 0709.0614.

[35] **NNPDF** Collaboration, R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo, "Parton distributions with QED corrections," *Nucl. Phys. B* **877** (2013) 290–320, 1308.0598.

[36] **NA3** Collaboration, J. Badier *et. al.*, "Experimental Determination of the pi Meson Structure Functions by the Drell-Yan Mechanism," *Z. Phys. C* **18** (1983) 281.

[37] **NA10** Collaboration, B. Betev *et. al.*, "Differential Cross-section of High Mass Muon Pairs Produced by a 194-GeV/$c\pi^{-}$ Beam on a Tungsten Target," *Z. Phys. C* **28** (1985) 9.

[38] **NA10** Collaboration, P. Bordalo *et. al.*, "Nuclear Effects on the Nucleon Structure Functions in Hadronic High Mass Dimuon Production," *Phys. Lett. B* **193** (1987) 368.

[39] J. S. Conway *et. al.*, "Experimental Study of Muon Pairs Produced by 252-GeV Pions on Tungsten," *Phys. Rev. D* **39** (1989) 92–122.

[40] **Saclay-CERN-College de France-Ecole Poly-Orsay** Collaboration, J. Badier *et. al.*, "Measurement of the K^-/π^- Structure Function Ratio Using the Drell-Yan Process," *Phys. Lett. B* **93** (1980) 354–356.

[41] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory," *Sov. J. Nucl. Phys.* **15** (1972) 438–450.

[42] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.," *Sov. Phys. JETP* **46** (1977) 641–653.

[43] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys. B* **126** (1977) 298–318.

[44] V. Bertone, S. Carrazza, and J. Rojo, "APFEL: A PDF Evolution Library with QED corrections," *Comput. Phys. Commun.* **185** (2014) 1647–1668, 1310.1394.

[45] S. Alekhin *et. al.*, "HERAFitter," *Eur. Phys. J. C* **75** (2015), no. 7 304, 1410.4412.

[46] M. Botje, "QCDNUM: Fast QCD Evolution and Convolution," *Comput. Phys. Commun.* **182** (2011) 490–532, 1005.1481.

[47] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, "LHAPDF6: parton density access in the LHC precision era," *Eur. Phys. J. C* **75** (2015) 132, 1412.7420.

[48] B. Andersson, G. Gustafson, and B. Soderberg, "A General Model for Jet Fragmentation," *Z. Phys. C* **20** (1983) 317.

[49] T. Sjöstrand, "Jet Fragmentation of Nearby Partons," *Nucl. Phys. B* **248** (1984) 469–502.

[50] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune," *Eur. Phys. J. C* **74** (2014), no. 8 3024, 1404.5630.

[51] A. Casher, H. Neuberger, and S. Nussinov, "Chromoelectric Flux Tube Model of Particle Production," *Phys. Rev. D* **20** (1979) 179–188.

[52] B. Andersson, G. Gustafson, and T. Sjöstrand, "Baryon Production in Jet Fragmentation and Υ Decay," *Phys. Scripta* **32** (1985) 574.

[53] J. R. Christiansen and P. Z. Skands, "String Formation Beyond Leading Colour," *JHEP* **08** (2015) 003, 1505.01681.

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

[55] C. Bierlich, G. Gustafson, and L. Lönnblad, "Collectivity without plasma in hadronic collisions," *Phys. Lett. B* **779** (2018) 58–63, 1710.09725.

[56] C. Bierlich, S. Chakraborty, G. Gustafson, and L. Lönnblad, "Setting the string shoving picture in a new frame," *JHEP* **03** (2021) 270, 2010.07595.

[57] E. Norrbin and T. Sjöstrand, "Production and hadronization of heavy quarks," *Eur. Phys. J. C* **17** (2000) 137–161, hep-ph/0005110.

[58] S. Wolfram, "Parton and Hadron Production in e^+e^- Annihilation," in *15th Rencontres de Moriond: Part I, High-Energy Hadronic Interactions*, 1980.

[59] J. Weil *et. al.*, "Particle production and equilibrium properties within a new hadron transport approach for heavy-ion collisions," *Phys. Rev. C* **94** (2016), no. 5 054905, 1606.06642.

[60] A. V. da Silva, C. Bierlich, D. Dobrigkeit Chinellato, and J. Takahashi, "Studying the Effect of the Hadronic Phase in Nuclear Collisions with PYTHIA and UrQMD," *Springer Proc. Phys.* **250** (2020) 319–323, 1911.12824.

[61] **Belle** Collaboration, S. K. Choi *et. al.*, "Observation of a narrow charmonium - like state in exclusive B+- —> K+- pi+ pi- J / psi decays," *Phys. Rev. Lett.* **91** (2003) 262001, hep-ex/0309032.

[62] **LHCb** Collaboration, R. Aaij *et. al.*, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays," *Phys. Rev. Lett.* **115** (2015) 072001, 1507.03414.

[63] **LHCb** Collaboration, R. Aaij *et. al.*, "Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$," *Phys. Rev. Lett.* **122** (2019), no. 22 222001, 1904.03947.

[64] **Particle Data Group** Collaboration, P. A. Zyla *et. al.*, "Review of Particle Physics," *PTEP* **2020** (2020), no. 8 083C01.

[65] S. J. Sciutto, "AIRES: A system for air shower simulations," astro-ph/9911331.

[66] S. J. Sciutto, "The AIRES system for air shower simulations: An Update," in *27th International Cosmic Ray Conference*, 6, 2001. astro-ph/0106044.

[67] A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft, "FLUKA: A multi-particle transport code (Program version 2005),".

[68] R. Brun, F. Bruyant, F. Carminati, S. Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban, "GEANT Detector Description and Simulation Tool,".

[69] **GEANT4** Collaboration, S. Agostinelli *et. al.*, "GEANT4–a simulation toolkit," *Nucl. Instrum. Meth. A* **506** (2003) 250–303.

[70] J. Allison *et. al.*, "Geant4 developments and applications," *IEEE Trans. Nucl. Sci.* **53** (2006) 270.

[71] J. Allison *et. al.*, "Recent developments in Geant4," *Nucl. Instrum. Meth. A* **835** (2016) 186–225.

[72] P. Paschalis, H. Mavromichalaki, L. I. Dorman, C. Plainaki, and D. Tsirigkas, "Geant4 software application for the simulation of cosmic ray showers in the Earth's atmosphere," *New Astron.* **33** (2014) 26–37.

[73] R. Sarkar, A. Roy, and S. K. Chakrabarti, "Simulation of cosmic rays in the Earth's atmosphere and interpretation of observed counts in an X-ray detector at balloon altitude near tropical region," *Adv. Space Res.* **65** (2020) 189–197.

8 Publications

Paper I

T. Sjöstrand, M. Utheim,
A Framework for Hadronic Rescattering in pp Collisions,
Eur. Phys. J. C **80**, 907 (2020)
e-Print: `arXiv:2005.05658` [hep-ph]
MCnet–20-11, LU TP 20–12

My first paper presents a new framework for hadronic rescattering in Pythia. A significant part of this framework is new models for hadron–hadron interactions at low energies, not only between common hadron species like nucleons, but also all uncommon ones, even including charmed and bottom hadrons.

On the technical side, I wrote most of the code for the rescattering algorithm and cross section calculations, while Torbjörn Sjöstrand wrote code for modelling low-energy interactions. In the article, I wrote most of the description of the rescattering model (Section 3), Torbjörn wrote most of the introduction, background and summary (Sections 1, 2 and 6), and model tests and comparisons to data (Sections 4 and 5) were written in collaboration.

Paper II

C. Bierlich, T. Sjöstrand, M. Utheim,
Hadronic Rescattering in pA and AA Collisions,
Eur. Phys. J. A **57**, 227 (2021)
e-Print: `arXiv:2103.09665` [hep-ph]
MCnet–21–03, LU TP 21–08

The second paper builds on the model introduced in Paper 1, and studies the effects of hadronic rescattering in the context of pA and AA collisions, using the Angantyr framework. While all authors contributed with analyses and in some capacity to all parts of the article, my main focus was on model tests (Section 3).

Paper III

P. J. Ilten, M. Utheim,
Forming Molecular States with Hadronic Rescattering,
To be submitted to *Eur. Phys. J. A*,

e-Print: arXiv:2108.03479 [hep-ph]
MCnet–21–13, LU TP 21–31

Paper III is a smaller paper that applies the model developed in Paper I to study exotic hadron formation in rescattering, motivated by the idea that exotic hadron can be seen as molecular states (e.g. $\Sigma_c^+ \bar{D}^{*0}$) that can form when the constituent hadrons interact near the pole mass of the exotic hadron. In this project, Philip Ilten contributed with expertise and background studies to determine which studies to undertake, while I wrote the code necessary for exotic hadron formation in rescattering and for analysis. The article was written in collaboration.

Paper IV

T. Sjöstrand, M. Utheim,
Hadron Interactions for Arbitrary Energies and Species, with Applications to Cosmic Rays,
To be submitted to *Eur. Phys. J. C*,
e-Print: arXiv:2108.03481 [hep-ph]
MCnet–21–14, LU TP 21–32

Unlike the other papers, Paper IV moves away from hadronic rescattering. The goal of the project was to develop a framework to facilitate simulating hadronic cascades in mediums. In Paper I we implemented low energy interactions between many uncommon hadron species, and a large part of this paper was to extend such interactions to perturbative energies, in particular defining PDFs for uncommon hadrons. My contribution focused on parameterizing these PDFs and implementing them in Pythia, while Torbjörn Sjöstrand wrote the necessary code to simulate interactions between the additional hadron species. In addition, I wrote a simple algorithm for a rudimentary simulation of hadronic cascades in the atmosphere, as a way to test our model.

Additional contributions to Pythia

In addition to the four main papers included in this thesis, I made additional contributions to Pythia, not presented in this thesis. This includes technical maintenance, especially around the release of Pythia 8.3, and writing parts of the upcoming Pythia 8.3 manual.

9 Acknowledgements

At the time of writing, the COVID-19 pandemic has been going on for almost a year and a half. There are many dear friends and family I haven't had a chance to see during this time, and with the workload, I haven't been able to stay in touch as much as I would have liked. This isolation has made me realize how important these people are to me, and so I would like to extend my love and thanks to everyone close to me. In the following paragraphs, I will acknowledge the people who have been directly involved with my PhD.

First and foremost, I would like to thank my supervisor Torbjörn Sjöstrand, for the guidance he has given me and the efforts he has put in to help me succeed. Coming from a background in mathematics, I had a lot to learn about particle physics when I first moved to Lund, and Torbjörn's vast knowledge on the field has been of invaluable help to this end. I would also like to thank my two other collaborators, Christian Bierlich and Philip Ilten, both of whom I enjoyed working with, and who have taught me a lot and helped me grow as an independent scientist.

Next, I would like to thank my roommates Yasir, Federica and Nina for being awesome people in general; in the future when I think back at my thesis, I will remember you and all the fun we had while I was writing. Two other roomies I lived with earlier and that I will remember fondly are Oscar and Adriana (and Mary, of course); thank you for the time together in Lund, and I hope we meet again in the future.

During the last few months of writing this thesis, several friends and colleagues helped my writing by providing valuable feedback. A special thanks to Malin Sjödahl, Smita Chakraborty and Ole Christian Hagestad for all the time they spent on giving me feedback. An additional thanks to Smita for the support and motivation she has given me during this intense period. Other people that I want to mention, who have been present in my life and made it more joyful in this final period, are Robin, Per Asbjørn, Eleni, Elia, and Julia. Two friends from outside of academia who deserve special mentions are Truls and Johan.

Finally I would like to thank some other colleagues and peers who have left me with particularly strong impressions, namely Andrew, Timea (special thanks for helping me with the QFT course!), Leif, Astrid, Harsh, and Patricia. I'm happy we have been part of each others' journies.

Paper 1

I

A Framework for Hadronic Rescattering in pp Collisions

Torbjörn Sjöstrand and Marius Utheim

Eur. Phys. J. C **80**, 907 (2020)
e-Print: `arXiv:2005.05658` [hep-ph]
MCnet–20-11, LU TP 20–12

Abstract

In this article, a framework for hadronic rescattering in the general-purpose Pythia event generator is introduced. The starting point is the recently presented space–time picture of the hadronization process. It is now extended with a tracing of the subsequent motion of the primary hadrons, including both subsequent scattering processes among them and decays of them. The major new component is cross-section parameterizations for a range of possible hadron–hadron combinations, applicable from threshold energies upwards. The production dynamics in these collisions has also been extended to cope with different kinds of low-energy processes. The properties of the model are studied, and some first comparisons with LHC pp data are presented. Whereas it turns out that approximately half of all final particles participated in rescatterings, the net effects in pp events are still rather limited, and only striking in a few distributions. The new code opens up for several future studies, however, such as effects in pA and AA collisions.

1 Introduction

One of the most unexpected discoveries at the LHC is that high-multiplicity pp events bear a striking resemblance to heavy-ion AA events. The first example was the observation of a "ridge", i.e. an enhanced particle production around the azimuthal angle of a trigger jet, stretching away in (pseudo)rapidity [1–3]. Even more spectacular is the smoothly increasing fraction of strange baryon production with increasing charged multiplicity, a trend that lines up with pA data before levelling out at the AA results [4, 5]. Further examples include non-vanishing v_2 azimuthal flow coefficients [2, 3, 6], strong peaks in hadron ratios such as Λ^0/K^0_S at around $p_\perp \approx 2$ GeV [7], and an $\langle p_\perp \rangle$ strongly increasing with particle mass [8], all suggesting some form of collective flow. A recent overview of relevant observations and related theoretical ideas and challenges can be found in Ref. [9].

One possible explanation for these phenomena is that a quark–gluon plasma (QGP) can be created in pp collisions. This runs counter to the conventional wisdom that, unlike in AA collisions, the pp environment does not offer sufficiently large volumes and long time scales for a QGP to form, see e.g. [10–12]. Nevertheless, such models have been developed, for instance the core–corona model implemented in EPOS [13]. In it a lower-density corona of colour strings can hadronize independently, whereas in a higher-density core the strings can melt into a QGP that hadronizes collectively. In its simplest form, a string here represents the colour confinement field between a separated colour triplet–antitriplet pair, typically formed in the collision and thereafter expanding mainly along the collision axis. More central pp collisions correlate both with a higher core fraction and a higher multiplicity, thus offering a mechanism for multiplicity-dependent event properties that can be continued on to AA collisions.

Alternatively, the similarity between pp and AA could be viewed as incentive to explore what phenomena could be explained without recourse to QGP formation. As examples, the formation of ropes with a higher colour charge than the string may explain a changed particle composition [14], while the shoving of overlapping strings can give collective flow [15]. Strings squeezed into a smaller transverse area could also offer a higher string tension and thereby a changed particle composition [16].

Whatever approach is taken, one issue is that both strings and particles are produced very closely packed, in fact physically overlapping to a large extent. This is nothing new, but is already a consequence e.g. of the Pythia model for MultiParton Interactions (MPIs) [17, 18] and the Lund string model view of particle production [19]. The former assumes that several strings are drawn out from a collision area of a typical proton size, and the latter that each of these strings individually has about the same transverse size. Even allowing for the transverse expansion of the string systems, the overlap of fragmenting strings and of primary produced hadrons in pp collisions is alarmingly high [20]. This opens up for the above-mentioned

modifications of the string properties, and would also suggest that hadrons can interact with each other (elastically or inelastically) on the way out from the production region surrounding the primary "scattering". This is what is referred to as hadronic rescattering.

So why has this overlap not attracted attention in traditional high-energy pp generators, such as Herwig [21, 22], Pythia [23, 24] or Sherpa [25, 26]? One practical reason is that close-packing corrections did not seem necessary to describe $\mathrm{pp}/\mathrm{p\bar{p}}$ data up to Tevatron energies, either because they were not there or (more likely) because nobody looked. Concerning rescattering in particular, another is that hadrons produced in a given space–time region of an event also tend to move in the same direction. The most obvious example of this is the ordering in rapidity with respect to the collision axis. This implies that hadronic rescattering tends to occur between pairs of rather low invariant mass and therefore should not upset the overall structure of the event, in particular if hadrons of different species are not distinguished. Furthermore, in high-$p_\perp$ jets the parton-shower evolution spreads out the colour strings, such that overlaps are far less frequent than in the low-$p_\perp$ region [16]. As we will see, rescattering indeed only appears to have a noticeable impact on a select few distributions in pp collisions.

The situation is different in heavy-ion physics, where the hadronic densities could be even higher, and the density drops slower per unit time for a larger expanding system, so there are more opportunities for rescattering on the way out. Several rescattering frameworks have been developed as part of the description of AA collisions, see e.g. the overview and comparison in Ref. [27]. The best known probably is UrQMD [28], which much of our current work is based upon. SMASH [29] is a recent addition still being actively developed. Luciae [30] / Paciae [31] has its roots in Lund, even if now disconnected. Many of these programs make use of Lund string fragmentation.

With the recent implementation of an explicit space–time picture for the hadronization in Pythia [20], it becomes possible to use e.g. UrQMD to simulate rescattering on Pythia generated events. This was recently done [32], with interesting results. Unavoidably it is a kludge, however: while Pythia 8 is written in C++, information has to be transferred to the UrQMD Fortran code, and then UrQMD in turn relies on the older Pythia 6 Fortran version for some tasks. Interfacing SMASH would have the advantage of being able to stay with C++, but again SMASH in its turn makes use of Pythia.

We therefore believe it would be worthwhile to develop and provide a purely internal implementation of hadronic rescattering. In this article we will present such a new framework, and show some of the first results obtained with it. This does not preclude the usage of and comparison with other packages, but rather that interfacing with such packages could be simplified. For instance, one could imagine implementing alternative cross section parameterizations while still retaining the underlying space–time tracing. As part of developing this framework, our work includes implementations of low energy hadron-hadron interac-

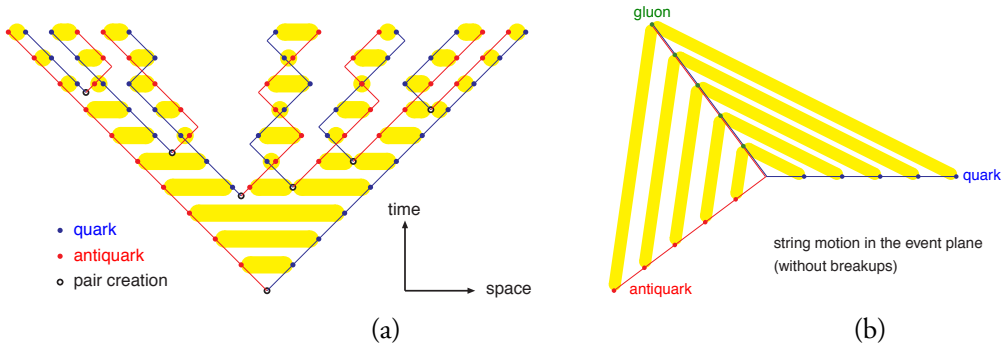


Figure I.1: (a) String breakup in a $q\bar{q}$ event. The points denote the location of quarks and antiquarks at snapshots in time, and the yellow regions the string pieces then stretched out between them. (b) String drawing in the plane of a $q\bar{q}g$ event.

tions. This means event generation in Pythia becomes available for beam energies all the way down to the mass threshold, a feature which may have other applications not related to rescattering.

The outline of this article is as follows. Section 2 reviews the space–time hadron production picture that provides the starting point for the subsequent rescattering. It also describes the algorithm for finding hadronic rescattering vertices and the evolution of the event through the rescattering phase. Section 3 describes the dynamics of low energy processes. This includes how such processes are implemented, and how total, partial and differential cross sections are modelled for the different processes. It represents the bulk of the new features that have been included into Pythia as a result of this work. Then Section 4 presents some model tests and model features, while Section 5 shows some comparisons with experimental data of relevance for the model. Finally Section 6 gives a summary and outlook.

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 space–time model

In this section we will review and extend the space–time picture for hadron production, and present how this picture is used as a starting point to trace collision vertices throughout the time evolution of the event.

2.1 Hadronization

The Lund string model is based on the assumption of linear confinement, 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 simplicity we may consider the process $\mathrm{e}^+\mathrm{e}^- \to \gamma^*/\mathrm{Z}^0 \to \mathrm{q}\bar{\mathrm{q}}$, where the quark–antiquark pair moves out along the $\pm z$ axis, see Figure I.1a. The linearity leads to a straightforward relationship between the energy–momentum and the space–time pictures:

$$\left|\frac{\mathrm{d}p_{z,\mathrm{q}/\bar{\mathrm{q}}}}{\mathrm{d}t}\right| = \left|\frac{\mathrm{d}p_{z,\mathrm{q}/\bar{\mathrm{q}}}}{\mathrm{d}z}\right| = \left|\frac{\mathrm{d}E_{\mathrm{q}/\bar{\mathrm{q}}}}{\mathrm{d}t}\right| = \left|\frac{\mathrm{d}E_{\mathrm{q}/\bar{\mathrm{q}}}}{\mathrm{d}z}\right| = \kappa \; . \tag{I.1}$$

It is necessary to keep track of signs: as the q-to-$\bar{\mathrm{q}}$ separation increases their energies decrease, with more and more of the energy instead stored in the intermediary string. At the maximal separation there would be no energy left for the quarks, and the string tension would then start to pull them together again, so that they would perform an oscillatory motion often referred to as a "yo-yo" motion.

If there is enough energy, the string between an original $\mathrm{q}_0\bar{\mathrm{q}}_0$ pair may break by producing new $\mathrm{q}_i\bar{\mathrm{q}}_i$ pairs, where the intermediate q_i ($\bar{\mathrm{q}}_i$) are pulled towards the $\bar{\mathrm{q}}_0$ (q_0) end, such that the original colour field is screened. This way the system breaks up into a set of n colour singlets $\mathrm{q}_0\bar{\mathrm{q}}_1 - \mathrm{q}_1\bar{\mathrm{q}}_2 - \mathrm{q}_2\bar{\mathrm{q}}_3 - \ldots - \mathrm{q}_{n-1}\bar{\mathrm{q}}_0$, that we can associate with the primary hadrons. Each $\mathrm{q}_i\bar{\mathrm{q}}_i$ pair is produced with zero energy and momentum at its common vertex, since the string does not contain any local concentrations of energy. The energy and momentum of a hadron $\mathrm{h}_i = \mathrm{q}_i\bar{\mathrm{q}}_{i+1}$ therefore is provided by the string intermediate to the $\mathrm{q}_i\bar{\mathrm{q}}_i$ and $\mathrm{q}_{i+1}\bar{\mathrm{q}}_{i+1}$ breaks. This gives $E_{\mathrm{h}_i} = \kappa(z_i - z_{i+1})$ and $p_{z,\mathrm{h}_i} = \kappa(t_i - t_{i+1})$. Note that $z_i > z_{i+1}$ since q_0 is moving in the $+z$ direction. If boosted to a frame where $t'_i = t'_{i+1}$, i.e. where the hadron is at rest, one obtains $m_{\mathrm{h}_i} = E'_{\mathrm{h}_i} = \kappa(z'_i - z'_{i+1})$.

Unlike the intermediate vertices, the $\mathrm{q}_0\bar{\mathrm{q}}_0$ pair starts with non-vanishing energy at the origin. The equivalent vertex for the q_0 instead is where it has lost its energy, which (in the massless approximation) occurs at $t = z = E_{\mathrm{q}_0}(t = 0)/\kappa$. This vertex can be used as the starting point for a recursive procedure, where the location of each consecutive vertex can be reconstructed from the E and p_z of the intermediate hadron. Knowing the momenta of all hadrons it is therefore possible to reconstruct all $\mathrm{q}_i\bar{\mathrm{q}}_i$ production vertices, or the other way around. Hadrons do not have a unique definition of a production "vertex", being extended objects, but a convenient choice is the average of the $\mathrm{q}_i\bar{\mathrm{q}}_i$ ones on either side of it [20]. Alternatives include an early or late choice, where the backward or forward light cones of the two $\mathrm{q}_i\bar{\mathrm{q}}_i$ vertices cross.

Several issues have here been swept under the carpet, since they do not directly affect the key relationship between the energy–momentum and the space–time pictures. One issue is that quarks with non-vanishing mass or $p_\perp$ should move along hyperbolae $E^2 - p_z^2 = m^2 + p_\perp^2 =$

$m_\perp^2$. When produced inside a string 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, e.g. by replacing m by $m_\perp$ in relations between E and p_z. Another issue is that the above notation only allows for meson production. 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, and there is no perfect way around that. Smearing factors will be introduced to largely remove the tension for the transverse degrees of freedom, and somewhat reduce it for the other ones. Either way, this semiclassical model does not introduce any clear systematic biases. Hence, there is no big problem in practice, since we are interested in average effects obtained by Monte Carlo sampling over a wide range of possible early histories.

The real practical hurdle is to go on from a simple straight string to a larger string system. Consider e.g. $\mathrm{e^+e^-} \to \gamma^*/\mathrm{Z}^0 \to \mathrm{q\bar{q}g}$. In the limit where the number of colours is large, the $N_C \to \infty$ approximation [33], a string will be stretched from the colour of the q to the anticolour of the g, and then on from the colour of the g to the anticolour of the $\bar{\mathrm{q}}$, Fig. I.1b. To first approximation the two string pieces each could be viewed as a boosted copy of a simple $\mathrm{q\bar{q}}$ system. The problems arise around the gluon kink, as follows. We already noted that a $\mathrm{q}/\bar{\mathrm{q}}$ turns around when it has lost its energy. When the same thing happens for a gluon, however, it is instead replaced by a new expanding string region made out of inflowing momentum from the q and $\bar{\mathrm{q}}$. Therefore there are actually three string regions in which breaks can occur, and the third one is especially important in the limit of a low-energy gluon. Note that QCD favours the emission of soft gluons, and that additionally a gluon is pulling out two string pieces and therefore loses energy twice as fast as a quark, so such third regions contribute a fair fraction of all hadron production. For systems with more than one intermediate gluon the string motion becomes even more complicated.

A framework to handle energy and momentum sharing in such complicated topologies was developed in Ref. [34], and was then extended to reconstruct matching space–time production vertices in [20]. (An earlier extension in [35] included several of the same main features, but could not handle as complicated systems as required for LHC applications.) Again it can be described as a recursive procedure, starting from one end of the string system, but now with additional rules how to pass from one string region to the next. The reader is referred to Ref. [20] for details.

In addition to the main group of open strings stretched between $\mathrm{q\bar{q}}$ endpoints, there are

two other common string topologies. One is a closed gluon loop, which can be viewed as an open string (with at least one intermediate gluon) where the q and $\bar{\mathrm{q}}$ endpoints are fused into a single gluon, which closes the colour flow. Once an initial $\mathrm{q}_0\bar{\mathrm{q}}_0$ breakup has been picked somewhere along the string, at random (within given rules), the further handling devolves back into the open string framework. 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. The junction moves by the net pull of the string, and is at rest only in a frame where the opening angle between each quark pair is 120°. Also in this case there may be gluons on the string between a quark and the junction. Each of the three legs may be hadronized according to the same basic rules as above, with some special care needed where they meet at the junction, around which a baryon is formed to carry the net baryon number of the system.

There is one further aspect added to the framework presented so far. For the energy–momentum picture in a $\mathrm{q}\bar{\mathrm{q}}$ system we started out with a pure two-dimensional representation in (E, p_z) space, but then added random Gaussian $p_\perp$ kicks motivated by the tunnelling mechanism. Alternatively we could have motivated such fluctuations by the uncertainty relationship: a string could be expected to have a radius roughly $\sqrt{2/3}$ that of the proton, since if $r_\mathrm{p}^2 = \langle x^2 + y^2 + z^2 \rangle$ then $\langle x^2 + y^2 \rangle = (2/3) r_\mathrm{p}^2$. Either argument gives $p_\perp$ kicks of the order 0.3 GeV for each $\mathrm{q}_i\bar{\mathrm{q}}_i$ pair, consistent with data. By contrast, the basic machinery sets all $\mathrm{q}_i\bar{\mathrm{q}}_i$ production vertices to have $x = y = 0$, which gives an unreasonably perfect lineup of the hadrons. For the studies in [20] we therefore introduced a Gaussian (x, y) smearing with a width according to the expressions above, and will continue to do so. By the additional smearing to be introduced in the next section, which partially might overlap, some reduction of the width would be motivated, however.

Unfortunately, complications may arise in multiparton systems, notably for those hadrons that have their two defining $\mathrm{q}_i\bar{\mathrm{q}}_i$ vertices in two different string regions, meaning there is no unique separation between transverse and longitudinal degrees of freedom. Occasionally this may give unreasonably large positive or negative $\tau^2 = t^2 - x^2 - y^2 - z^2$. A few safety checks have been introduced to catch and correct such mishaps as well as possible.

2.2 Multiparton interaction vertices

The framework described above assumes that all partons start out from the same space–time production vertex, as would be the case e.g. in $\mathrm{e}^+\mathrm{e}^- \to \mathrm{Z}^0 \to \mathrm{q}\bar{\mathrm{q}}$. In pp the colliding hadrons are extended objects, however. The Lorentz-contracted hadrons pass through each other at a fairly well-defined time, conventionally $t = 0$, but over a transverse region of hadronic sizes. In the overlap region several parton-parton interactions can occur, as described by the MPI framework in Pythia [17, 18].

The probability for an interaction at a given transverse coordinate (x, y) can be assumed related to the time-integrated overlap of the parton densities of the colliding hadrons in that area element. Let the partons be described by a Lorentz contracted probability distribution $P_{\mathrm{LC}}(x, y, z)$, which in its rest frame reduces to a spherically symmetric $P(r)$ with $r^2 = x^2 + y^2 + z^2$. Setting the two incoming beam particles A and B to move along the z axis with velocity $\pm v$, separated by $\pm b/2$ in the x direction, where b is the impact parameter, this overlap ("eikonal") reads

$$\begin{aligned}\mathcal{O}(x, y; b) &\propto \int\int P_{\mathrm{LC},A}\left(x - \frac{b}{2}, y, z - vt\right) P_{\mathrm{LC},B}\left(x + \frac{b}{2}, y, z + vt\right) \mathrm{d}z\,\mathrm{d}t \\ &\propto \int P_A\left(x - \frac{b}{2}, y, z_A\right) \mathrm{d}z_A \int P_B\left(x - \frac{b}{2}, y, z_B\right) \mathrm{d}z_B\,, \end{aligned} \tag{I.2}$$

the latter by suitable variable transformation. The answer can be further simplified in case of a Gaussian distribution $P(r) \propto \exp(-r^2/r_0^2)$:

$$\begin{aligned}\mathcal{O}(x, y; b) &\propto \int \exp\left(\frac{(x - b/2)^2 + y^2 + z_A^2)}{r_0^2}\right) \mathrm{d}z_A \\ &\quad \times \int \exp\left(\frac{(x + b/2)^2 + y^2 + z_B^2)}{r_0^2}\right) \mathrm{d}z_B \\ &\propto \exp\left(-\frac{2r_\perp^2}{r_0^2}\right) \exp\left(-\frac{\mathrm{b}^2}{2r_0^2}\right)\,, \end{aligned} \tag{I.3}$$

where $r_\perp^2 = x^2 + y^2$. That is, for a Gaussian proton the overlap region is an azimuthally symmetric Gaussian, with no memory of the collision plane, and the total overlap is a Gaussian in b. The r_0 parameter can be approximately related to the proton radius r_{p} by $\langle r^2 \rangle = \langle x^2 + y^2 + z^2 \rangle = 3r_0^2/2 = r_{\mathrm{p}}^2$. The default in Pythia is a constant proton radius value $r_{\mathrm{p}} \approx 0.85$ fm for the distribution of partons. With increasing energy, and a related increase in the number of MPIs per collision, the effective edge of interacting partons is pushed outwards and thus collision cross sections can go up.

The Gaussian is a very special case, 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, $P(r) \propto \Theta(r_0 - r)$, where Θ is the step function, 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, $P(r) \propto \exp(-r/r_0)$.

In the Pythia MPI machinery the overlap distribution $\mathcal{O}(b) = \int\int \mathcal{O}(x, y; b)\,\mathrm{d}x\,\mathrm{d}y$ can be chosen and tuned according to a few different forms. The current default is $\mathcal{O}(b) \propto \exp((b/b_0)^p)$ with $p = 1.85$, i.e. close to but not quite Gaussian. A similar shape and tune is obtained with a double Gaussian $P(r)$, where a smaller-radius second Gaussian can be

viewed as representing hot spots inside the proton. In both cases a stronger-than-Gaussian peaking of $\mathcal{O}(b)$ at $b = 0$ is required to get a sufficiently long tail out to largest charged multiplicities in LHC and Tevatron minimum-bias events.

The $P(r)$ and $\mathcal{O}(b)$ distributions as described so far are likely to be significant simplifications, however. If one views the evolution from a simple original parton configuration via initial-state cascades into a set of interacting partons, then there are likely to arise complicated patterns and correlations. One such framework is presented in Ref. [36], where an implementation of Mueller's dipole model [37, 38] for the two colliding hadrons are used to assign MPI production vertices. These then turn out to give clearly non-isotropic distributions. In the future the relevant code for these assignments will be made available, but using it comes at a cost in terms of a considerably slower event generation.

For now, we have therefore settled for a simplified framework with enough flexibility for our purposes. In it the MPIs locations by default are selected according to the Gaussian $\exp(-2r_\perp^2/r_0^2)$, but optionally this can be modified in either of two ways. Either the x coordinates are scaled by a factor r_ϵ and the y ones by $1/r_\epsilon$, or else the Gaussian is multiplied by a φ modulation factor

$$\frac{\mathrm{d}N}{\mathrm{d}\varphi} \propto 1 + \epsilon\cos(2\varphi) \; . \tag{I.4}$$

Here $r_\epsilon > 1$ or $\epsilon > 0$ means an enhancement in the collision plane and $r_\epsilon < 1$ or $\epsilon < 0$ out of it. Asymmetries in the spatial distribution also arise from the Monte Carlo sampling of a finite number of MPIs, and these may be even more important.

This machinery is used to select the (x, y) coordinates of the MPI vertices at $t = z = 0$. Only a fraction of the full beam-particle momentum is carried away by the MPIs, leaving behind one or more beam remnants [39]. These are initially distributed according to the basic $\exp(-r_\perp^2/r_0^2)$ shape around the center of the respective beam. By the random fluctuations, and by the interacting partons primarily being selected on the side leaning towards the other beam particle, the "center of gravity" will not be located at the $x = \pm b/2, y = 0$ positions originally assumed. 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. As a small improvement on a uniform shift, remnants located closer to the other remnant are shifted more, so as to deplete the overlap region more. This is achieved by assigning each remnant a weight

$$\left(1 + \frac{b}{r_\mathrm{p}} \exp\left(\frac{\pm x}{r_\mathrm{p}}\right)\right)^{-1} \tag{I.5}$$

proportional to its eventual shift, where x is relative to the respective beam center with the other beam displaced $\mp b$ in the x direction. 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 [20]. We will 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. Specifically, a radiated parton is assigned a location at $t = 0$ that is smeared by $\Delta r_\perp$ relative to its mother parton according to a two-dimensional Gaussian with a width inversely proportional to its $p_\perp$. The constant of proportionality can be set freely, but should obviously be such that $\Delta r_\perp \, p_\perp \sim \hbar$. So as not to obtain unreasonable $\Delta r_\perp$ shifts, the $p_\perp$ is set to be at least 0.5 GeV in this context, comparable to the cut-off scale of the FSR showers. 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 $N_C \to \infty$ 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. 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 of the character already described in Section 2.1.

There is one key difference, however, namely that the strings now can be stretched between partons that do not originate from the same vertex. Even in the simplest case, a q connected with a $\bar{\mathrm{q}}$ from a different MPI, there is a new situation not studied previously, where the vertex separation should be equivalent to a piece of string already at $t = 0$. For the energy–momentum picture it is traditionally assumed that its effects are sufficiently small that they can be neglected. If the effects of a 1 fm $\approx$ 1 GeV special term is to be spread over many hadrons, then the net effect on each hardly would be noticeable.

For the space–time picture we do want to be more careful about the effects of the transverse size of the original source. The bulk of the effects determining the hadronic production vertices do come from the framework of Section 2.1, and therefore we will be satisfied if we can introduce a relevant amount of smearing on hadron production, without necessarily fully describe effects for the individual hadron. This is achieved as follows.

For a simple $\mathrm{q}\bar{\mathrm{q}}$ string, such as in Figure I.1a, the relevant length of each hadron string piece is related to its energy. For a given hadron, define E_{hq} ($E_{\mathrm{h}\bar{\mathrm{q}}}$) as half the energy of the hadron plus the full energy of all hadrons lying between it and the q ($\bar{\mathrm{q}}$) end, and use this as a measure of how closely associated a hadron is with the respective endpoint. Also let

$\mathbf{r}_{\perp \mathrm{q}}$ ($\mathbf{r}_{\perp \bar{\mathrm{q}}}$) be the (anti)quark transverse production coordinates. Then define the hadron production vertex offset to be

$$\Delta \mathbf{r}_{\perp \mathrm{h}} = \frac{E_{\mathrm{h}\bar{\mathrm{q}}}\, \mathbf{r}_{\perp \mathrm{q}} + E_{\mathrm{hq}}\, \mathbf{r}_{\perp \bar{\mathrm{q}}}}{E_{\mathrm{hq}} + E_{\mathrm{h}\bar{\mathrm{q}}}} = \frac{(E_{\mathrm{tot}} - E_{\mathrm{hq}})\, \mathbf{r}_{\perp \mathrm{q}} + E_{\mathrm{hq}}\, \mathbf{r}_{\perp \bar{\mathrm{q}}}}{E_{\mathrm{tot}}} \,, \tag{I.6}$$

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

This procedure is then generalized to more complicated string topologies. In a $\mathrm{q} - \mathrm{g}_1 - \mathrm{g}_2 - \ldots - \bar{\mathrm{q}}$ string, one may define E_{hq} as above. If $E_{\mathrm{hq}} < E_{\mathrm{q}} + E_{\mathrm{g}_1}/2$ the hadron is viewed as produced between the q and g_1, and the offset can be found as above, only with $E_{\bar{\mathrm{q}}}$ replaced by $E_{\mathrm{g}_1}/2$. If instead $E_{\mathrm{q}} + E_{\mathrm{g}_1}/2 < E_{\mathrm{hq}} < E_{\mathrm{q}} + E_{\mathrm{g}_1} + E_{\mathrm{g}_2}/2$ then the excess energy $E_{\mathrm{hq}} - E_{\mathrm{q}} - E_{\mathrm{g}_1}/2$ determines the admixture of $\mathbf{r}_{\perp \mathrm{g}_1}$ and $\mathbf{r}_{\perp \mathrm{g}_2}$, and so on, stepping through region after region, for hadron after hadron, until the $\bar{\mathrm{q}}$ end is reached. For junction topologies the same kind of approach can be used to iterate from each leg towards the central junction. The two lowest-energy legs are considered first, and an $\mathbf{r}_\perp$ towards which the third string is iterated is formed by the relative unused energy fractions of the first two. That way a junction baryon can receive contributions from all three legs.

There are two obvious shortcomings. Firstly, the approach does not take into account the higher regions, handled in the complete string motion, e.g. made up out of q and g_2 momentum, where the hadron offset could be a more complex combination of three different parton offsets. Secondly the sharing according to energy is not Lorentz covariant. Nevertheless, we believe this approach to provide a sensible approximation to the smearing effects one may expect. There is also a third, less obvious problem, namely what to do with closed gluon loops. There the hadronization is begun at a random point, where the location of this point currently is not stored anywhere. The algorithm as presented so far will start at another point and therefore give a mismatch. We have not considered this a big issue for now, since the default CR algorithm will dissolve almost all such closed loops, and again the key issue is to provide some relevant amount of smearing without attaching too deep a meaning to each separate correction to the dominant hadronization picture.

2.3 The space–time picture of hadronic rescattering

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 potential collision is studied in the center-of-momentum frame of the two particles, with motion

along the $\pm z$ direction, i.e.

$$\begin{aligned} p_1 &= (E_1, 0, 0, p) \ , \\ p_2 &= (E_2, 0, 0, -p) \ . \end{aligned} \tag{I.7}$$

If they are not produced at the same time, the position of the earlier particle is offset to the creation time of the later particle. Particles moving away from each other already at this common time, i.e. with $z_1 > z_2$, 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, and the two particle species. There is no solid theory for the b dependence of P, so we will consider two different shapes. The default model is a Gaussian dependency,

$$P(b) = P_0 e^{-b^2/b_0^2}, \tag{I.8}$$

where P_0 is referred to as the opacity, a free parameter that is 0.75 by default, and the characteristic length scale is

$$b_0 = \sqrt{\frac{\sigma}{P_0 \pi}} \ , \tag{I.9}$$

where σ is the cross section. It is assumed that the only dependency on the energy and the particle species is through σ, which will be discussed in great detail in Section 3. Typical values of b_0 are around 1-2 fm for the most common processes. An alternative model is a grey disk with interaction probability

$$P(b) = P_0 \, \Theta(b - b_0), \tag{I.10}$$

where Θ is the Heaviside step function. The $P_0 = 1$ case gives the often-used black disk limit. In both these cases, the parameter b_0 is chosen so that

$$\int_0^\infty 2\pi b \, P(b) \, \mathrm{d}b = \sigma \ . \tag{I.11}$$

This normalization ensures that if b is chosen uniformly on a large disk, the total probability of an interaction is the same for both models. In reality, with a finite effective region, one may expect the Gaussian shape to give fewer scatterings.

If it is determined that the particles will interact, the interaction time is defined as the time of closest approach in the rest frame. 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 an effective delay before the new hadrons are formed and can begin to interact. For the other processes, such as elastic scattering or an intermediate resonance decay, there is the option to have effective formation times before new interactions are allowed. One reason for why one would want this is that it takes some time for the new hadrons to break free from the volume formerly occupied by the mothers and form their own new (spatial) wave functions.

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. During rescattering, unstable particles can decay, with the fastest-decaying ones having lifetimes comparable to the timescales of rescattering. For these particles, an invariant lifetime τ is picked at random according to an exponential $\exp(-\tau/\tau_0)$, where $\tau_0 = 1/\Gamma$ is the inverse of the width. This is done for each short-lived hadron, and 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:

Firstly, the procedure is not Lorentz invariant, since the time-ordering of interactions is defined on the lab frame of the full collision, i.e. the CM frame for LHC events. We do not expect this to be a major issue: even if the time ordering would change depending on the frame chosen, it would not matter in choosing between two potential interactions with a spacelike separation, and only for a fraction of those with a timelike one. This has been studied and confirmed within existing rescattering approaches [28, 29, 40]. We will also present a check in Section 4.4, where we confirm that the effect on observable quantities is negligible. More consistent time orderings have been proposed [41, 42], but are nontrivial to implement and have not been considered here.

Secondly, 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. If one considers a closed system in thermal equilibrium, where $2 \to n$ processes are allowed, indeed $n \to 2$ at commensurate rates would be a natural ingredient to maintain that balance. The system is rapidly expanding in pp collisions, so for our current studies it should not be a big issue. One place where it could make a difference is in baryon rates, where pair annihilation outweighs pair creation within the current setup. In the future $3 \to n$ collisions could be identified by isolating cases where a hadron has two very closely

separated potential $2 \to n$ interactions, which then could be joined into one. This would also introduce an alternative argument for a formation time, as the borderline between separated and joined processes.

Thirdly, introducing rescattering will change the shape of events, which of course is the point of the exercise, but it also affects distributions we do not want to change. One example, related to the second limitation above, is that the charged multiplicity will increase, which has to be compensated by a tuning of other parameters. In this article only a simple retune is made specifically for pp. More properly one should go back to $\mathrm{e^+e^-}$ annihilation events and retune the fragmentation of a simple string there, with rescattering effects included, before proceeding to pp. In $\mathrm{e^+e^-} \to \mathrm{Z^0} \to \mathrm{q\bar{q}}$ events, however, the bulk of rescattering should be related to nearest neighbours in rank, i.e. in order along the string. So, if such rescatterings are not simulated, then fragmentation parameters should not have to be changed significantly. A shortcut to avoid a bigger retune therefore is to forbid nearest-rank neighbours from rescattering also in pp events, and this is one model variation we will consider.

Fourthly, all possible subprocesses are assumed to share the same 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.

Finally, the model only considers the effect of hadrons colliding with hadrons, not those of strings colliding/overlapping with each other or with hadrons. The former is actively being studied within Pythia, as a shoving/repulsion of strings [15, 43]. Both shove and rescattering act to correlate the spatial location of strings/hadrons with a net push outwards, giving rise to a radial flow. In reality the two could be combined, with shove acting before hadronization and rescattering after. The two effects do not add linearly, however, since an early shove leads to a more dilute system of strings and primary hadrons, and thereby less rescattering. Thus it will become a nontrivial task to distinguish the effects of the two possible phenomena, not made any simpler if also string–hadron interactions were to be included in the mix.

3 The hadronic rescattering model

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 the total cross section into a sum of partial cross sections, one for each possible process, as these are used to represent relative frequencies for each process to occur. In this section, we discuss the possible processes we have implemented in our framework, including how their partial

cross sections are calculated, and how those processes are simulated.

As we will see, 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^*. The possible processes that can occur depend heavily on the particle types involved. In our model, the following types of processes are available:

- Elastic interactions are ones where the particles do not change species, i.e. $AB \to AB$. In our implementation, these are considered different from elastic scattering through a resonance, e.g. $\pi^+\pi^- \to \rho^0 \to \pi^+\pi^-$ (in reality there are likely to be interference terms that make this separation ambiguous). In experiments, usually all $AB \to AB$ events are called elastic because it is not possible to tell which underlying mechanism was involved. Therefore, when comparing with data for elastic cross sections, we do include contributions from resonance formation.

- Resonance formation typically can be written as $AB \to R \to CD$, 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. This is assumed to require the annihilation of at least one $\mathrm{q\bar{q}}$ 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 $AB \to XB$ or $AB \to AX$ and double $AB \to X_1X_2$. Here X represents a massive excited state of the respective incoming hadron, and there is no net colour 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 $AB \to A^*B$, $AB \to AB^*$ or $AB \to A^*B^*$. 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.

- Nondiffractive 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.

All total and partial cross sections have a nontrivial energy dependence. Whereas we have made an effort to cover a fair amount of detail, it is not feasible to give all processes full attention in the first release of this framework, not even in the proportionately few cases where experimental data exist. Our hope is that since rescatterings will not be observable on an individual basis and instead the average effects they induce is what will be of interest, we can live with imperfections here and there so long as they do not generate non-negligible systematic biases. Refinements could be introduced over time without affecting the rescattering machinery as such. In Section 4.5 we will study the rates of different particle types participating in rescattering and at which energies most interactions occur, giving an indication of which cross sections are the most important for future refinement.

In the continued discussion, some common simplifications should be noted.

- Cross sections are invariant when all particles are replaced by their antiparticles. Whenever we talk about any particular cross section for two particles, it is always implicit that the exact same procedure is used to calculate the cross section for their antiparticles.

- Many measured cross sections approximately scale in accordance with the Additive Quark Model (AQM) [44, 45], i.e. 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 proportional to their constituent masses, this gives an effective number of interacting quarks in a hadron of approximately

$$n_{\mathrm{q,AQM}} = n_{\mathrm{u}} + n_{\mathrm{d}} + 0.6\, n_{\mathrm{s}} + 0.2\, n_{\mathrm{c}} + 0.07\, n_{\mathrm{b}} \,. \tag{I.12}$$

  For lack of alternatives, many unmeasured cross sections are assumed to scale in proportion to this.

- The neutral Kaon system is nontrivial, with strong interactions described by the $\mathrm{K}^0/\bar{\mathrm{K}}^0$ states and weak decays by the $\mathrm{K}^0_{\mathrm{S}}/\mathrm{K}^0_{\mathrm{L}}$ ones. The oscillation time is of the order of the $\mathrm{K}^0_{\mathrm{S}}$ lifetime, far above the rescattering scales of interest in this article. Therefore an intermediate "decay" invariant time of 10^9 fm has been introduced for $\mathrm{K}^0/\bar{\mathrm{K}}^0 \to \mathrm{K}^0_{\mathrm{S}}/\mathrm{K}^0_{\mathrm{L}}$, well above hadronization scales but also well below decay ones. While the bulk of Kaon production is into the strong eigenstates, a fraction is into the weak ones, such as $\phi \to \mathrm{K}^0_{\mathrm{S}}\,\mathrm{K}^0_{\mathrm{L}}$. Cross sections for $\mathrm{K}^0_S/\mathrm{K}^0_{\mathrm{L}}$ with a hadron are given by the mean of the cross section for K^0 and $\bar{\mathrm{K}}^0$ with that hadron. When the collision occurs, the $\mathrm{K}_{\mathrm{S,L}}$ is converted into either K^0 or $\bar{\mathrm{K}}^0$, where the probability for each is proportional to the total cross section for the interaction with that particle.

Finally, keep in mind that we here concern ourselves with cross sections for collisions at low CM energies, with most rescatterings occurring below 2 GeV, and very few above 5 GeV, as we will see.

3.1 Total cross sections

The total cross section is needed by the rescattering algorithm to determine how close two hadrons need to be to interact. In the rescattering algorithm, each hadron pair (including the products of rescatterings) is checked for potential interactions, and thus naively $\mathcal{O}(n^2_{\text{primary}})$ total cross sections must be calculated. Quick checks that can exclude a fair fraction of all pairs at an early stage are essential to keep time consumption at a manageable level. In particular, we have made an effort to ensure that total cross sections can be calculated efficiently, and that partial cross sections are only calculated for a hadron pair when it has been determined that they should interact.

A brief summary of total cross sections is provided in Table I.1. Figure I.2 shows the total and elastic cross sections for some important processes where PDG data is available [46].

Table I.1: Summary of total cross section descriptions. Here, N is used to denote a nucleon (p or n), B a baryon and M a meson.

Case	Method
NN, < 5 GeV	Fit to data
NN, > 5 GeV	HPR_1R_2 parameterization
Other BB	AQM (UrQMD) parameterization
$\mathrm{p\bar{p}}$, < 5 GeV	Ad hoc parameterization
$\mathrm{p\bar{p}}$, > 5 GeV	HPR_1R_2 parameterization
Other $\mathrm{B\bar{B}}$	AQM rescaling of $\mathrm{p\bar{p}}$
$\pi\pi$ and $\mathrm{K}\pi$	Parameterization based on [47, 48] and [49]
NK^-, $\mathrm{N\bar{K}}^0$	Resonances + ad hoc parameterization
NK^+, NK^0	Ad hoc parameterization
MB/MM with resonances	Resonances + elastic
Other MB/MM	HPR_1R_2 if available, otherwise AQM

Baryon-baryon

For NN collisions below 5 GeV, the total cross section is found by an interpolation of experimental data [46]. The nn cross section is taken to be the same as the pp one. Above 5 GeV, the cross section is found using the HPR_1R_2 parameterization [46],

$$\sigma_{\text{tot}} = P + H\log^2\left(\frac{s}{s_0}\right) + R_1\left(\frac{s}{s_0}\right)^{\eta_1} + R_2\left(\frac{s}{s_0}\right)^{\eta_2}, \tag{I.13}$$

where:

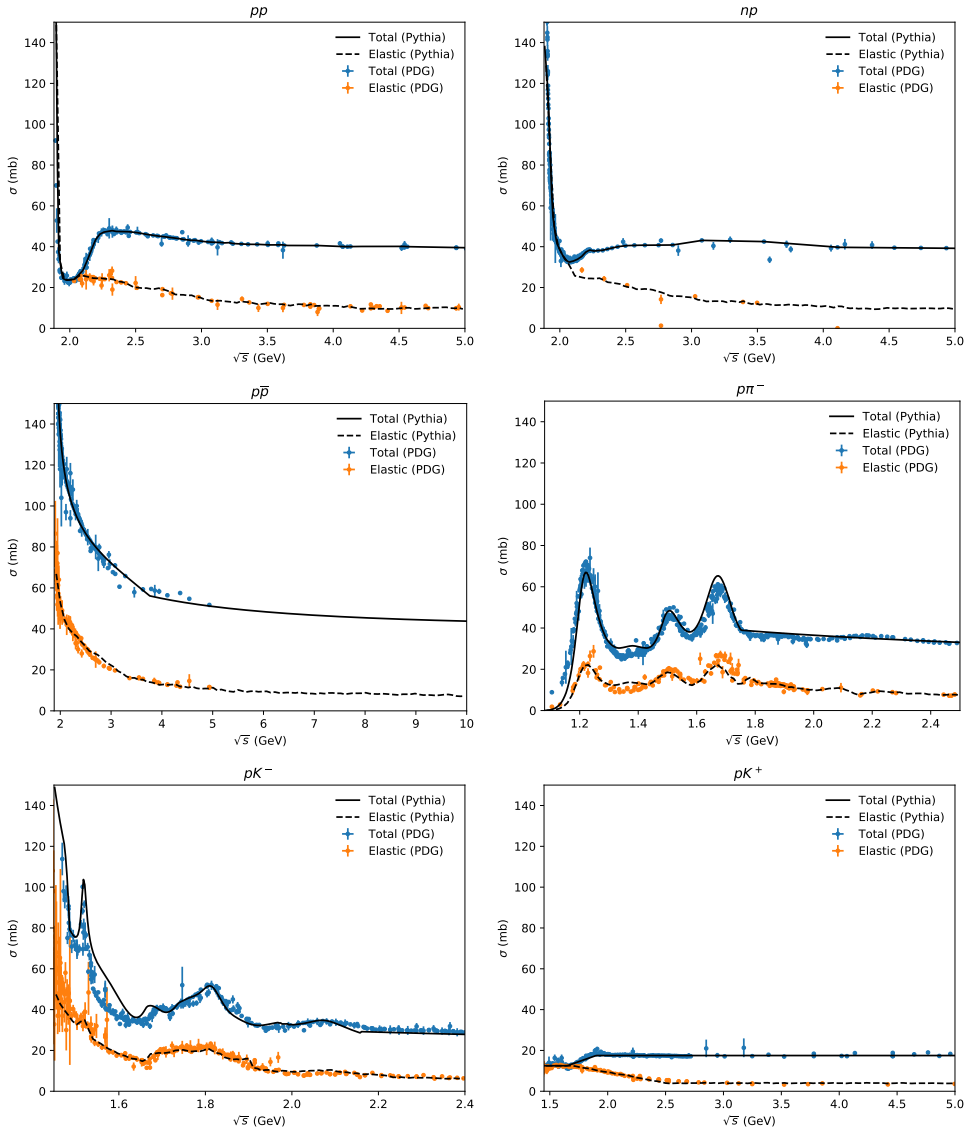


Figure I.2: Total and elastic cross sections for some important processes. The elastic cross sections for $\mathrm{p}\pi^-$ and pK^- include elastic scattering through a resonance, $AB \to R \to AB$, which notably do not correspond to the elastic cross sections calculated in Section 3.2.

- P, R_1 and R_2 depend on the specific particle species, as shown in Table I.2.
- s_0 depends on the masses of A and B and is given by $(m_A + m_B + M)^2$, where $M = 2.1206$ GeV is a constant.

Table I.2: Parameters for the HPR_1R_2 parameterization, for processes used in our rescattering framework. All numbers are in units of mb. N stands for either p or n and $\bar{\mathrm{K}}$ stands for either K^- or $\bar{\mathrm{K}}^0$.

Process	P	R_1	R_2
pp/nn	34.41	13.07	-7.394
pn	34.71	12.52	6.66
$\bar{\mathrm{p}}\mathrm{p}$	34.41	13.07	7.394
$\mathrm{N}\pi^{\mp}$	18.75	9.56	$\pm$1.767
$\mathrm{p}\bar{\mathrm{K}}$	16.36	4.29	3.408
$\mathrm{n}\bar{\mathrm{K}}$	16.31	3.70	1.826

- $H = \pi(\hbar c)^2/M^2 = 0.2720$ mb, $\eta_1 = 0.4473$ and $\eta_2 = 0.5486$ are constants.

In other baryon–baryon cases, the cross section is found using the AQM ansatz as

$$\sigma_{\mathrm{AQM},AB} = (40\ \mathrm{mb})\ \frac{n_{\mathrm{q,AQM},A}}{3}\ \frac{n_{\mathrm{q,AQM},B}}{3}\ . \tag{I.14}$$

Baryon-antibaryon

For $\mathrm{B}\bar{\mathrm{B}}$, we parameterize the cross section as a function of the absolute value of the center-of-mass momentum p_{CM} of the colliding hadrons. For $\mathrm{p}\bar{\mathrm{p}}$ below $p_{\mathrm{CM}} < 6.5$ GeV, we use the UrQMD parameterization [28]:

$$\sigma_{\mathrm{tot}}(\mathrm{p}\bar{\mathrm{p}}) = \begin{cases} 271.6e^{-1.1\,p^2}\ , & p < 0.3\ , \\ 75.0 + 43.1\,p^{-1} + 2.6\,p^{-2} - 3.9\,p\ , & 0.3 < p < 6.5\ , \end{cases} \tag{I.15}$$

For $p_{\mathrm{CM}} > 6.5$ GeV, we use HPR_1R_2. The boundary at 6.5 GeV has been chosen to give a smooth transition between the two regions, and is slightly different from the boundary at 5 GeV used by UrQMD. For all other baryon-antibaryon interactions, the total cross section is found using the same parameterization, but rescaling by an AQM factor,

$$\sigma_{\mathrm{tot}}(\mathrm{B}\bar{\mathrm{B}}) = \frac{\sigma_{\mathrm{AQM,B}\bar{\mathrm{B}}}}{\sigma_{\mathrm{AQM,p}\bar{\mathrm{p}}}}\,\sigma_{\mathrm{tot}}(\mathrm{p}\bar{\mathrm{p}})\ , \tag{I.16}$$

where σ_{AQM} is given in eq. (I.14).

In some cases no quarks can annihilate, e.g. for $\Delta^{++}(\mathrm{uuu}) + \overline{\Delta}^{+}(\bar{\mathrm{d}}\bar{\mathrm{d}}\bar{\mathrm{d}})$. In these cases, the annihilation cross section (see Section 3.4) is subtracted from the total one.

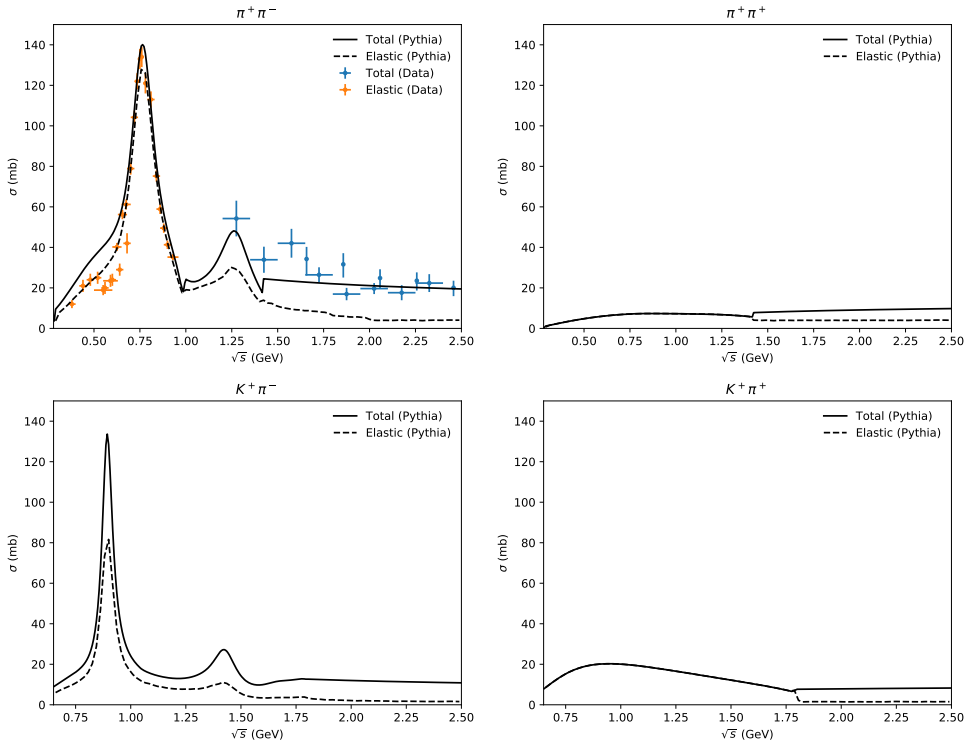


Figure I.3: Total and elastic cross sections for $\pi\pi$ and Kπ interactions. We see that resonances exist for $\pi^+\pi^-$ and $K^+\pi^-$, but not for $\pi^+\pi^+$ and $K^+\pi^+$. The elastic cross sections include cross sections for elastic scattering through a resonance. For $\pi^+\pi^-$, the elastic data comes from [50, 51] and total data comes from [52, 53]. Note that in some theory calculations the concept of elastic is extended to related processes, e.g. $\pi^+\pi^- \to \pi^0\pi^0$ may count as part of a broader $\pi\pi \to \pi\pi$ "elastic" process. If we had taken that viewpoint, the elastic cross sections for $\pi^+\pi^-$ and $K^+\pi^-$ would have equalled the total cross section at low energies.

Meson-hadron

The most common meson-meson interactions are $\pi\pi$ and Kπ. In these two cases, the total cross sections are found using the calculations of Peláez et al. [47–49]. Below 1.42 GeV for $\pi\pi$ and below 1.8 GeV for Kπ, values of the total cross sections have been tabulated and are found using interpolation, for the sake of efficiency. Above these thresholds, the cross section is parameterized as

$$\sigma_{\text{tot}}(AB) = \frac{4\pi^2\left(\beta_P s + \beta_\rho s^{\alpha_\rho} + \beta_2 s^{\alpha_{R2}}\right)}{\sqrt{(s-(m_A-m_B)^2)(s-(m_A+m_B)^2)}}, \tag{I.17}$$

where, $\alpha_\rho = 0.53$, $\alpha_{R2} = 2\alpha_\rho - 1 = 0.06$, and the β parameters depend on the exact process as given in Table I.3. Total and elastic cross sections for $\pi\pi$ and Kπ interactions are shown in Figure I.3.

Table I.3: Parameter values for the $\pi\pi$ and Kπ cross sections, as used in eq. (I.17). In the case of Kπ, I refers to the sum of the third isospin components for the incoming particles. The two $I = 1/2$ cases are equivalent, except for Clebsch-Gordan coefficients.

Case	β_P	β_ρ	β_2
$\pi^\pm\pi^\mp$	0.83	1.01	0.013
$\pi^\pm\pi^0$	0.83	0.267	-0.0267
$\pi^0\pi^0$	0.83	0.267	0.053
$\pi^\pm\pi^\pm$	0.83	-0.473	0.013
K$\pi^\pm, I = 1/2$	6.9032	8.2126	0.0
K$\pi^0, I = 1/2$	3.4516	4.1063	0.0
K$\pi, I = 3/2$	10.3548	-5.76786	0.0

For some of the remaining meson-hadron interactions, explicit resonances are implemented. In these cases, at low energies (below ~ 2 GeV, depending on the specific interaction), the total cross section is given by the elastic cross section plus the sum of resonance cross sections,

$$\sigma_{\mathrm{tot}} = \sigma_{\mathrm{el}} + \sum_{\mathrm{resonances}} \sigma_{\mathrm{res}} \ , \tag{I.18}$$

where σ_{el} and σ_{res} will be described in the following sections. There is an option in Pythia to also calculate the $\pi\pi$ and Kπ cross sections this way instead of using the default methods of Ref. [48, 49], but there are two drawbacks of using this approach. In terms of physics, it is less accurate because it does not take into account interference effects between resonances. And in terms of computational efficiency it is slower, which can have a significant impact on performance that is exacerbated by how common these interactions are.

One important case with a lot of data is $\mathrm{p/n} + \mathrm{K}^-/\bar{\mathrm{K}}^0$. Summing resonances does not accurately match data at low energies, so an additional contribution has been added, based on formulae from UrQMD. Furthermore we add an explicit elastic contribution not present in UrQMD in order to get an even better fit. Above 2.16 GeV, we use the *HPR*$_1$*R*$_2$ parameterization. The case $\mathrm{p/n} + \mathrm{K}^+/\mathrm{K}^0$ is also important and much data exists, but in this case resonances cannot form since there are no common quark–antiquark pairs to annihilate. We use an ad hoc parameterization to fit these cross sections to data at low energies. Specifically, the total cross section is given by 12.5 mb below 1.65 GeV and 17.5 mb above 1.9 GeV, with a linear transition in the intermediate range. The total and elastic cross sections for both these NK cases are shown in Figure I.2.

The last special case is $N\pi$ which uses the *HPR*$_1$*R*$_2$ parameterization above the resonance region. All other cases use the AQM parameterization above the resonance region. For those processes where resonances are not available, AQM is instead used at all energies.

3.2 Elastic scattering

In this section we discuss the directly elastic processes $AB \to AB$, leaving aside scattering through a resonance, $AB \to R \to AB$. A summary of σ_{el} descriptions is provided in Table I.4.

Table I.4: Summary of elastic cross section descriptions. Here, N is used to denote a nucleon, B a baryon and M a meson. For Kπ below 1.8 GeV, I refers to the sum of the third isospin component of the incoming particles.

Case	Method
pp/nn/pn, < 5 GeV	Fit to data
pp/nn/pn, > 5 GeV	CERN/HERA parameterization
Other BB	AQM parameterization
$\mathrm{p\bar{p}}$	UrQMD parameterization
Other $\mathrm{B\bar{B}}$	Rescaling $\mathrm{p\bar{p}}$
$\pi\pi$, < 1.42 GeV	Parameterization by Peláez et al. [48]
$\pi\pi$, > 1.42 GeV	Constant 4 mb
Kπ, $I = 1/2$, < 1.8 GeV	No scattering except through resonances
Kπ, $I = 3/2$, < 1.8 GeV	Parameterization by Peláez et al. [49]
Kπ, > 1.8 GeV	Constant 1.5 mb
Nπ, < 4 GeV	Fit to data
Nπ, > 4 GeV	CERN/HERA parameterization
NK	Ad hoc parameterization
Other MB/MM	AQM parameterization

For pp, nn, and pn, the elastic cross section is fitted to PDG data below 5 GeV [46], which is assumed to be the same as the total cross section up to 2.1 GeV. Above 5 GeV, σ_{el} is parameterized as a function of laboratory momentum p_{lab}, according to the CERN/HERA parameterization [54] with the general form

$$\sigma_{\text{HERA}}(p) = a + b\,p^n + c\log^2 p + d\log p\,, \tag{I.19}$$

with parameters given in Table I.5. For all other BB cases, the elastic cross section is given by an elastic AQM-style parameterization [28],

$$\sigma_{\text{AQM,el}} = 0.039\,\sigma_{\text{AQM,tot}}^{3/2}\,. \tag{I.20}$$

The CERN/HERA parameterization is also used for $\mathrm{p\bar{p}}$ for $p_{\text{lab}} > 5$ GeV, albeit with different parameters. Below this lab momentum, we use another ad hoc parameterization from UrQMD [28],

$$\sigma_{\text{el}}(\mathrm{p\bar{p}}) = \begin{cases} 78.6\,, & p < 0.3\,, \\ 31.6 + 18.3\,p^{-1} - 1.1\,p^{-2} - 3.8\,p\,, & 0.3 < p < 5\,. \end{cases} \tag{I.21}$$

For all other baryon-antibaryon cases, the elastic cross section is found by rescaling the $\mathrm{p\bar{p}}$ cross section, using an AQM factor in the same way as for total cross sections.

Table I.5: CERN/HERA parameters

Case	a	b	n	c	d
NN	11.9	26.9	-1.21	0.169	-1.85
$p\bar{p}$	10.2	52.7	-1.16	0.125	-1.28
$N\pi$	0	11.4	-0.4	0.079	0

For elastic cross sections involving mesons, there are several special cases. For $\pi\pi$, we separate our calculation into two regions, below and above 1.42 GeV, as for the total cross section. Below, the purely elastic cross section is found by parameterizing the d-wave contribution from Peláez et al. [47, 48]. This parameterization can be seen in Figure I.3, where it is equal to the total $\pi^+\pi^+$ cross section since no resonances can be formed in that case. The other $\pi\pi$ cases get the same contribution, except with a scale factor that depends on the exact case. Above 1.42 GeV, a constant elastic cross section of 4 mb is consistent with the parameterization of Ref. [48] when the contribution from resonances is taken into account. For $K\pi$, we divide the region into below and above 1.8 GeV. Below this threshold, for total isospin $I = 1/2$, the whole elastic cross section is well described by scattering through a resonance. For total isospin $I = 3/2$, resonances cannot form, and we instead use a parameterization by Ref. [49]. Above 1.8 GeV, we use a constant 1.5 mb for all cases.

In $N\pi$ interactions, the non-resonant elastic cross section vanishes below around 1.8 GeV. Between this energy and up to 4 GeV, we add a non-resonant contribution by interpolating data. Above 4 GeV, we use the CERN/HERA parameterization.

The last special case is NK^+/NK^0. This uses a simple fit to data, using 12.5 mb below 1.7 GeV and 4.0 mb above 2.5 GeV, with a linear transition in between. In all remaining cases, the AQM parameterization given in eq. (I.20) is used.

The angular distribution for non-resonant $AB \to AB$ is specified by the selection of the t value according to an exponential $\exp(B_{\text{el}}t)$, where the slope is given by

$$B_{\text{el}} = 2b_A + 2b_B + 2\alpha' \ln\left(\frac{s}{s_0}\right) . \tag{I.22}$$

Here $b_{A,B}$ is 2.3 GeV^{-2} for unflavoured baryons and 1.4 GeV^{-2} for mesons, $\alpha' = 0.25\ \text{GeV}^{-2}$ is the slope of the pomeron trajectory, and $s_0 = 1/\alpha' = 4\ \text{GeV}^2$ [55, 56]. The $b_{A,B}$ values are rescaled by AQM factors for strange or heavier hadrons, while α' is assumed universal.

Note that, strictly speaking, the σ_{tot}, σ_{el}, B_{el} and ρ (the ratio of the real to imaginary parts of the forward scattering amplitude) should be connected by the optical theorem. Here we make no attempt to model ρ or to exactly fulfil the optical theorem, which would have been quite messy in the low-energy resonance region. Note that an $L = 0$ resonance would decay isotropically, meaning a more complicated overall angular distribution when interference between elastic and resonance contributions is considered. We have checked, however, that

the optical theorem is approximately obeyed above the resonance region, assuming that ρ is not giving large effects.

3.3 Resonance formation

Explicit resonance formation has been implemented for $\pi\pi$, Kπ, Nπ, Nη, Nω, $\Sigma\pi$, ΣK, $\Lambda\pi$, ΛK, and $\Xi\pi$. This includes all isospin configurations of these particles where resonances exist (e.g. $\Sigma^+\pi^-$, but not $\Sigma^-\pi^-$). For the formation of a particular resonance R the cross section is given by a nonrelativistic Breit–Wigner [46]

$$\sigma_{AB\to R} = \frac{\pi}{p_{\mathrm{CM}}^2} \frac{(2S_R+1)}{(2S_A+1)(2S_B+1)} \frac{\Gamma_{R\to AB}\Gamma_R}{(m_R-\sqrt{s})^2 + \frac{1}{4}\Gamma_R^2} \,, \tag{I.23}$$

where S is the spin of each particle, p_{CM} is the CM momentum of the incoming particles, $\Gamma_{R\to AB}$ is the mass-dependent partial width, and Γ_R is the total mass-dependent width of R, found by summing the partial widths. The partial widths of a particle at mass m are given by UrQMD as

$$\Gamma_{R\to AB}(m) = \Gamma_{R\to AB}(m_0) \frac{m_0}{m} \frac{\left\langle p^{2l+1}(m)\right\rangle}{\left\langle p^{2l+1}(m_0)\right\rangle} \frac{1.2}{1.0 + 0.2\frac{\left\langle p^{2l}(m)\right\rangle}{\left\langle p^{2l}(m_0)\right\rangle}} \,, \tag{I.24}$$

where m_0 is the nominal mass of the particle and $\Gamma_{R\to AB}(m_0)$ is the nominal width, both known from experiment, and l is the angular momentum of the outgoing two-body system. The final factor ensures that widths do not blow up at large masses. The phase space factors are given by

$$\left\langle p^{2l+1}(m)\right\rangle = \iint p_{\mathrm{CM}}^{2l+1}(m, m_A, m_B)\, A(m_A)\, A(m_B)\, \mathrm{d}m_A\, \mathrm{d}m_B \,, \tag{I.25}$$

where

$$p_{\mathrm{CM}}(m, m_A, m_B) = \frac{\sqrt{(m^2-(m_A+m_B)^2)(m^2-(m_A-m_B)^2)}}{2m} \tag{I.26}$$

and $A(m)$ are the mass distribution functions, given by a Breit–Wigner,

$$A(m) = \frac{1}{2\pi} \frac{\Gamma(m)}{(m^2-m_0^2)^2 + \frac{1}{4}\Gamma^2(m)}, \tag{I.27}$$

which reduces to $A(m) = \delta(m-m_0)$ for particles with zero width. Note that although the mass distribution depends on mass-dependent widths, which again depend on the mass distribution of other particles, there is no circular dependency since particle widths can only depend on the widths of lighter particles.

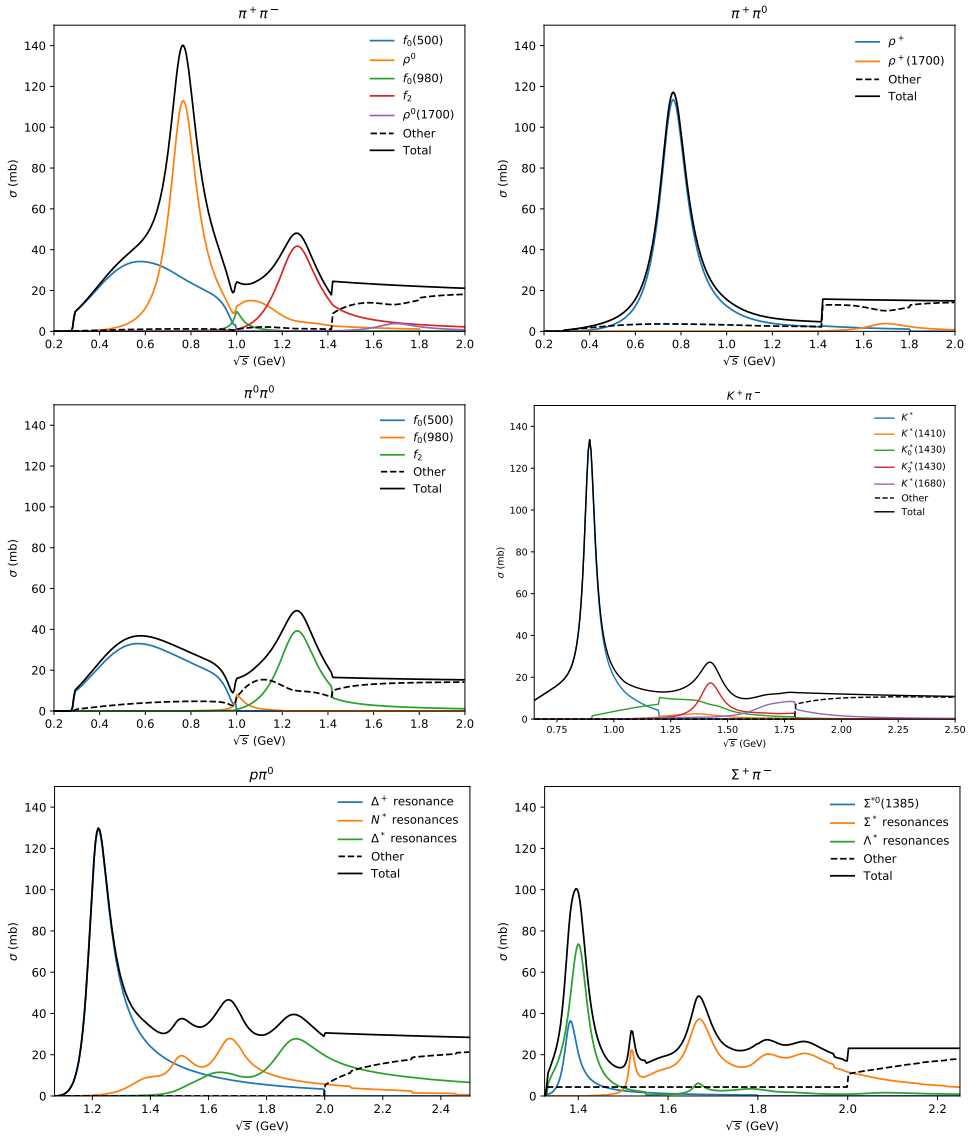


Figure I.4: Resonant cross sections for some important cases, with partial cross sections for each resonance. For $p\pi^0$ and $\Sigma^+\pi^-$ there are many resonances, and we have divided them into groups for readability. The "other" cross sections include elastic, diffractive and non-diffractive.

Figure I.4 shows the resonant cross sections for some important cases. For the $\pi\pi$ cases there is a small elastic cross section below 1.42 GeV, corresponding to a d-wave contribution. For $K^+\pi^-$ there is no direct elastic cross section at low energies, but a significant fraction of

the resonances formed will decay back to the initial state particles, cf. Figure I.3. We also observe a discontinuous behaviour at some points. One reason for this is that resonance particles are assigned a restricted mass range outside which they cannot be formed, which is particularly noticeable for example for $\mathrm{p}\pi^0 \to \Delta^+$ at 2.0 GeV. Another reason for a non-smooth behaviour is the fact that the total cross section is parameterized using the more sophisticated machinery of [47–49] and the resonance cross sections are scaled to sum to this value. This is especially noticeable for $\pi^+\pi^- \to \rho^0$, where the total cross section is significantly larger than the sum of resonance cross sections in the range around 1.0-1.2 GeV, and is why the cross section for $\pi^+\pi^- \to \rho^0$ has a second peak in that region instead of looking like a regular Breit-Wigner. Both these kinds of discontinuities are visible in the $\mathrm{K}^+\pi^-$ cross sections, at the K^* cutoff at 1.2 GeV.

One exceptional case is the formation of $f_0(500)$ resonances in $\pi^+\pi^-$ or $\pi^0\pi^0$ interactions. The nature of the $f_0(500)$ meson is not fully understood and it has certain exotic properties, notably its width is about the same as its mass. For this reason, eq. (I.23) does not describe its formation well. We find the relevant cross sections by interpolating values calculated based on the work by Peláez et al. [47, 48]. After the $f_0(500)$ has been produced, it is treated as any other meson, including in its decay.

The formula for mass-dependent partial widths works only for two-body decays. These are the dominant ones for most resonances we consider, but some hadrons have three- or four-body decays, for instance $\rho^0 \to \pi^+\pi^-\pi^+\pi^-$. For such particles, we calculate the mass-dependent partial widths for the two-body channels according to eq. (I.24), but assume that the multibody channels have a constant width for the purposes of calculating the total width needed in eq. (I.23).

In the space–time description, the resonance is created at the average location of the two incoming hadrons at the interaction time in the collision CM frame. The resonance is then treated as any unstable particle with a mean lifetime that is assumed to be $\tau = 1/\Gamma(m_0)$, even if the resonance is off-shell. If all decay channels of the resonance are two-body decays, then eq. (I.24) is used to calculate the branching ratios. In this case, the masses of the outgoing particles are picked according to

$$\mathrm{d}\Gamma_{R\to AB} \sim p_{\mathrm{CM}}^{2l+1}(m, m_A, m_B)\, A(m_A)\, A(m_B)\, \mathrm{d}m_A\, \mathrm{d}m_B\,. \tag{I.28}$$

If there is one or more multibody decay channels, the particle is instead decayed using the existing Pythia machinery.

3.4 Annihilation

In $\mathrm{B\bar{B}}$ collisions the baryon number can be annihilated, so that only mesons remain in the final state. For $\mathrm{p\bar{p}}$, below 2.1 GeV, annihilation counts for all inelastic processes, so below

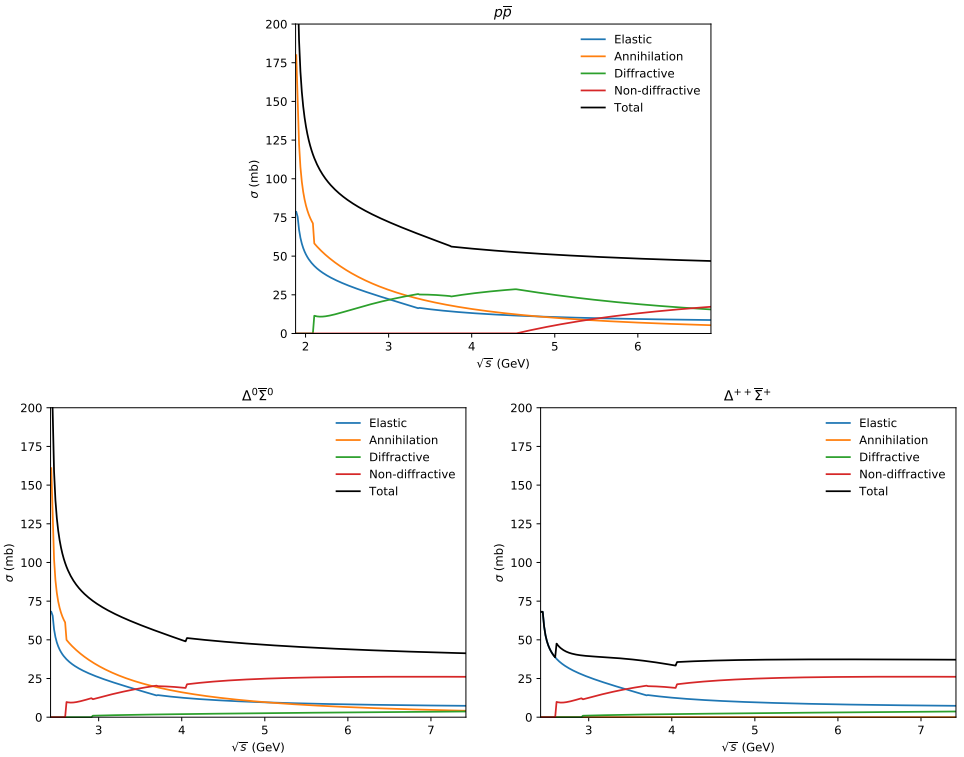


Figure I.5: Partial cross sections for $p\bar{p}$, $\Delta^0\overline{\Sigma}^0$ and $\Delta^{++}\overline{\Sigma}^+$. We see that $\Delta^0\overline{\Sigma}^0$ is simply a rescaling of the $p\bar{p}$ case, except it gets different diffractive and non-diffractive contributions because $p\bar{p}$ implements explicit resonances. For $\Delta^{++}\overline{\Sigma}^+$ annihilation is not possible, so the annihilation cross section is subtracted from the total, significantly changing its shape.

this threshold,

$$\sigma_{\text{ann}} = \sigma_{\text{tot}} - \sigma_{\text{el}}. \tag{I.29}$$

Above the threshold, it is given by a parameterization by Koch and Dover [57],

$$\sigma_{\text{ann}} = 120\frac{s_0}{s}\left(\frac{A^2 s_0}{(s-s_0)^2 + A^2 s_0} + 0.6\right), \tag{I.30}$$

where $s_0 = 4m_p^2$ and $A = 0.05$ GeV. For other $B\bar{B}$, this is rescaled in the same way as for the total cross section. Note that the cross section is taken to be the same regardless of whether the baryons have one, two or three quarks in common, but if there are none then currently no annihilation is assumed, even though in principle it would be possible to decompose a $B\bar{B}$ system with no $q\bar{q}$ pairs in common into three separate $q\bar{q}$ strings. Figure I.5 shows the cross sections for $p\bar{p}$, $\Delta^0\overline{\Sigma}^0$ and $\Delta^{++}\overline{\Sigma}^+$.

When an annihilation process occurs, one or two quark-antiquark pairs are annihilated. If two or more pairs are available, the probability for a second annihilation is given by a free parameter, by default 0.2, to represent a small but existing rate. No complete annihilation of all three pairs is performed, since the rate presumably is small and since it then would be necessary to recreate a new pair, making little net difference. The pair(s) to be annihilated is (are) chosen uniformly among all possible combinations. If only one quark pair remains, a single string is stretched between the q and $\bar{\mathrm{q}}$, along the original collision axis. If two pairs remain, a random pairing is done to form two separate strings. The procedure for sharing momentum is similar to the one described below in Section 3.6. The possibility of having a single string stretched between a diquark–antidiquark pair is omitted, since then a new baryon–antibaryon pair would be produced.

3.5 Diffractive processes

Diffractive cross sections in the continuous regime are calculated using SaS ansatz [56, 58]. The basic version of SaS is designed to deal only with processes involving p, $\bar{\mathrm{p}}$, π, ρ, ω and ϕ (as needed for $\mathrm{pp}/\gamma\mathrm{p}/\gamma\gamma$ collisions), and only for collision energies above 10 GeV. It is here extended to all baryons by applying an AQM rescaling factor to the corresponding p cross sections. For mesons a similar rescaling to π (= ρ) cross sections is performed, except that here ϕ is retained as the template for $\mathrm{s\bar{s}}$ interactions. The η and η' cross sections thus are the appropriate mixes of π and ϕ ones.

The differential cross section for single diffraction $AB \to XB$ is taken to be of the form

$$\mathrm{d}\sigma_{XB} \propto \frac{\mathrm{d}M_X^2}{M_X^2} \left(1 - \frac{M_X^2}{s}\right) \exp(B_{XB}\, t)\, \mathrm{d}t \,, \tag{I.31}$$

where

$$B_{XB}(s) = 2b_B + 2\alpha' \ln\left(\frac{s}{M_X^2}\right) \,, \tag{I.32}$$

with b_B and α' as for elastic scattering. The constant of proportionality involves hadron–pomeron and triple-pomeron couplings, specified for the few template processes and then multiplied by AQM factors. The diffractive mass spectrum is taken to begin at $M_{X,\mathrm{min}} = m_A + 2m_\pi = m_A + 0.28$ GeV and extend to the kinematical limit $M_{X,\mathrm{max}} = E_{\mathrm{CM}} - m_B$, while t can take values within the full allowed range [23]. Above $E_{\mathrm{CM,min}} = 10$ GeV the integrated cross section has been parameterized. Below this scale, our studies show that a shape like

$$\sigma_{XB}(E_{\mathrm{CM}}) = \sigma_{XB}(E_{\mathrm{CM,min}}) \left(\frac{E_{\mathrm{CM}} - M_{X,\mathrm{min}} - m_B}{E_{\mathrm{CM,min}} - M_{X,\mathrm{min}} - m_B}\right)^{0.6} \tag{I.33}$$

provides a good representation of the behaviour down to the kinematic threshold. Note that m_A and m_B are the actual masses of the colliding hadrons, not those of the corresponding template process.

Single diffraction $AB \to AX$ is obtained by trivial analogy with $AB \to XB$. For double diffraction $AB \to X_1X_2$ the cross section reads

$$\mathrm{d}\sigma_{XX} \propto \frac{\mathrm{d}M_1^2}{M_1^2}\,\frac{\mathrm{d}M_2^2}{M_2^2}\left(1-\frac{(M_1+M_2)^2}{s}\right)\left(\frac{s\,m_\mathrm{p}^2}{s\,m_\mathrm{p}^2+M_1^2\,M_2^2}\right)\exp(B_{XX}\,t)\,\mathrm{d}t\,, \quad \text{(I.34)}$$

where

$$B_{XX}(s) = 2\alpha' \ln\left(e^4 + \frac{s\,s_0}{M_1^2\,M_2^2}\right)\,, \quad \text{(I.35)}$$

again with $s_0 = 1/\alpha'$. For the behaviour below 10 GeV, our studies suggest that

$$\sigma_{XX}(E_\mathrm{CM}) = \sigma_{XX}(E_\mathrm{CM,min})\left(\frac{E_\mathrm{CM}-M_{X_1,\mathrm{min}}-M_{X_2,\mathrm{min}}}{E_\mathrm{CM,min}-M_{X_1,\mathrm{min}}-M_{X_2,\mathrm{min}}}\right)^{1.5} \quad \text{(I.36)}$$

is a suitable form.

So far we only considered the continuum production, which dominates for large diffractive masses. For small masses, diffractive cross sections can also include the formation of explicit resonances, and the contribution from these should be added to the continuum contribution. In our framework, this can occur as $\mathrm{NN} \to \mathrm{NN}^*$ or $\mathrm{NN} \to \mathrm{N}\Delta^*$ (single diffractive), or $\mathrm{NN} \to \Delta\mathrm{N}^*$ or $\mathrm{NN} \to \Delta\Delta^*$ (double diffractive), and similarly when one baryon is replaced by its antibaryon. Higher excitations are implicitly part of the continuum diffractive treatment and not considered here. The cross section for $AB \to CD$ is given by Ref. [28]

$$\sigma_{AB\to CD} = (2S_C+1)\,(2S_D+1)\,\frac{1}{s}\,\frac{\langle p_{CD}\rangle}{\langle p_{AB}\rangle}\,|\mathcal{M}|^2\,, \quad \text{(I.37)}$$

where S is the spin of each particle, $\mathcal{M}$ is the matrix element, and $\langle p_{ij}\rangle$ are phase space factors given by eq. (I.25) (assuming $l = 0$). In practice, this expression will sometimes lead to the sum of partial cross sections being larger than the total one. In those situations, we rescale the excitation cross sections (leaving other partial cross sections unchanged) so that the sum of partial cross sections is equal to the total.

For the matrix elements, we use the same as UrQMD [28]. For $\mathrm{NN} \to N\Delta$ it is given by

$$|\mathcal{M}|^2 = A\,\frac{m_\Delta^2\Gamma_\Delta^2}{(s-m_\Delta^2)^2+m_\Delta^2\Gamma_\Delta^2}\,, \quad \text{(I.38)}$$

where $m_\Delta = 1.232$ GeV and $\Gamma_\Delta = 0.115$ GeV are the nominal mass and width of Δ, and the coefficient is $A = 40000$. For $\mathrm{NN} \to \Delta\Delta$, the matrix element is a constant

$|\mathcal{M}|^2 = 2.8$. Finally, for the remaining classes, the matrix element takes the form

$$|\mathcal{M}|^2 = \frac{A}{(m_C - m_D)^2 (m_C + m_D)^2} \,, \tag{I.39}$$

where m_C and m_D are the nominal masses for the outgoing particles (which will never be the same for these classes, so the matrix element cannot diverge), and the coefficient A is $A = 6.3$ for NN $\to$ NN*, $A = 12$ for NN $\to N\Delta^*$, and $A = 3.5$ for NN $\to \Delta N^*$ and NN $\to \Delta\Delta^*$.

In eq. (I.37), the only dependence on outgoing masses comes from the phase space term. Thus, the masses of the outgoing particles are distributed according to

$$\frac{\mathrm{d}\sigma_{AB\to CD}}{\mathrm{d}m_C \mathrm{d}m_D} \sim p_{\mathrm{CM}}(E_{\mathrm{CM}}, m_C, m_D)\, A(m_C)\, A(m_D) \,, \tag{I.40}$$

from eq. (I.25). The t behaviour is assumed to be given by an exponential slope with the same B_{XB}/B_{XX} as in the continuum single/double diffraction for the given diffractive masses.

Calculating the integrals in eq. (I.25) during event generation would be debilitatingly slow. Therefore, we tabulate the cross sections for each process up to 8 GeV and use interpolation to get the total and partial excitation cross sections. For energies above this threshold, the expansion

$$p_{\mathrm{CM}}(E_{\mathrm{CM}}, m_C, m_D) = \frac{1}{2} E_{\mathrm{CM}} \left(1 - \frac{m_C^2 + m_D^2}{E_{\mathrm{CM}}^2} + \mathcal{O}(E_{\mathrm{CM}}^{-3}) \right) \tag{I.41}$$

shows that p_{CM} is approximately constant with respect to m_C and m_D when $E_{\mathrm{CM}} \gg m$. At the same time, the mass distributions $A(m)$ vanish at large m. Thus, in this limit, the phase factor can be approximated as

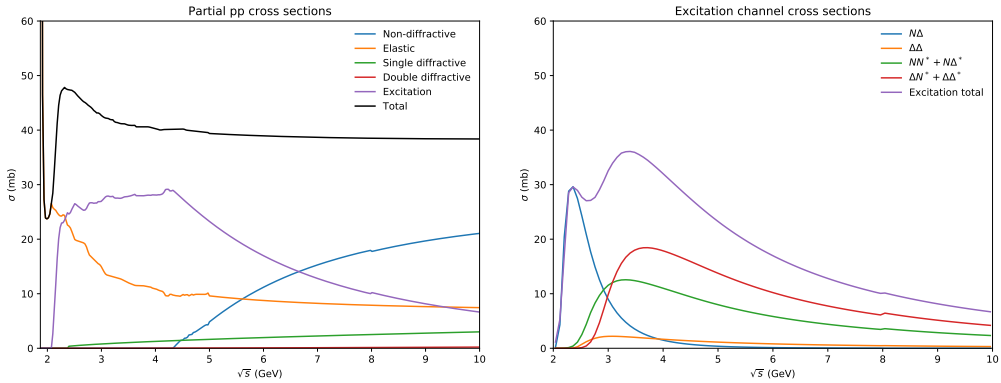
$$\langle p_{CD} \rangle \approx p_{\mathrm{CM}}(E_{\mathrm{CM}}, m_{C,0}, m_{D,0}) \int \mathrm{d}m_C\, A(m_C) \int \mathrm{d}m_D\, A(m_D), \tag{I.42}$$

By integrating A ahead of time, the cross sections can be calculated efficiently during runtime also above the tabulated region.

For other incoming hadron combinations, we fall back on the simpler smooth low-mass enhancement implemented in SaS to compensate for the lack of explicit resonances. For $AB \to XB$ the differential cross section in eq. (I.31) is multiplied by a factor

$$c_{\mathrm{res}} \frac{(m_A + M_{\mathrm{res},0})^2}{(m_A + M_{\mathrm{res},0})^2 + M_X^2} \,. \tag{I.43}$$

Here $c_{\mathrm{res}} = 2$ and $M_{\mathrm{res},0} = 2\ \mathrm{GeV} - m_{\mathrm{p}}$ have been chosen to provide a decent description of the low-mass enhancement in pp collisions at medium-high energies. For energies below



(a) (b)

Figure I.6: (a) All partial cross sections for pp interactions. (b) Excitation cross sections, according to eq. (I.37). Note the small jump at 8 GeV, at the boundary between the tabulated and parameterized regions.

10 GeV this part of the cross section can be described in the same spirit as the continuum part in eq. (I.33), but the power is changed from 0.6 to 0.3. Double diffraction can be handled in the same spirit. Three terms contribute, where either side A, side B or both are enhanced by a factor like eq. (I.43). In eq. (I.36) the power is changed from 1.5 to 1.25 for the first two and to 1.0 for the last one.

The kinematics of events is provided by the mass and t selections outlined above. The decays of the explicit low-mass resonances are assumed to be isotropic. In the other cases a diffractive system is handled as a string stretched between two parts of the incoming hadron. A baryon is split into a diquark plus a quark at random, where the former/latter is moving in the forwards/backwards direction in the rest frame of the hadron. Here forwards is the direction the hadron will be moving out along, once boosted to the collision CM frame. A meson is correspondingly split into a quark plus an antiquark, but here the choice of which is moving forwards is taken to be random. The two string ends are given relative $p_\perp$ kicks of nonperturbative size, however, such that the string alignment along the collision axis is smeared.

Figure I.6a shows all partial cross sections for pp collisions. We see that the single diffractive cross section is very small compared to other cross sections, and the double diffractive one almost vanishes. The excitation cross section is here shown separately from the cross sections describing diffraction in the continuous region. Note that below around 4.5 GeV, the excitation cross section is set equal to the difference $\sigma_{\text{tot}} - \sigma_{\text{el}}$ instead of following the form given by eq. (I.37). The full shape of the excitation cross sections are shown in Figure I.6b.

3.6 Nondiffractive processes

Nondiffractive cross sections are found by subtracting all other partial cross sections from the total cross section,

$$\sigma_{\mathrm{nondiff}} = \sigma_{\mathrm{tot}} - \sigma_{\mathrm{el}} - \sigma_{\mathrm{diff}} - \sigma_{\mathrm{res}} - \sigma_{\mathrm{ann}}. \tag{I.44}$$

At large energies the nondiffractive processes dominate the total cross section, but at low energies they can have a small or even vanishing cross section. Since it is defined as the difference between the total and the other partial cross sections, it can sometimes have a fluctuating energy dependence with no clear physics explanation.

A nondiffractive event is associated with the exchange of a gluon between the two incoming hadrons, where the gluon carries negligible momentum but leads to a rearranged colour topology. To this end, each initial hadron is separated into a colour (a quark or an antidiquark) part and an anticolour (an antiquark or a diquark) part. For a baryon the selection of the diquark part is done according to the $SU(6)$ decomposition (in three flavours times two spins), while the meson subdivision is trivial. After the colour-octet gluon exchange, the colour end of one hadron forms a colour singlet with the anticolour end of the other hadron, and vice versa. (Cases with more complicated colour-charge topologies are suppressed and are neglected here.) This leads to two strings being stretched out between the two octet-state "hadrons".

Consider the collision in its rest frame, with hadron A (B) moving in the $+z$ ($-z$) direction. In that frame, the colour and anticolour objects of each hadron are assumed to have an opposite and compensating $p_\perp$. This is chosen according to a Gaussian with the same width as used to describe the $p_\perp$ smearing in string breakup vertices. In the breakup context a width of $\langle p_\perp^2 \rangle \approx (0.35\ \mathrm{GeV})^2$ is motivated by a tunnelling mechanism, but a number of that magnitude for the parton motion inside a hadron could equally well be viewed as a consequence of confinement in the transverse directions by way of the Heisenberg uncertainty relations.

Including (di)quark masses, the transverse masses $m_{\perp A1}$ and $m_{\perp A2}$ of the two A hadron constituents are defined. Next a z_A value is picked that splits the A lightcone momentum $p_A^+ = E_A + p_{zA}$ between the two, $p_{A1}^+ = z_A p_A^+$ and $p_{A2}^+ = (1 - z_A) p_A^+$ [39]. For a meson $z = x_1/(x_1 + x_2)$, where the x_i are picked at random according to $(1 - x_i)^{0.8}/\sqrt{x_i}$. For a baryon first each of the three quarks are assigned an x_i according to $(1 - x_i)^{2.75}/\sqrt{x_i}$. If z_A is associated with the diquark, made out of the first two quarks, then $z_A = 2(x_1 + x_2)/(2(x_1 + x_2) + x_3)$. Note that here the diquark tend to take most of the momentum, not only because it consists of two quarks, but also by an empirical enhancement factor of 2. The p_{Ai}^- can now be obtained from $p^+ p^- = m_\perp^2$, and combined to give an effective mass m_A^* that the A beam remnant is associated with: $m_A^{*2} = m_{\perp A1}^2/z + m_{\perp A2}^2/(1 - z)$. The same procedure can be repeated for the B hadron, but with $p^+ \leftrightarrow p^-$. Together, the

criterion $m_A^* + m_B^* < E_{\mathrm{CM}}$ must be fulfilled, or the whole selection procedure has to be restarted. (Technically, some impossible values can be rejected already at earlier stages.) 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 $\mathrm{u\bar{u}}$ or $\mathrm{d\bar{d}}$ 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.

One reason why $m_A^* + m_B^* < E_{\mathrm{CM}}$ might fail is if the constituent transverse masses are too big. Thus, after a number of failed attempts, their values are gradually scaled down to increase the likelihood of success. This, on the other hand, increases the risk of obtaining two strings with low invariant masses. A further check is therefore made that each string has a mass above that of the lightest hadron with the given flavour content, and additionally that the mass excess is at least a pion mass for one of the two strings.

The two strings can now be hadronized, but often one or both have small masses. To this end the ministring framework, used when at most two hadrons can be formed from a string, has been extended to try harder. Several different approaches are used in succession, until one of them works. The order is as follows.
(1) Several attempts are made to produce two hadrons from the string by a traditional string break in the middle.
(2) If not, a hadron is formed consistent with the endpoint flavour content. Four-momentum is shuffled between it and one of the partons of the other string, so as to put the hadron on mass shell while conserving the overall four-momentum. Since the string with lowest mass excess is considered first, the two partons of the other string should normally be available.
(3) If no allowed shuffling is found, then a renewed attempt is made to produce two hadrons by a string break, but this time the two lightest hadrons of the given flavour content are chosen.
(4) If that does not work, one lightest hadron is formed from the endpoint flavours and the other is set to be a π^0.
(5) It still no success, then go back to forming one hadron, but the lightest possible, and again shuffle momentum to a parton.
(6) Finally, the problem may occur also for the string with higher mass excess, i.e. after the

first string was hadronized, and possibly took some four-momentum in the process. Then a collapse to one hadron (at random or eventually the lightest) with the recoil taken by another hadron is attempted.

3.7 The transition to high-energy processes

We have now described a framework for low energy hadron-hadron interactions. Our motivation for doing this has been to apply it to rescattering, but in principle, having this framework means that it is now possible to generate events in PYTHIA at these low energies. Despite all the technical details, the structure of the resulting events is quite simple. At most two objects (either hadrons or strings) are created in the first step of the process. The strings are stretched out almost perfectly along the collision axis and fragment into hadrons with only small nonperturbative $p_\perp$ kicks.

This is in contrast to the high-energy framework used to simulate the primary LHC pp collision, e.g. in inelastic nondiffractive processes. Here the multiparton interactions machinery very much is based on perturbation theory, where each interaction requires the use both of hard matrix elements and parton distribution functions (PDFs), giving scattered partons over a wide range of $p_\perp$ scales, even if the lower scales dominate. Many string pieces are stretched criss-cross in the event, and fragment into the high-multiplicity initial state that the rescattering framework will be applied to. If one uses this perturbative framework at lower and lower energies the average number of MPIs will decrease, as will their typical $p_\perp$ scale. Gradually the idea of applying a perturbative approach becomes less appealing. Technically the machinery can be applied down to 10 GeV CM energy, but is then highly questionable. Furthermore, many of the cross sections described here do not scale correctly at higher energies. For a high-energy $\mathrm{pp}/\mathrm{p\bar{p}}$ primary collision four different models are available [59]. Only one of them explicitly covers some more collision types, but extensions by AQM rescaling could be possible.

Therefore it is tempting to interpolate between the two descriptions. There is now such an option available. In it, the fraction of perturbatively handled events rises from the threshold energy $E_{\mathrm{thr}} = 10$ GeV as

$$P_{\mathrm{pert}} = 1 - \exp\left(-\frac{E_{\mathrm{CM}} - E_{\mathrm{thr}}}{E_{\mathrm{wid}}}\right) , \tag{I.45}$$

where $E_{\mathrm{wid}} = 10$ GeV is a measure of the size of the transition region. This is actually the same form as already used previously to transition between a nonperturbative and a perturbative description of diffraction, with the diffractive system mass replaced by E_{CM} [59, 60].

How this transition works in practice is illustrated in Figure I.7a, for the energy dependence

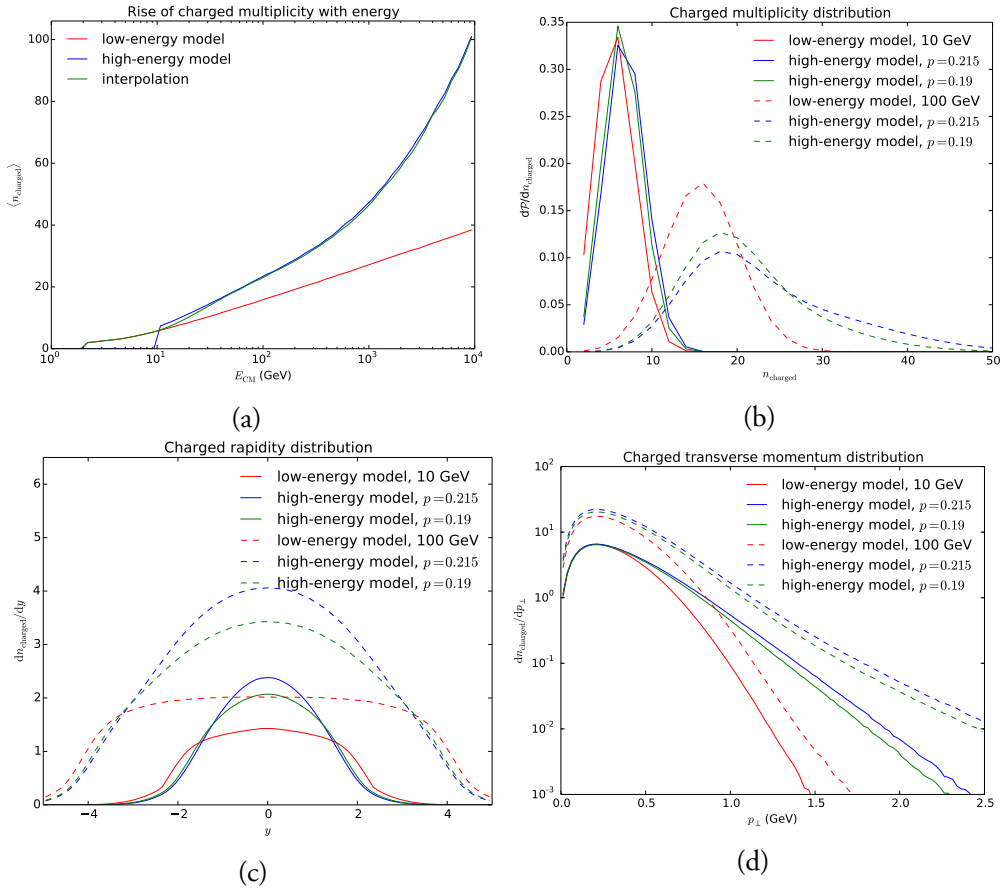


Figure I.7: (a) Energy dependence of the average charged multiplicity in nondiffractive pp collisions. (b,c,d) Comparison of charged multiplicity, rapidity and transverse momentum distributions for 10 and 100 GeV nondiffractive pp collisions.

of the charged multiplicity in nondiffractive events. In this figure the difference between the low-energy and high-energy model multiplicities is not so large in the transition range 10 – 30 GeV, but the importance of the perturbative components obviously increases with energy. Zooming in on the behaviour at the 10 GeV threshold and at an energy above it, at 100 GeV, Figure I.7b,c,d show some differential distributions. At 10 GeV the limited phase space does not allow for high multiplicities, while a longer perturbatively-induced tail is apparent at 100 GeV. Nevertheless, the MPI activity is reflected in a shift towards central rapidities and the presence of a high-$p_\perp$ tail already at 10 GeV.

The perturbative model results have been obtained with the default Monash tune [61], which mainly is based on comparisons with LEP, Tevatron and LHC data. One should therefore be aware that the extrapolation to lower energies is not without its problems. As

an example, the key parameter of the MPI framework is the $p_{\perp 0}$ one, that regularizes the divergence of the perturbative $2 \to 2$ cross sections in the limit $p_\perp \to 0$. It is assumed to have an energy dependence that scales like $p_{\perp 0} \propto E_{\mathrm{CM}}^{p}$ (but more complicated forms could be considered). The default values, with $p = 0.215$, gives $p_{\perp 0} = 0.56$ and 0.91 GeV, respectively, at 10 and 100 GeV. If p is changed to 0.19, then instead $p_{\perp 0} = 0.66$ and 1.02 GeV, respectively, at the low energies, assuming a fixed $p_{\perp 0}$ value at 7 TeV. The result of such a modest change is illustrated in Fig. I.7b,c,d. Qualitatively the difference to the low-energy model remains, but quantitatively it is visibly reduced.

One may also note that the string drawing can be quite different in the two cases. In the nonperturbative model the pp events always are represented by two strings, each stretched between a quark and a diquark. When MPIs are included, it becomes frequent that two quarks are kicked out of the same proton, more so at low energies where the high-x valence-quark part of PDFs is probed. This leads to so-called junction topologies, where the baryon number can wander more freely in the event [62]. Technically, this makes the hadronization of low-energy events more messy, and may require repeated attempts to succeed.

In diffraction, the excited masses M_X vary between events, also for a fixed CM energy. To handle perturbative activity inside the diffractive system then would seem to require a time-consuming re-initialization of the MPI framework for each new diffractive system. Instead, at the beginning of a run, an initialization is done for a set of logarithmically spaced diffractive masses, and numbers relevant for the future generation are saved in arrays. By interpolation, required numbers can then be found for any mass during the subsequent event generation. This approach has now been extended also to be available for nondiffractive processes, if so desired. This means that pp collisions can be simulated essentially from the threshold to LHC energies and beyond without any need to re-initialize. The prize to pay is a somewhat longer initialization step at the beginning of a run, but still in the range of seconds rather than minutes. One current limitation is that it is numbers for the MPI generation that are stored, so it is not now possible to pick a specific hard process for handling in the same way.

Another limitation is that the perturbative framework requires access to PDFs for the colliding hadrons, which restricts us to p, n and (with big uncertainties) π. Additionally PDFs are available for the photon and the pomeron, the latter used in diffraction, and in that sense they can be handled on equal footing with hadrons. A further restriction is that Pythia can only be set up for one combination of incoming beams at a time, so as to handle the perturbative processes. The simpler nonperturbative machinery used for rescatterings has no such restriction, of course.

4 Model tests

In this section, we will study the properties of the rescattering model. We start with studying how rescattering affects simple observables such as $p_\perp$ spectra, charged multiplicity, jet structure, and the potential for collective flow. We also look at how event properties change when rescattering is performed in a Lorentz boosted frame, in order to verify that the frame-dependence described in Section 2.3 does not significantly alter the final state.

Next, we look at the rates at which different particle types participate in rescattering and the rates at which the different types of processes occur. Finally, we consider the free parameters and model choices that have gone into the framework, and study the effect of changing those.

4.1 Basic effects of rescattering

As the most basic check, Figure I.8 shows how charged multiplicity, rapidity spectra, transverse momentum spectra, and invariant production times are affected by rescattering. We see that rescattering increases charged multiplicity, which is obviously expected when one considers the fact that we have implemented $2 \to n$, $n \geq 3$ interactions, but not interactions involving multiple incoming particles. The rescatter-affected hadrons have a broader multiplicity distribution than those not involved: events that start out with a low number of primary hadrons have a smaller rescattering probability than average, and vice versa.

In the same vein, the rescattered fraction is larger for central rapidities, where there are more hadrons to begin with, and this is also where inelastic rescatterings give a multiplicity increase. An interesting observation is that higher-$p_\perp$ hadrons seldom participate in rescattering, Figure I.8c. The natural explanation is that these hadrons typically are produced at larger transverse distances by (mini)jet fragmentation, where the particle density is reduced by having fewer overlapping MPI systems than at small $r_\perp$. Notable is also the slight net decrease at high $p_\perp$ by rescattering, (over)compensated by the increase at small $p_\perp$. Finally, and quite logically, rescattering kicks in with some delay in invariant time, since a sufficient amount of primary hadrons have to be produced first.

The point of introducing rescattering is to change some event properties, but not all changes are relevant rescattering signals, since some could easily be compensated by a retuning of many other parameters. In particular, the average (charged) event multiplicity is such a signal. Indeed, the fact that it is changed by rescattering means that a retune is necessary in order to restore it to the experimentally well-known value. The MPI framework, which is the main driving force in generating the multiplicity spectrum, is sufficiently uncertain to easily absorb the rescattering effects on the multiplicity. More specifically, when we study the effects of rescattering, the $p_{\perp 0}$ parameter of the MPI framework,

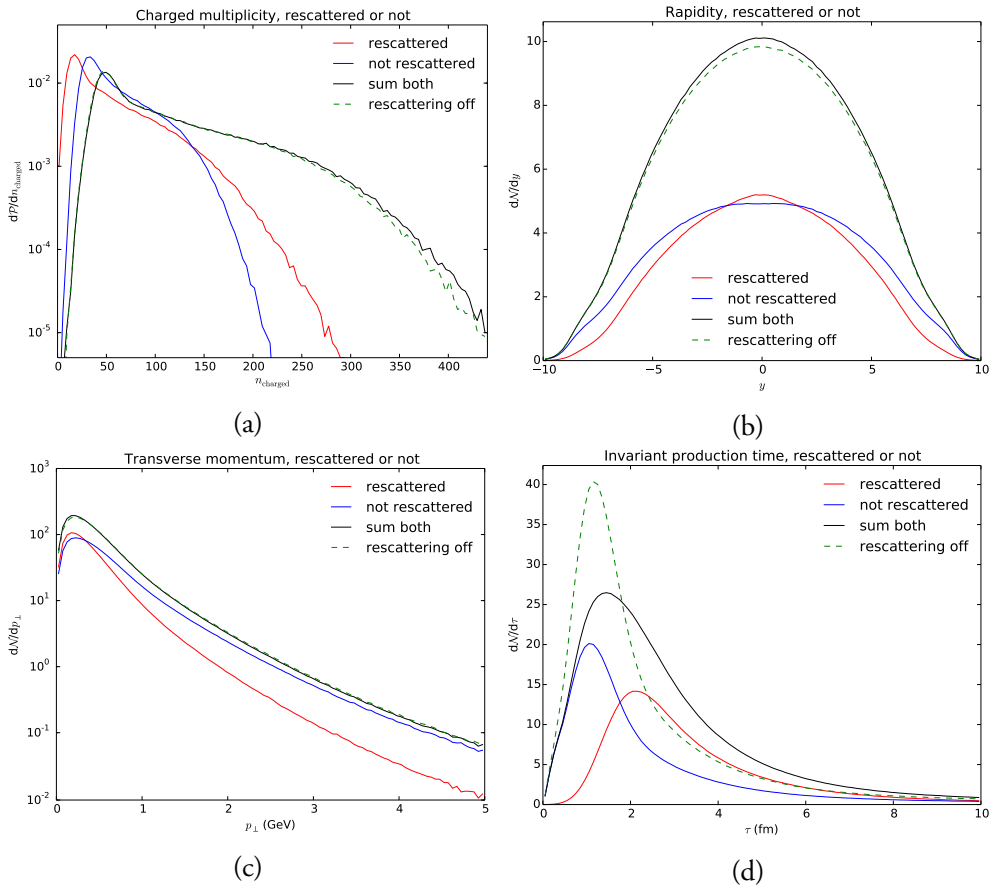
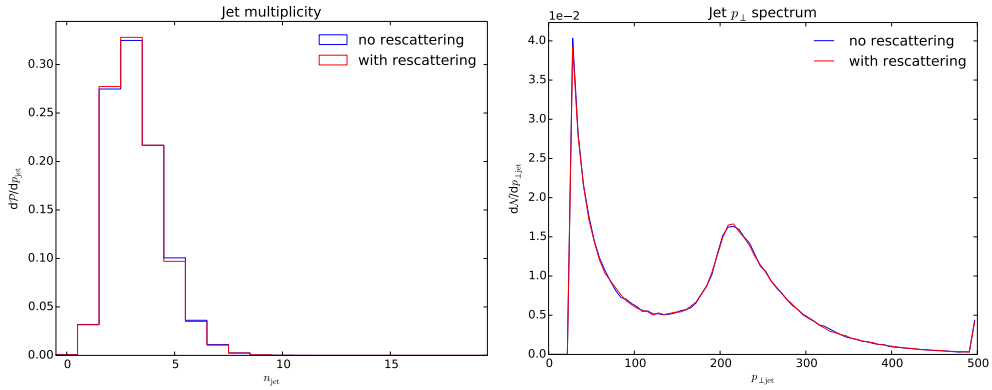


Figure I.8: (a) Multiplicity, (b) rapidity, (c) transverse momentum, and (d) invariant production time spectra of charged final-state hadrons, subdivided into those that have been involved in rescatterings and those that have not, in 13 TeV nondiffractive pp events. As reference a comparison is also made with events without rescatterings.

`MultipartonInteractions:pT0Ref`, is adjusted to restore the average charged multiplicity in the $\eta < 2.5$ range to the no-rescattering value. Its default value in Pythia is `pT0Ref = 2.28` (GeV), and we have found that setting it to `pT0Ref = 2.345` restores charged multiplicity to the correct value. We will use this value in all subsequent studies, unless otherwise noted. In the future, a more detailed retune would be desirable.

4.2 Jets

We have already argued that high-$p_\perp$ particles are less affected by rescattering than low-$p_\perp$ ones, and hence jets should remain essentially unchanged. This also turns out to be the case.



(a) (b)

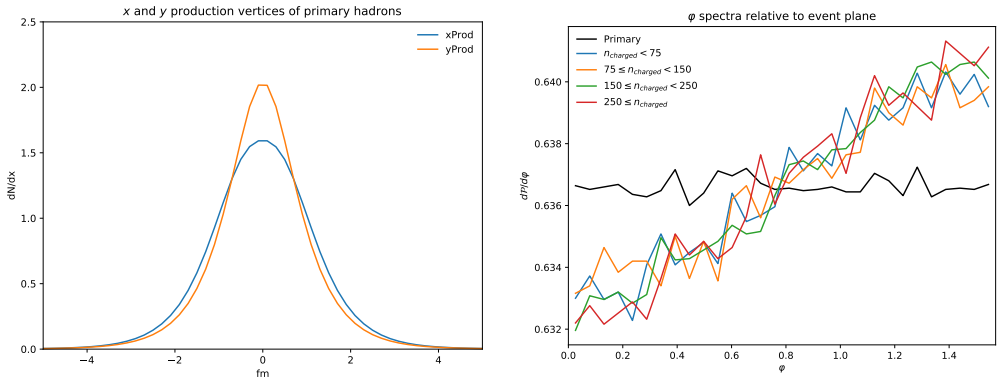
Figure I.9: Production rates (a) and inclusive $p_\perp$ spectra (b) of jets in 13 TeV pp collisions, as further described in the text. The uptick in the last bin of (b) is because all jets with $p_\perp > 500$ GeV have been put there.

As an example, QCD two-jet production with $p_\perp > 200$ GeV hard collisions at 13 TeV was studied, and anti-$k_\perp$ jets found for a 0.7 radius and a 25 GeV lower cut-off [63]. We then find that the particle multiplicity inside a jet with rescattering on is about 2% higher than with rescattering off. This increase is almost uniformly spread from the center to the periphery of the jet. The $p_\perp$-weighted jet profile is almost identical, however. Studying the jet rate itself, there is a small net reduction in the number of jets when rescattering is allowed, Figure I.9a. The difference is too small to be visible in the jet $p_\perp$ spectrum, Figure I.9b. A closer inspection shows that the jet rate above 150 GeV, i.e. in the domain of the two hard jets, is unchanged within statistics. Below that scale, however, i.e. mainly additional jets from parton showering, there is a drop by about 2% in the rate. This is most likely related to a slight leakage of hadrons out of the jet cone, shifting jet energies ever so slightly downwards. Such tiny differences could easily be tuned away, so in the end we conclude that jet properties are not measurably affected.

4.3 Collective flow

One of the telltale signs of collective behaviour is an anisotropy in the azimuthal angle of outgoing particle momenta. Here we perform a preliminary study to see whether rescattering can produce azimuthal flow at all.

In order to obtain a systematic flow, two things are required: an initial spatial anisotropy and a mechanism for collective behaviour. In this toy study an anisotropy is obtained by selecting the primary pp collisions to have their impact parameter aligned along the x



(a) (b)

Figure I.10: (a) x and y coordinates of primary hadrons, showing an initial anisotropy. (b) Azimuthal direction of momentum for outgoing hadrons, binned according to charged multiplicity. The angle is the acute angle to the event plane, $\varphi \in [0, \pi/2]$. The plot includes the spectrum for the primary hadrons, which illustrates that there is no flow before a collective behaviour has been induced by rescattering.

axis, and choosing MPI vertices according to a Gaussian distribution multiplied by a φ modulation factor with $\epsilon = 0.5$ (see Section 2.2). The resulting x-y anisotropy of primary hadron production is illustrated in Figure I.10a. This causes an elliptic flow, as shown in Figure I.10b, where the φ angle of final particle momenta is relative to the x axis (which we know to be our event plane). By the symmetry of the initial anisotropy, the shape of the spectrum should depend only on the acute angle to the event plane, $0 < \varphi < \pi/2$, and we reduce the spectrum to this range to obtain better statistics.

The flow is aligned in the y-direction, consistent with the higher density gradient in this direction. Results are binned according to the charged multiplicity, which is correlated with the impact parameter. A low multiplicity is associated with peripheral events, for which the spatial anisotropy may be strong, but collective behaviour is suppressed by the low density. A high multiplicity, on the other hand, indicates a central event with much rescattering, but a low impact parameter so a less strict azimuthal alignment. In our simple study these two effects largely cancel to give comparable asymmetries independently of the multiplicity.

Unfortunately, the aforementioned study has been made under the unrealistic advantage of a known event plane. In practice one would rather study e.g. two-particle azimuthal correlations. Furthermore, the initial anisotropy has been made implausibly large for illustratory purposes. When the simulation is repeated with more reasonable assumptions, we no longer observe any signs of flow. Therefore this brief study should be regarded as a proof of concept, and we hope to return to flow studies in the context of heavy-ion collisions, where a strong spatial anisotropy occurs naturally.

4.4 Lorentz frame dependence

The time ordering of rescatterings is not Lorentz invariant but, we do not expect this to be a major issue, since most potential rescatterings cannot influence each other. To confirm this more thoroughly, we boost the events by three units of rapidity either along or transverse to the collision axis, perform rescattering in this boosted frame, then boost back afterwards. Some results of performing this procedure, compared with the ones in the normal CM frame, are shown in Figure I.11. One may first note that the number of rescatterings and their invariant mass distribution are essentially unchanged. The rapidity spectrum of rescatterings however is somewhat deformed by the forward boost, where rescatterings would begin at around $y = -3$. Such rescatterings thus in part preempt ones at larger times in that frame. The same applies for the space–time pseudorapidity, $\eta = (1/2)\,\ln((t+z)/(t-z))$. If instead the boost is transverse, the effects on the y and η spectra are even smaller. Here collisions on the $-x$ side of the event get an earlier start than those on the $+x$ one, giving a $\pm 2\%$ modulation in the azimuthal distributions of rescatterings (not shown). These effects average out in other distributions, however, so that the $p_\perp$ and $r_\perp = \sqrt{x^2 + y^2}$ rescattering spectra are almost unchanged by transverse and longitudinal boost alike.

At the end of the day, the real test is whether observable properties are affected or not. Figure I.11e,f show that the final-state charged-hadron rapidity and $p_\perp$ spectra are almost completely insensitive to the choice of rest frame. The same also applies for other distributions we have studied, such as the azimuthal dependence, or the separate $\pi/\mathrm{K}/\mathrm{p}$ spectra. The breach of Lorentz frame independence therefore is a negligible issue for our studies.

4.5 Rescatter rates

In this section we study how common rescatterings are, both overall and subdivided by hadron species and by process types. The average number of rescatterings per (inelastic) nondiffractive pp event is shown as a function of the collision energy in Figure I.12a. It is compared to the primary hadron multiplicity, i.e. the hadrons produced directly from the fragmenting strings, and to the final charged multiplicity. Note that these latter two are almost equal; the multiplicity increase from the decays of primary hadrons is compensated by the decrease from the exclusion of neutral particles. This largely holds also on an event-by-event level, so we may use the observable charged multiplicity as a simple measure of number of primary hadrons that may rescatter. As an order-of-magnitude, the average number of rescatterings $\langle n_{\mathrm{rescatter}} \rangle$ is about half that of the primary multiplicity $\langle n_{\mathrm{primary}} \rangle$. While the number of potentially colliding pairs increases like n^2_{primary}, the dashed line represents a fit according to a much slower $\langle n_{\mathrm{primary}} \rangle^{1.2}$. The reason is that the system size also increases with energy. Obviously so in the longitudinal direction, but also in the transverse one, by an increasing MPI perturbative activity spreading production vertices

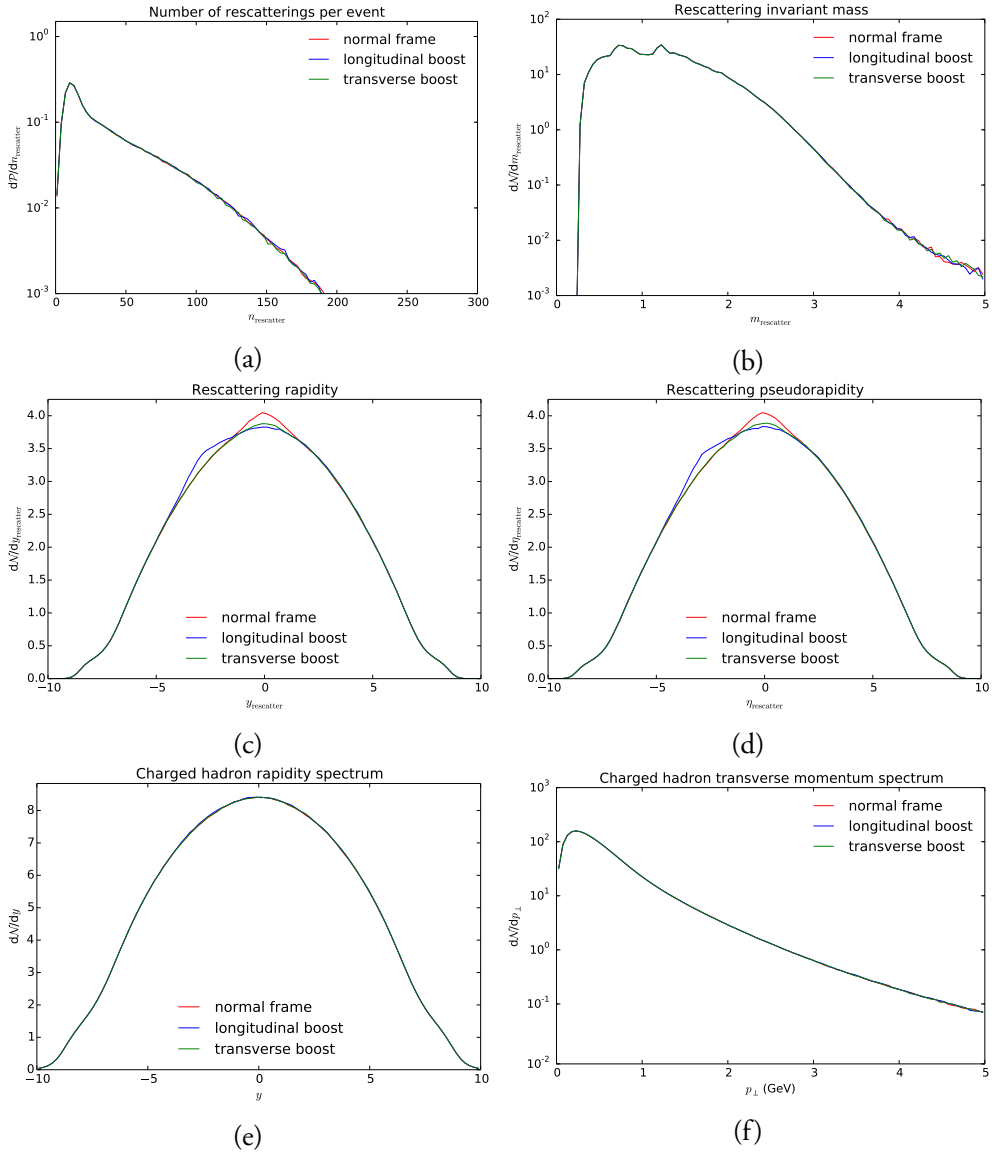


Figure I.11: Effects of modified time ordering on 13 TeV pp nondiffractive collisions, where ordering is either in the normal rest frame, or in a frame boosted either longitudinally or transversely by three units of rapidity. (a) Number of rescatterings. (b) Invariant mass distribution of rescatterings. (c) Rapidity distribution of rescatterings. (d) Distribution in $\eta = (1/2)\,\ln((t+z)/(t-z))$ of rescatterings. (e) Rapidity distribution of final charged hadrons. (f) Transverse momentum spectrum of final charged hadrons.

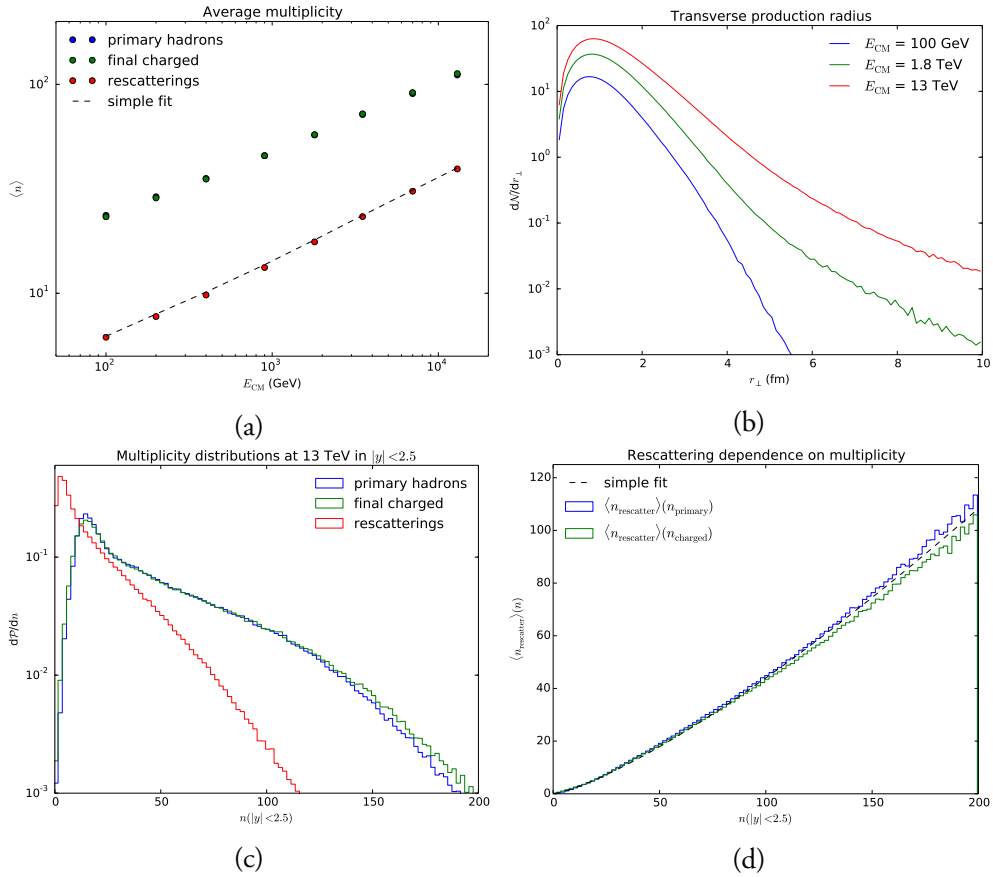


Figure I.12: (a) Energy dependence of multiplicities in nondiffractive pp collisions. (b) Primary hadron production in $r_\perp = \sqrt{x^2 + y^2}$ at three energies. (c) Distribution in the numbers of primary hadrons, charged hadrons and rescatterings in the central $|y| < 2.5$ region of 13 TeV nondiffractive pp collisions. (d) Multiplicity dependence of the number of rescatterings in events as above.

over a larger transverse area, Figure I.12b.

Zooming in on the central rapidity region of 13 TeV nondiffractive events, the different kinds of multiplicity distributions are displayed in Figure I.12c, and the rescattering rate as a function of the primary or charged multiplicity in Figure I.12d. In the latter, a simple fit $\langle n_{\mathrm{rescatter}} \rangle \propto n_{\mathrm{primary}}^{1.3}$ has been inserted to guide the eye, showing a similar scaling as for the energy dependence. The power 1.3 also describes the dependence in the event as a whole, without the $|y| < 2.5$ restriction.

With well over a hundred different hadron species that can be produced, the number of different colliding hadron pairs are in the thousands, even if most of them are quite rare. To

Table I.6: Number of collisions per 13 TeV nondiffractive pp event, of different incoming particle combinations, where particles have been grouped so as to avoid too fragmented a view. M represents other meson species and B other baryon ones. All combinations with a rate below 0.1 have been summed into the "other" group.

incoming	rate	incoming	rate	incoming	rate
$\pi + \pi$	12.63	$\mathrm{K} + \mathrm{N}$	0.39	$\eta/\eta' + N$	0.19
$\pi + \rho$	4.59	$\rho + \rho$	0.38	$\pi + \mathrm{B}$	0.18
$\pi + \mathrm{K}$	3.84	$\rho + \mathrm{N}$	0.36	$N + \Delta$	0.16
$\pi + \mathrm{N}$	3.44	$\rho + \omega/\phi$	0.34	$\pi + \Sigma^*$	0.15
$\pi + \omega/\phi$	2.08	$\rho + \eta/\eta'$	0.30	$\rho + \Delta$	0.14
$\pi + \eta/\eta'$	1.80	$\pi + f_0(500)$	0.29	$\eta/\eta' + \omega/\phi$	0.14
$\pi + \mathrm{K}^*$	1.33	$\mathrm{K} + \omega/\phi$	0.27	$\pi + \mathrm{M}$	0.12
$\pi + \Delta$	1.10	$\mathrm{K} + \mathrm{K}$	0.26	$\mathrm{K} + \Delta$	0.11
$\rho + \mathrm{K}$	0.54	$\pi + \Lambda$	0.25	$\mathrm{K}^* + \mathrm{N}$	0.11
$\pi + \Sigma$	0.46	$\omega/\phi + \mathrm{N}$	0.24		
$\mathrm{N} + \mathrm{N}$	0.46	$\mathrm{K} + \eta/\eta'$	0.23		
$\mathrm{K} + \mathrm{K}^*$	0.41	$\rho + \mathrm{K}^*$	0.20	other	1.87

give some feel, Table I.6 shows the most common groups of hadron pairs. Here π represents all pions, K all Kaons ($\mathrm{K}^{\pm}$, K^0, $\bar{\mathrm{K}}^0$, $\mathrm{K}^0_{\mathrm{S,L}}$), N all nucleons (p, n, $\bar{\mathrm{p}}$, $\bar{\mathrm{n}}$), and so on. As can be seen, $\pi\pi$ rescatterings dominate by far, constituting about a third of all rescatterings, while π with anything else constitutes another third. This highlights the importance of accurate cross sections for processes involving pions.

Collisions are also characterized by which type of process occurs, Table I.7. The resonant, elastic and nondiffractive types dominate by far. Baryon–antibaryon annihilation is small but not negligible for the baryon subclass of particles. Diffraction and excitation require more phase space to occur, and therefore become suppressed.

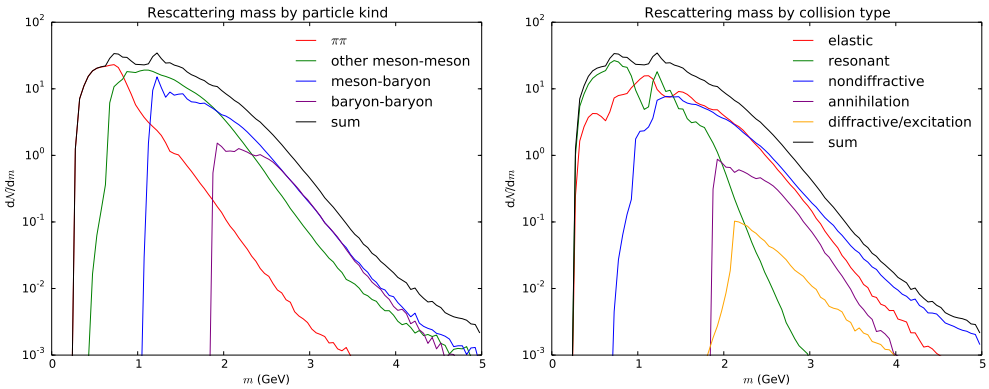
It is also interesting to study the invariant mass spectrum of collisions, Figure I.13. There is a natural steep fall-off with mass for two particles to come close to each other, because of the way the fragmentation process correlates the space–time and energy–momentum pictures. Near each mass threshold there is also a phase-space suppression factor. On top of that the individual cross sections can give a more serrated shape for each collision type separately, mainly from resonance contributions, but these largely average out in the overall picture.

4.6 Model variations

As part of the new framework, several parameters and settings have been introduced. In this section, we study how changing these settings affects rescattering phenomenology. In particular, as a simple and direct test, we present how each main model setting impacts the average number of rescatterings per event. In addition to these new settings, we also study

Table I.7: Number of collisions of different types per 13 TeV nondiffractive pp event.

Process type	rate
resonant	17.80
elastic	14.08
nondiffractive	6.92
annihilation	0.49
diffraction + excitation	0.05



(a) (b)

Figure I.13: Invariant mass distributions of rescattering pairs in 13 TeV nondiffractive pp events. (a) Grouped by incoming hadron kinds. (b) Grouped by process type.

existing settings that could have an effect on rescattering. A summary of settings and their overall effects is given in Table I.8, with the average number of rescatterings for different variations shown in Table I.9. In more detail, the effect of the settings are as follows:

- `Rescattering:impactModel` describes how the rescattering probability depends on the impact parameter b. The default (1) is a Gaussian fall-off, while the alternative (0) is a sharp edge, see eqs. (I.8) and (I.10). In a uniform medium the two alternatives are normalized to result in equal rescattering rates, as given by the cross section. In practice we see that the Gaussian option gives more long-range interactions, Figure I.14a, as expected, but overall a somewhat reduced rescattering rate. This is because the particle density falls off from the central collision axis, such that there are fewer pairs at large than at small impact parameter to begin with. The fact that the Gaussian option gives a lower rescattering rate means that the loss of events in the important 0.3–0.7 fm region for the Gaussian model is not compensated for by including longer-range interactions.

- `Rescattering:opacity` is the rescattering probability at $b = 0$, i.e. P_0 of eqs. (I.8) and (I.10). A lower opacity reduces the probability of close interactions, but increases

Table I.8: List of model choices and parameters used to study the range of possible rescattering effects, with their effect on the rescattering rate. Parameter names are as defined in the PYTHIA user interface. See the text for more detailed information.

Setting	Default	Effect on rescattering rate
`Rescattering:impactModel`	1 (Gaussian)	Black disk gives more
`Rescattering:opacity`	0.9	Larger values give more
`Rescattering:quickCheck`	on	Turning it off gives more
`Rescattering:nearestNeighbours`	on	Turning it off gives less
`Rescattering:tauRegeneration`	1.	Larger values give less
`HadronVertex:mode`	0	± 1 gives much more/less
`HadronVertex:kappa`	1.	Larger values give more
`HadronVertex:xySmear`	0.5	No significant effect
`PartonVertex:modeVertex`	2 (Gaussian)	Has a small effect
`PartonVertex:ProtonRadius`	0.85	Larger value gives less
`PartonVertex:EmissionWidth`	0.1	No significant effect

the range of interactions. This gives fewer rescatterings, for the same reason as above.

- `Rescattering:quickCheck` enables a simple check that tests whether two hadrons are moving away from each other at their respective time of creation in the CM frame of the event, and if so does not study further whether a rescattering is possible. This is faster than the more time-consuming full check, where the hadron pair is boosted to their common rest frame and the earliest particle is offset to a common time of creation before checking whether the hadrons move away from each other. Performing the quick check first reduces the total execution time by about a factor of two, since the number of hadron pairs to consider in an LHC event may be of the order of 10 000, whereof the vast majority are moving away from each other by any criterion (note that the full check is still performed on pairs that pass the simple check). The simple check rejects about 5% of the collisions that would have been accepted by the full check, but these false rejections typically are close to the (unphysically sharp) accept/reject border, and do not make a significant impact on rescattering distributions. For these reasons the quick check is on by default.

- `Rescattering:nearestNeighbours` allows hadrons that are produced as nearest neighbours along a string to rescatter against each other, see Section 2.3. The number of rescatterings goes up when on, but net effects do not change in proportion, since nearest-neighbour pairs are more likely to move in the same direction anyway.

- `Rescattering:delayRegeneration` and `Rescattering:tauRegeneration` are based on the assumption that it takes some formation time for a scattered hadron to build up a new wave function, and that during that time it has a reduced likelihood to scatter again. If `delayRegeneration` is switched on, this time is chosen at random according to an exponential distribution with average proper time (in fm)

Table I.9: Average number of rescatterings per event, when varying different settings individually. Events are `SoftQCD:nonDiffractive` processes at 13 TeV, using `MultipartonInteractions:pT0Ref = 2.345`.

Setting	$n_{\text{rescatter}}$
Default	39.2
`Rescattering:impactModel = 0`	45.5
`Rescattering:opacity = 0.8`	37.3
`Rescattering:opacity = 1.0`	40.8
`Rescattering:quickCheck = off`	40.8
`Rescattering:nearestNeighbours = off`	25.4
`Rescattering:tauRegeneration = 0.0`	45.4
`Rescattering:tauRegeneration = 2.0`	38.4
`HadronVertex:mode = -1`	64.0
`HadronVertex:mode = 1`	21.7
`HadronVertex:kappa = 0.8`	32.8
`HadronVertex:kappa = 1.2`	44.4
`HadronVertex:xySmear = 0.3`	40.2
`HadronVertex:xySmear = 0.7`	39.1
`PartonVertex:modeVertex = 1`	39.6
`PartonVertex:protonRadius = 0.7`	39.3
`PartonVertex:protonRadius = 1.0`	39.1
`PartonVertex:EmissionWidth = 0.0`	39.6
`PartonVertex:EmissionWidth = 0.2`	39.2

given by the `tauRegeneration`. Hadrons produced from string fragmentation are not affected, since they get their time offset from the hadronization process itself, roughly corresponding to an average τ of 1.5 fm. Setting $\tau_{\text{regen}} = 1$ fm reduces the number of rescatterings by about 10% relative to an instantaneous regeneration. The effect seems to saturate however, and increasing it to 2 fm does not make much further difference.

- `HadronVertex:mode` defines where the hadron vertex is placed in string hadronization. By default, hadrons are defined to be produced at the average location of the two string breaks that define it (see Figure I.1). By setting `HadronVertex:mode = 1`, the production vertex is shifted forward in time to the point where the two colour endpoints meet for the first time, and setting it to -1 shifts it backwards in time by that same amount. These variations have a significant effect on the density of primary produced hadrons, changing the number of rescatterings by about 50%. For this reason we do not vary this setting in our studies, but instead use `HadronVertex:kappa`, which gives similar but milder effects, as explained below.

- `HadronVertex:kappa` is the string tension, by default $\kappa \approx 1$ GeV/fm, eq. (I.1). Increasing κ compresses the production vertices and thus gives more rescattering. While the concept of a string tension is central in the hadronization framework, its

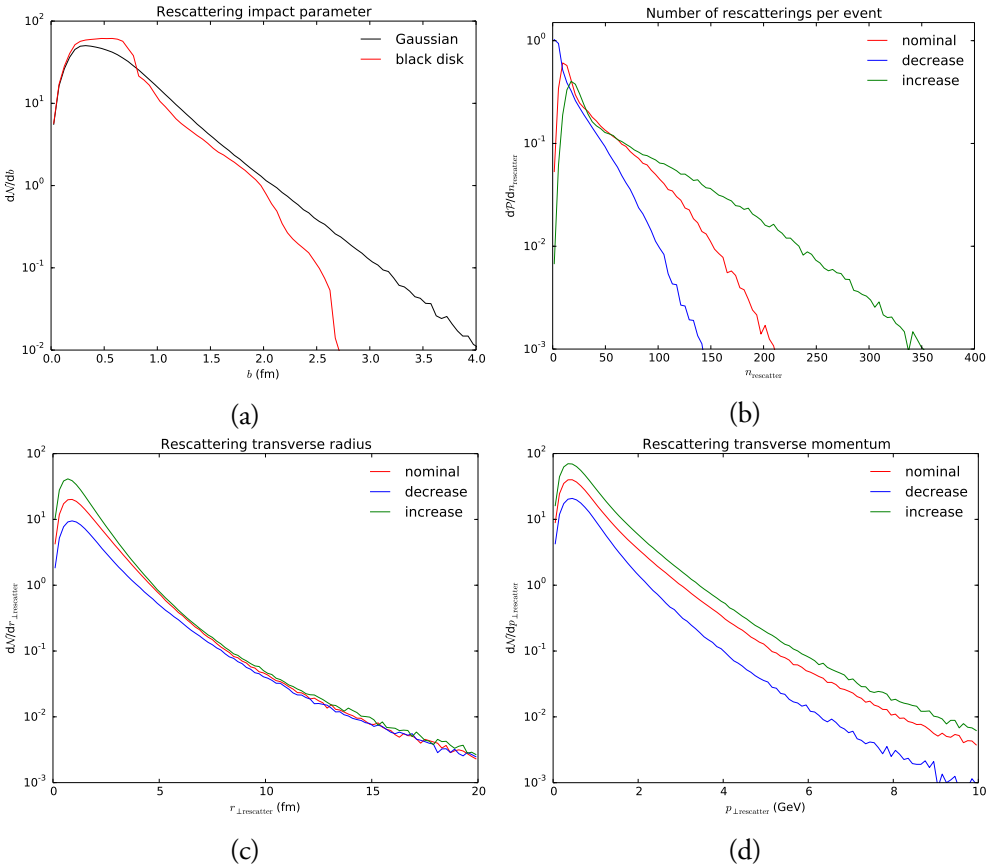


Figure I.14: (a) The impact-parameter distribution of rescatterings for the different impact models. (b) Number of rescatterings per event. (c,d) Distribution of rescatterings in $r_\perp$ and $p_\perp$. Results are for 13 TeV nondiffractive pp events.

exact value has not been relevant for the energy–momentum-related properties of an event. We allow for a generous $\pm 20\%$ variation to also cover some uncertainty in how to define the hadron production vertex, as described above.

- `HadronVertex:xySmear` is the width of a Gaussian smearing of string breakup vertices in the plane perpendicular to the string, see Section 2.2. Increasing this slightly increases the transverse offsets of the primary produced hadron vertices, but does not have significant overall effects on rescattering.

- `PartonVertex:modeVertex` picks the shape of the overlap region between the two incoming protons, as used to pick the location of MPI vertices, see Section 2.2. Different shapes give some variation in rescattering features, but they are small ones for most properties, and it is hard to quantify the difference between the various

Table I.10: List of model settings used to explore the range of possible rescattering effects. Here "increase" and "decrease" denote alternatives with more or less amount of rescattering relative to the default "nominal" values.

Setting	decrease	nominal	increase
`Rescattering:impactModel`	1	1	0
`Rescattering:opacity`	0.8	0.9	1.0
`Rescattering:quickCheck`	on	on	off
`Rescattering:nearestNeighbours`	off	on	on
`Rescattering:tauRegeneration`	2.	1.	0.
`HadronVertex:kappa`	0.8	1.	1.2
`PartonVertex:ProtonRadius`	1.0	0.85	0.7
`MultipartonInteractions:pT0Ref`	2.305	2.345	2.385

shapes. For this reason, we do not vary this setting in subsequent model tests. It is however a way to introduce spatial anisotropy in the primary hadron distribution, which is necessary for azimuthal flow.

- `PartonVertex:ProtonRadius` is the three-dimensional proton radius, which then gets converted to a two-dimensional one for the distribution of MPI production vertices, eq. (I.3). Increasing/reducing this by 0.15 fm will increase/reduce the transverse radius of rescattering vertices by about 0.10 fm, and higher values give a slightly lower number of rescatterings.

- `PartonVertex:EmissionWidth` is the constant of proportionality for smearing of the transverse production vertices generated by partons showers, which are assumed to be inversely proportional to the $p_\perp$ of the parton. Varying this within a reasonable range has no significant effect on rescattering.

For comparison purposes, one nominal scenario is defined as our best assumption on relevant settings, and in addition two extremes with decreased or increased rescattering rate, Table I.10. For each case, `pT0Ref` has been tuned as shown in the table in order to restore charged multiplicity.

The resulting variations of rescattering rates are shown in Figure I.14b. The rate difference mainly arises around small transverse radii, Figure I.14c (and early invariant times, not shown). By contrast, in properties such as the transverse momentum, Figure I.14d, or invariant mass of the collision systems (not shown), the variations more affect the normalization than the shape of the distributions. Comparisons to data will be given in Section 5.2.

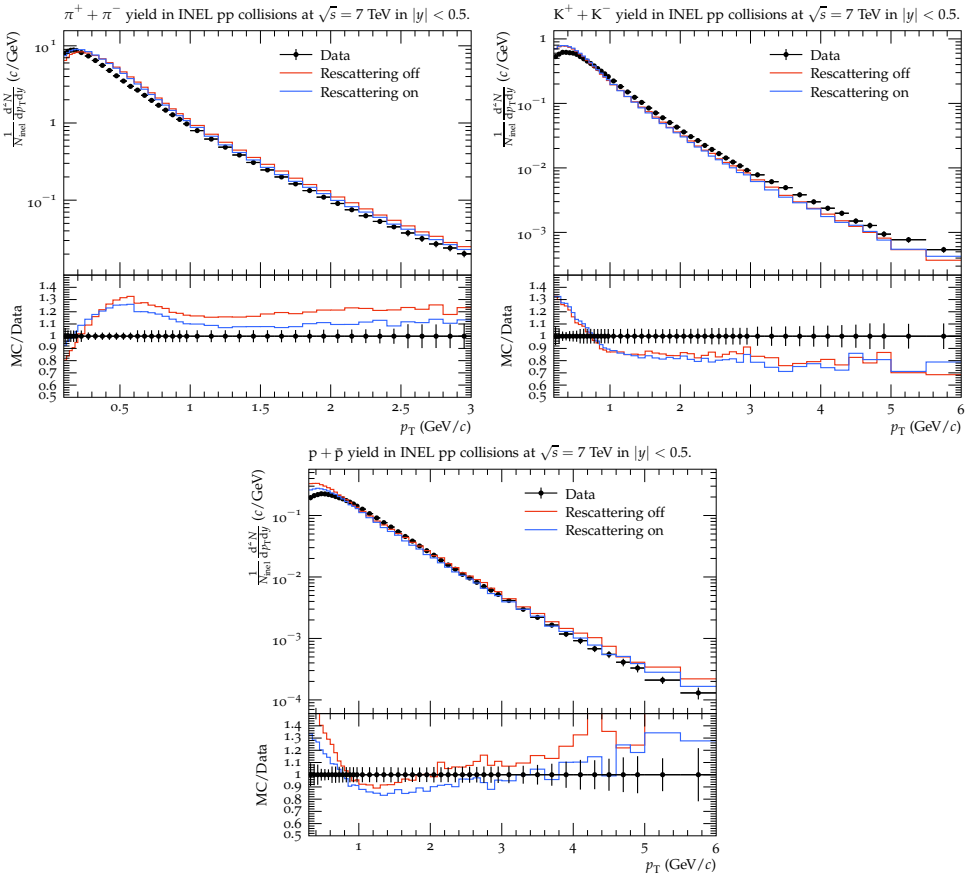
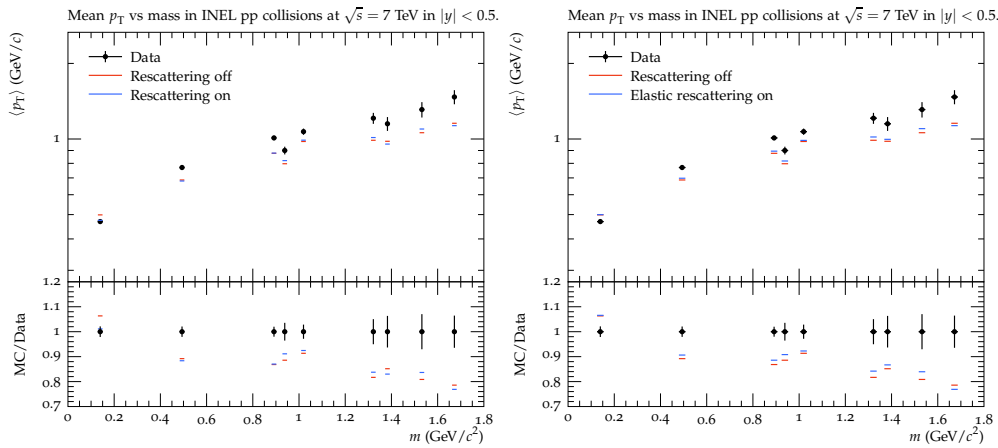


Figure I.15: $p_\perp$ spectra for $\pi^\pm$, $K^\pm$ and $p/\bar{p}$, compared with data from ALICE [4, 8].

5 Comparison with data

While the standard Pythia generally gives a good description of LHC pp data, there are some well-known discrepancies. One such is the shape of low-$p_\perp$ spectra of pions, Kaons and protons. Especially the poor description of the pion spectrum for $p_\perp < 0.5$ GeV has direct consequences for a number of other distributions [64], e.g. when the pseudorapidity spectrum is studied either for $p_\perp > 0.1$ GeV or $p_\perp > 0.5$ GeV charged particles. In this section, we study how these spectra are changed by rescattering, using Rivet [65] to generate plots and comparisons to data. Results are shown initially for the default rescattering model, then for alternative parameter choices within this model, and eventually for model variations of the primary hadron production. Finally, we briefly consider the $p_\perp$ spectrum for the Λ^0/K^0_S ratio. As before, the `pT0Ref` parameter is retuned to ensure the same charged multiplicity in all scenarios studied.



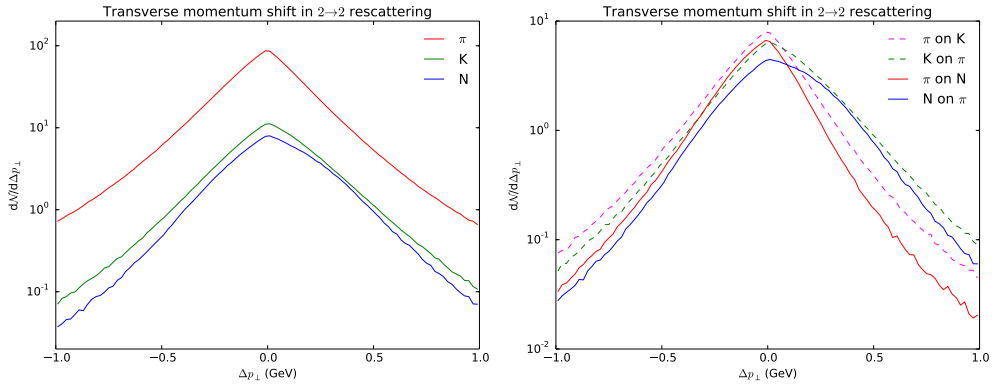
(a) (b)

Figure I.16: Average $p_\perp$ for different particle species, ordered by mass, with data from ALICE [4, 8]. The included particles are $\pi^\pm$, $K^\pm$, $K^*(892)^\pm$, p, $\phi(1020)$, Ξ^-, $\Sigma^*(1385)^\pm$, $\Xi^*(1530)^0$ and Ω^-. (a) Comparison of rescattering to no rescattering. (b) Comparison between the two when all rescatterings are forced to be elastic. Here we use the default `pT0Ref = 2.28`, since elastic scattering does not change charged multiplicity.

5.1 The effects of rescattering on transverse momentum speectra

Figure I.15 shows the $p_\perp$ spectra for pions, Kaons and protons, with and without rescattering. We see that rescattering gives a better fit to data for pions and protons, especially at low $p_\perp$, while for Kaons rescattering moves the $p_\perp$ spectrum away from data. The average $p_\perp$ for various particle species is shown in Figure I.16a, and here again there is an improvement for π and p, but a slight deterioration for K.

If we consider only elastic collisions, one would expect that rescattering should push lighter particles towards lower $p_\perp$ and heavier particles to higher $p_\perp$. This is because lighter particles generally move faster and will catch up with and push the heavier ones outwards, a phenomenon sometimes referred to as "pion wind". The actual momentum shifts in elastic rescatterings (including through resonances) is shown in Figure I.17. Here we see a positive shift both for K and N. This becomes more apparent if one considers only $K\pi \to K\pi$ and $N\pi \to N\pi$ scatterings, Figure I.17b, where the heavier K/N on the average gains $p_\perp$ at the expense of the lighter π. A closer study reveals that the strongest $p_\perp$ shifts comes from resonance production, i.e. K^* and Δ intermediate states. There are two reasons for this. Firstly, these resonances give large cross sections in a mass range where the flux of colliding pairs is large in the first place, and thus dominate over elastic scattering (in the processes discussed here). Secondly, elastic scattering is peaked in the forward direction, i.e. at small



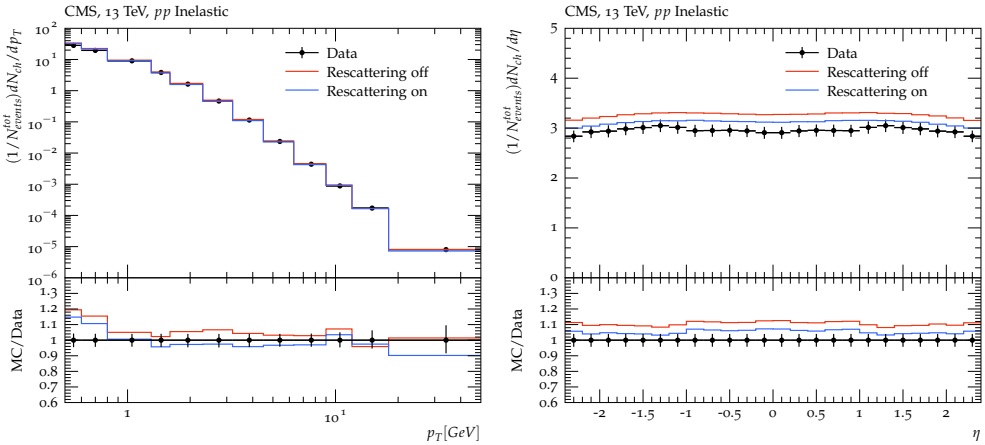
(a) (b)

Figure I.17: Shift of transverse momentum by $2 \to 2$ elastic or resonant processes, where positive numbers correspond to an increased $p_\perp$ in the collision. (a) Inclusive shifts for π, K and N (including antiparticles). (b) Shifts in $\mathrm{K}\pi \to \mathrm{K}\pi$ and $\mathrm{N}\pi \to \mathrm{N}\pi$ scatterings.

momentum transfers, while an s-channel spin 0 resonance decays isotropically in its rest frame.

In Figure I.16b, we look at $\langle p_\perp \rangle$ shifts when only elastic scattering is permitted. Specifically, this is done by calculating each total cross section as before, but setting the elastic cross section equal to the total one (thus excluding elastic scattering through a resonance). In this case, the $\langle p_\perp \rangle$ increases for all heavy particles except for Ω, which is so rare so this can simply be explained by statistical fluctuations. For particles such as p and Σ, the change in $\langle p_\perp \rangle$ is less than before, highlighting the fact that elastic scattering through a resonance gives the strongest momentum transfers. (As a side note, an unexpected observation is that the average pion $p_\perp$ actually increases very slightly, which turns out to be a consequence of the narrow rapidity window $|y| < 0.5$ used in the experimental analysis; the average does decrease if all rapidities are included.)

So why then is the mean $p_\perp$ reduced for Kaons when inelastic interactions are allowed? The answer is that in processes classified as inelastic, especially non-diffractive processes, we make a significant effort to ensure that at least three particles are produced, so as to avoid the elastic channel. Such interactions have to share the $p_\perp$ between more outgoing than incoming particles, which leads to a reduced average. In principle, the opposite kind of interactions would be possible, where three (or more) incoming particles could fuse to give two outgoing ones, presumably then with an increased $p_\perp$. We have not implemented these kinds of processes in the first version of our framework, but their potential effect on the Kaon $p_\perp$ spectrum should make them a priority in future work.



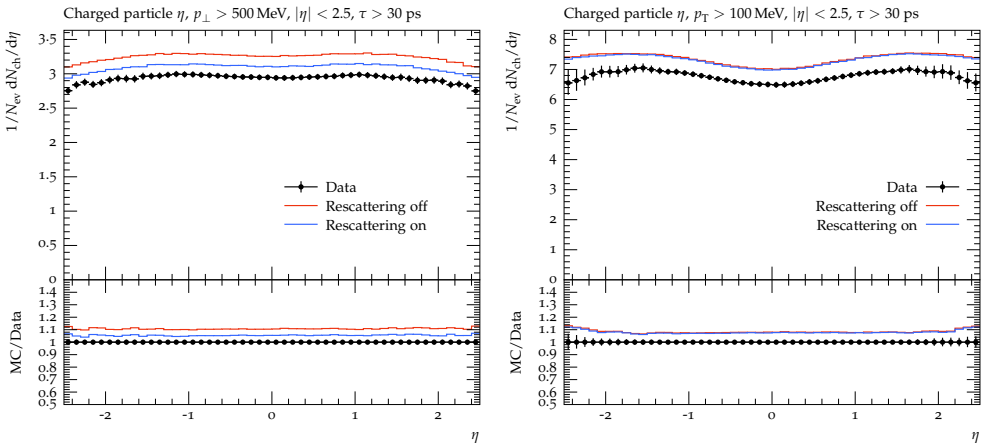
(a) (b)

Figure I.18: $p_\perp$ and η spectra compared with data from CMS [66]. Charged particles with $p_\perp > 500$ MeV and $|\eta| < 2.4$ are considered.

Another observation from Figure I.16a is that the mean $p_\perp$ of Σ^* is also reduced. In addition to the aforementioned effect of $2 \to n$ scattering, we have also observed that resonances formed during rescattering tend to have a lower $p_\perp$ than those produced directly from string fragmentation. From phase space considerations, it is less likely for two random high-$p_\perp$ particles to have an invariant mass in the resonance range than for two low-$p_\perp$ ones. The effect is especially large where the mass difference between the resonance and the particles forming it is small, such as for the Σ^* baryons. These particles still tend to gain $p_\perp$ when they themselves participate in rescattering, as we see in Figure I.16b.

The total $p_\perp$ spectrum for all charged particles is shown in Figure I.18a, and is improved overall by rescattering. The charged-particle pseudorapidity spectra in Figure I.18b and I.19a show that when a cut $p_\perp > 500$ MeV is used, rescattering shifts the spectrum down by an approximately fixed amount, to a better agreement with data. However, this improvement is not visible in Figure I.18b, where the cut is $p_\perp > 100$ MeV. This suggests that the "true" pseudorapidity spectra are mostly unaffected by rescattering, but because of $p_\perp$ shifts, rescattering has an indirect effect on the observed spectrum. The takeaway from this is that data affected by low-$p_\perp$ particle production are likely to be better described when rescattering is included.

In summary, rescattering does what it is expected to in elastic scattering, i.e. slows down lighter hadrons and speeds up heavier ones. The disappointing aspect is that we have observed other mechanisms that work in the other direction. Finding ways to compensate for these effects should be addressed in future work.



(a) (b)

Figure I.19: Charged particle η spectra compared with data from ATLAS [67, 68], with cuts (a) $p_\perp > 500$ MeV, and (b) $p_\perp > 100$ MeV.

5.2 Model dependence of transverse momentum spectra

Given the central role of the $p_\perp$ spectra, it is highly relevant to understand how sensitive they are to rescattering model variations. To this end, we can compare the default rescattering scenario with the two alternatives listed in Table I.10. These two are selected to minimize or maximize the number of rescatterings, within reasonable extremes for each relevant setting.

The results are shown in Figure I.20. What we observe is that the effects on the $p_\perp$ spectra tend to scale with the amount of rescattering. This is especially clear for π and p, where the minimum/maximum amount of rescattering give smaller/larger effects than the default values, respectively. At the same time, the maximum setup gives a relatively small further improvement over the default rescattering one. It is therefore meaningful to stay with the default scenario, rather than trying to use more extreme choices to come closer to data.

5.3 The thermal model alternative

The rate of $q\bar{q}$ string breaks is traditionally assumed to involve a suppression factor $e^{-\pi m_{\perp q}^2/\kappa}$: since the string does not contain any local concentrations of mass, a quark needs to tunnel out as a virtual particle until it has "eaten up" enough string length to correspond to its transverse mass [19]. This gives a Gaussian $p_\perp$ spectrum to quarks and, by addition, to hadrons. The derivation is done for a single string in isolation, however, whereas the reality at hadron colliders is that the typical event contains several more-or-less overlapping strings. This may modify the primary particle production processes, which set the starting

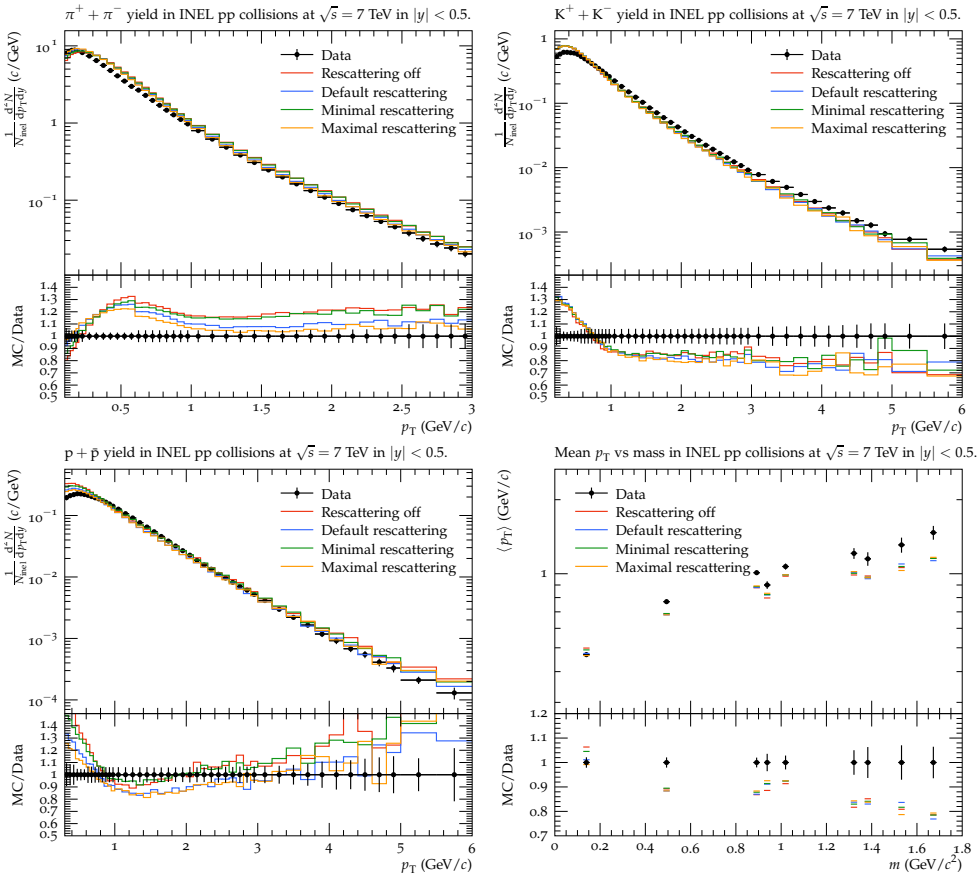


Figure I.20: $p_\perp$ spectra for $\pi^\pm$, $K^\pm$ and $p/\bar{p}$ and average $p_\perp$ for various particles, for different parameter configurations.

stage for the continued rescattering and decay processes we have considered in this article. Empirically, an exponential spectrum $\exp(-m_{\perp\mathrm{had}}/T)$ was early on proposed as a parameterization of hadron collision data, where $m_{\perp\mathrm{had}}$ is the transverse hadron mass and T could be associated with a temperature e.g. in the Hagedorn approach [69–71]. Interestingly, an effectively exponential fall-off could arise also starting from the Gaussian one, by assuming that the string tension is fluctuating along the string length, also in the absence of other strings [72].

Based on such ideas, a "thermal model" option has been included as an alternative in Pythia [16]. Unlike purely statistical models, however, it is strictly based on the string model, with local flavour and $p_\perp$ conservation. To this end, each $q\bar{q}$ breakup is associated with a (modified Bessel) $p_\perp$ distribution such that the two-dimensional convolution results in an $\exp(-p_{\perp\mathrm{had}}/T)$ spectrum. In each fragmentation step, an old q flavour is always known

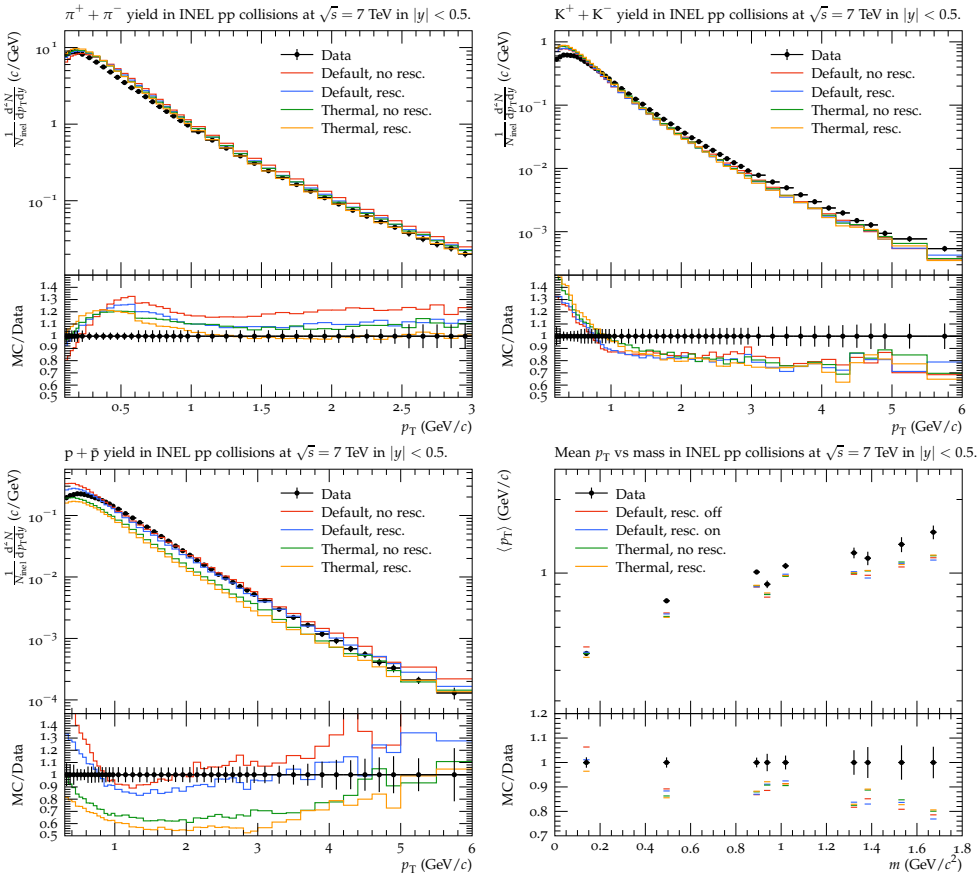


Figure I.21: $p_\perp$ spectra for $\pi^\pm$, $K^\pm$ and $p/\bar{p}$ and average $p_\perp$ for various particles, with comparing the Gaussian to the thermal model. When using the thermal model, `pT0Ref` has been tuned to 2.47 without rescattering and 2.52 with rescattering on, in order to maintain the correct n_{charged}.

when the new one is selected and a new hadron is formed out of the two. Each new quark and hadron possibility is assigned a relative weight $\exp(-m_{\perp\text{had}}/T)$, times relevant spin and symmetry factors, and these weights are used to make the random choice. The relative rate of diquark/baryon production requires a free parameter, while an additional s-quark suppression factor is needed to achieve better agreement with observed production rates. The suppression of multistrange hadrons is underestimated, however, whereas the standard string model overestimates it, suggesting that "the truth" may lie somewhere in between.

A key aspect of the $\exp(-m_{\perp\text{had}}/T)$ weight is that heavier primary hadrons obtain a larger $\langle p_\perp \rangle$ than lighter ones. While it does enhance low-$p_\perp$ pion production and deplete ditto baryon one, relative to the traditional string model, the effects are not large enough to

explain the data [16]. It is therefore interesting to combine the thermal model with rescattering, to check whether the two together give a larger combined improvement than each individually. The results of this comparison are shown in Figure I.21, where the Gaussian model is compared to the thermal model, both with and without rescattering. The effects of the thermal model are similar to the effects of rescattering, with an improvement for the mean $p_\perp$ of pions and protons and a deterioration for Kaons. For pions, the correction from the combination of the two in fact overshoots the $\langle p_\perp \rangle$ data, so that either of them individually gives a better result than the two combined, even if the pion $p_\perp$ spectrum itself looks rather reasonable. We also see that the $p_\perp$ spectrum for protons is less accurate, especially at higher $p_\perp$s. For these reasons, the results of using the thermal model are not particularly encouraging, at least not without a more thorough retuning.

5.4 Close-packing

Apart from the possibility of a randomly fluctuating string tension, one may also expect systematic effects on the tension in a denser string environment, which can be modelled in different ways. One option implemented in Pythia is that of colour ropes [14], wherein several more-or-less parallel strings can fuse into a "rope". The combined colour charge of this rope, as given by the Casimir operator, then gives a scaling-up factor applied to the string tension. When the rope breaks, the difference in charge before and after the break gives the effective charge involved in that $q\bar{q}$ production step. The other option is based on the assumption that a close-packing of strings gives them a smaller transverse area each, but preserves their separate identities [16]. Also in this option the string tension is increased, but in principle as a smooth function of the amount of squeezing rather than in the discrete steps of the rope. In practice, there need not be any big difference between these two options, but in this study we choose the second one for simplicity.

In this model, the creation of a new hadron is begun by an exploratory step ahead, so that the number of strings overlapping the rapidity range of the intended next hadron can be estimated. This local string number is then raised to some (tuned) power to give a rescaling factor for the string tension. To this basic picture some damping is introduced for particle production at large $p_\perp$, which typically occurs at larger transverse radii, away from the denser region. Note that the current implementation predates the introduction of space–time coordinates for the hadronization process, such that there now is room for improvements, but not ones that are likely to give a qualitatively changed behaviour for the properties studied here.

The close-packing modification can be used either for the standard string model or for the thermal alternative, by a rescaling either of κ or of T. In Figure I.22, we have used the former one. The trend here is that close-packing tends to increase $p_\perp$ for all particles, which means an improvement for all heavier hadrons, especially Kaons whose $p_\perp$ spectrum now

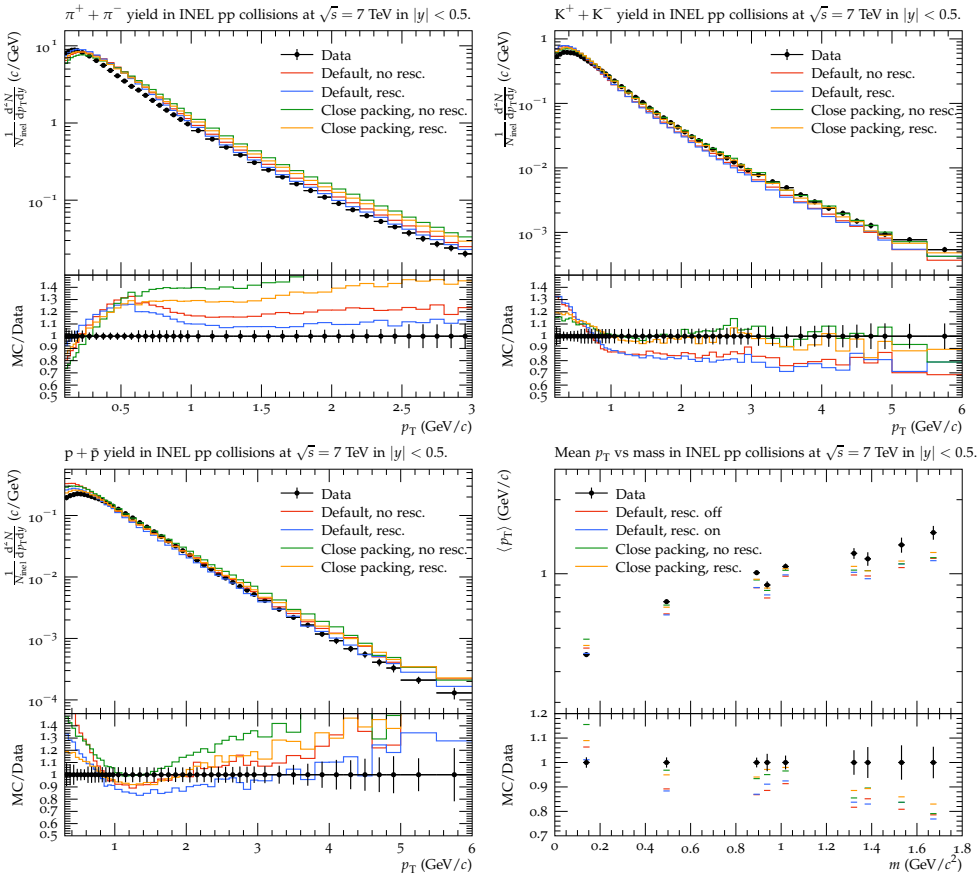


Figure I.22: $p_\perp$ spectra for $\pi^\pm$, $K^\pm$ and $p/\bar{p}$ and average $p_\perp$ for various particles, for the Gaussian model with and without close-packing corrections. When using the close-packing corrections, `pT0Ref` has been tuned to 2.18 without rescattering and 2.25 with rescattering on, in order to restore the correct n_{charged}.

follows data remarkably well above 1 GeV. However, this also means that the spectrum is worsened pions, and looking at their spectrum, the effect is quite severe. This deterioration is partially compensated for by rescattering, but not completely. This makes the close-packing option unsuited as it stands. A retuning of fragmentation parameters might ameliorate the situation, but that is beyond the scope of the current study.

5.5 The role of vector mesons

One of the standard assumptions is that the $p_\perp$ spectrum in $q\bar{q}$ string breaks is the same, independent of the quark species. This needs not be the case, and higher-order corrections

could well favour slightly different $p_\perp$ values for strange quarks [73, 74], but for now we assume it to hold. Similarly, primary pseudoscalar and vector mesons are assumed to have the same $p_\perp$ spectra. The correct relative fraction of the two kinds of mesons is not known a priori, however, and for many hadrons it is difficult to measure their production rates, especially those with large widths. The prime example is the ρ, which we have seen contributes non-negligibly to the total rescattering rate. Since the ρ has a higher mass than the Kaon, elastic ρK collisions would tend to reduce the Kaon $p_\perp$, partially counteracting the gain from Kπ collisions.

As a simple test of the significance of heavy primary hadrons, we have studied a toy scenario where no vector mesons at all are produced in the primary string fragmentation, but still can occur as intermediate states during rescattering. The resulting $p_\perp$ spectra are shown in Figure I.23. No attempt at a complete retune has been made, so it is the change by rescattering that is most interesting, not the overall agreement. Not unexpectedly, the $\langle p_\perp \rangle$ is wildly off for $K^*(892)$ and $\phi(1020)$, which now cannot be produced in the primary process. The "pion wind" effect is still there, in that rescattering shifts pions to smaller $p_\perp$ and protons to larger. For Kaons the $\langle p_\perp \rangle$ is still decreased by rescattering, providing further support that the primary mechanism for the Kaon $p_\perp$ loss is through $2 \to n$ processes, rather than from Kaon collisions with heavier particles.

It could have been informative also to go in the other direction, and include primary production of higher resonances, with orbital or radial excitations. Measurements at LEP show that such mesons are produced at a non-negligible rate [46]. And yet, their explicit inclusion tend to reduce the goodness of fit to many other properties, presumably because the assumed isotropic decay distributions do not represent the correct physics. Instead a higher-mass state could be viewed as a longer-than-normal string piece, with a decay along this string direction, just as if these products come directly from the string. Therefore we do not expect primary production of higher resonances to change $p_\perp$ properties appreciably, but currently do not have the full machinery necessary to test this assumption.

5.6 Other transverse momentum spectra

So far we have focused on $p_\perp$ spectra for pions, Kaons and protons. However, another experimental observation that pertains to collective behaviour is the peak for example in the Λ^0/K^0_S ratio around $p_\perp \approx 2$ GeV. In Figure I.24, the ratios for Λ^0/K^0_S and Ξ^-/Λ^0 are shown. Unfortunately rescattering does not provide an improvement. If anything it causes a deterioration, by reducing the relative number of Λ^0 and Ξ^- baryons through the baryon-antibaryon annihilation mechanism. As before, an inclusion of $3 \to 2$ processes could help alleviate the problem, but hardly give full agreement. In general, baryon production has been one of the more complicated and least successful aspects of the string fragmentation framework, already in the simpler $\mathrm{e}^+\mathrm{e}^-$ environment, and remains so.

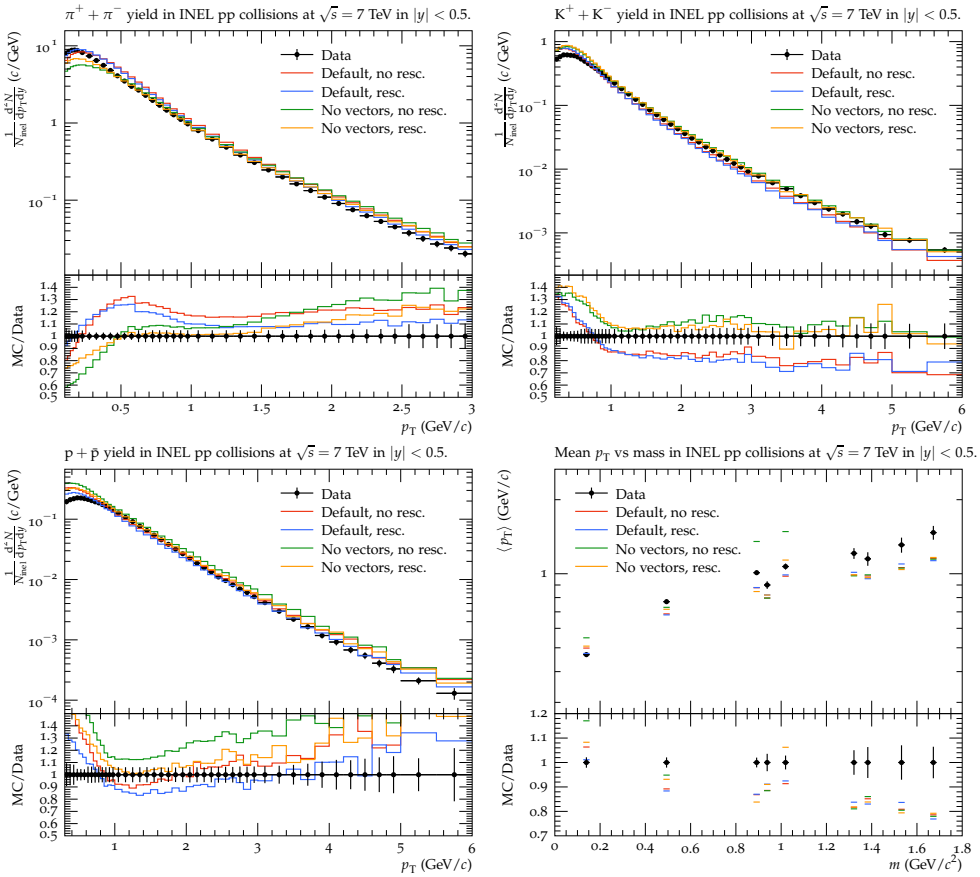


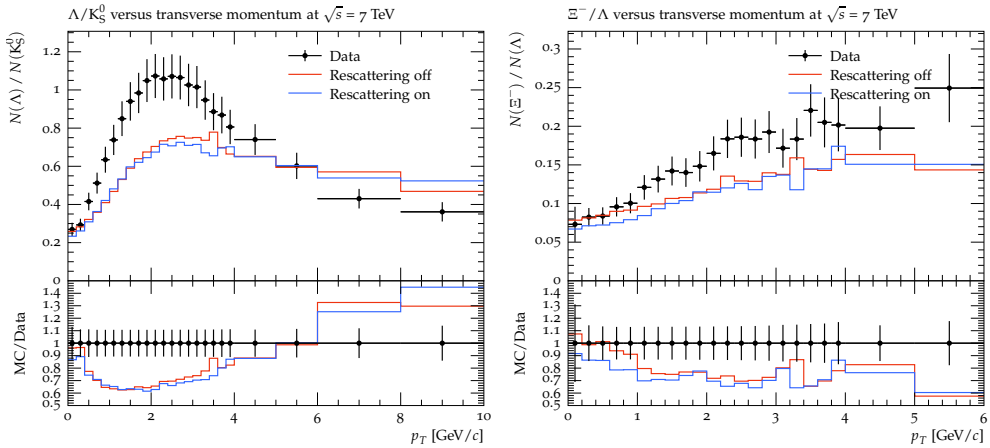
Figure I.23: $p_\perp$ spectra for $\pi^\pm$, $K^\pm$ and $p/\bar{p}$ and average $p_\perp$ for various particles, comparing rescattering to no rescattering, when no vector mesons are produced in the primary hadronization.

6 Summary and Outlook

Hadronic rescattering is inevitable in the dense hadronic systems produced in high-energy pp collisions. What less understood is the rate at which it happens, and the detailed modelling of the processes involved is open to discussion.

In this article we have developed and studied a framework for hadronic rescattering in pp collisions. This involves three main aspects:

1. The space–time tracing of the motion of hadrons, with interleaved scatterings and decays. The starting point here is our picture for the space–time production of hadrons. Thereafter the motion of these hadrons is traced and possible crossings iden-



(a) (b)

Figure I.24: $p_\perp$ ratios of (a) Λ^0 to K_S and (b) Ξ^- to Λ^0.

tified. The technical challenge is the fast growth of the number of hadron pairs to check, which can make have a significant impact on computing speed, even though most of these pairs never interact.

2. The cross section for different collision processes. This is where most of the development effort has gone, and most of the new code can be found. Much of the input has been from external sources, such as UrQMD ansätze, the calculations by Peláez et al., the HPR_1R_2, CERN/HERA and SaS parameterizations, and experimental data. We have tried to combine and extend these parts sensibly. For hadron pairs not described in any other way, the Additive Quark Model is invoked to provide order-of-magnitude cross sections, also for charm and bottom hadrons.

3. The production of the new hadrons in these collisions. This is done either through explicit few-body channels, like elastic scattering or resonance formation, or through the existing string fragmentation machinery. The typical collisions energies are so small, however, that extra efforts have to be made to translate these tiny strings into acceptable final states.

Each of the three components are open to further refinements, but the new framework presented here should offer a good starting point for various studies as is. Other frameworks overlapping with ours already exist. To the extent feasible, one obvious future task would be to compare with other rescattering implementations, starting from the same initial hadron configuration.

Nevertheless, what we bring now is a cohesive implementation, where the full power of the traditional Pythia energy–momentum description is extended by the recent match-

ing space–time picture and the new rescattering components, without the need to bridge disparate codes. This framework can then be applied to pp collisions of any kind, from minimum-bias to high-$p_\perp$ physics. As far as we know, no other single program can offer as much.

The main emphasis in this study has been to develop and test the framework, and to explore and understand how it behaves in general terms. Some applications to LHC pp studies have also been presented. In particular we note that rescattering contributes to some aspects of collective flow, notably a "pion wind" that slows down pions and speeds up protons and (most) other baryons. This helps remedy one of the glaring discrepancies of the traditional Pythia setup in comparisons with data. Unfortunately, the effects are not large enough to fully resolve the discrepancies. Worse, the Kaon $p_\perp$ spectrum is not modified appreciably, owing to a balance between speedup from the pion wind and slowdown from $2 \to n, n \geq 3$ processes. For this reason, one interesting topic for future study is the modelling of $3 \to 2$ and related processes. There are also other phenomena, like azimuthal flow, where rescattering appears to give only a very small contribution.

Thus it is obvious that further mechanisms will be needed to reach agreement with a number of observables. We have here briefly explored some potential options, such as a randomly fluctuating string tension, i.e. the "thermal" model, and a larger string tension in a dense-string environment. Other ideas remain to be mixed in, such as string shoving. It may be disappointing not to be in a situation where one simple model describes it all, but the reality is that any physical process that can happen will also do so, at some level.

The framework and its individual components have a higher applicability than the one presented in this article, and we envisage several follow-up studies. The most obvious one is to step up from pp to pA and AA. This should be straightforward, since Pythia already contains the Angantyr framework for heavy-ion collisions [75]. In a first step, we would study the effects of rescattering on its own, without any other mechanisms for collective flow. In a second step, one could combine it with other effects, such as shove and rope formation, which also contribute to flow effects.

One relevant AA study has already been done [32], based on Pythia/Angantyr and its space–time picture, but interfacing UrQMD to handle the rescattering. Physics comparisons between the two approaches will be useful on its own, but additionally we hope that we can offer a more user-friendly framework, thereby simplifying the future experimental study of rescattering effects.

Although this article has mainly focused on rescattering, it should not be overlooked that the underlying framework, which allows for collisions for different beam particles and collision energies from the mass threshold and upwards, has other potential use cases. It could for example come in handy for other applications, such as the simulation of cosmic ray showers in the atmosphere and of hadronic showers in detectors. Currently this flexibil-

ity only works for soft collisions, however. In order to fully include perturbative QCD aspects, such as jets and MPIs, it is necessary to specify meaningful PDFs for all colliding hadron species. Relevant combinations then have to be stored such that it is easy to switch between them. A special aspect is that, whereas collider physics mainly addresses particle production at central rapidities, the evolution of hadronic showers is especially sensitive to the production of the most forward hadrons, which therefore has to be carefully modelled.

In the current article, there has been no effort at a detailed retuning of all model parameters, but only a modest revision of $p_{\perp 0}$ to retain the same total charged multiplicity as before when rescattering is switched on. A future exercise would be to do a full-fledged retuning. This could start with $\mathrm{e}^+\mathrm{e}^-$ annihilation events at LEP, where no big effects are expected. Even small ones would be of interest, however, since they could also add one more source of uncertainty in W mass determinations [76], in addition to colour reconnection [77] and Bose-Einstein [78].

In conclusion, we hope that the current article and the new Pythia capabilities will be interesting for the experimental community, and also open up for further developments and studies. By experience we know that new generator capabilities tend to inspire both expected and unexpected applications.

Acknowledgements

Thanks to J.R. Peláez, A. Rodas and J. Ruiz de Elvira for providing us with Mathematica code for their $\pi\pi$ and πK cross sections. Work supported in part by the Swedish Research Council, contract number 2016-05996, and in part by the MCnetITN3 H2020 Marie Curie Innovative Training Network, grant agreement 722104. This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No 668679.

References

[1] **CMS** Collaboration, V. Khachatryan *et al.*, "Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC," *JHEP* **09** (2010) 091, `arXiv:1009.4122 [hep-ex]`.

[2] **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** no. 17, (2016) 172301, `arXiv:1509.04776 [hep-ex]`.

[3] **CMS** Collaboration, V. Khachatryan *et al.*, "Evidence for collectivity in pp collisions at the LHC," *Phys. Lett.* **B765** (2017) 193–220, arXiv:1606.06198 [nucl-ex].

[4] **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.* **C75** no. 5, (2015) 226, arXiv:1504.00024 [nucl-ex].

[5] **ALICE** Collaboration, J. Adam *et al.*, "Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions," *Nature Phys.* **13** (2017) 535–539, arXiv:1606.07424 [nucl-ex].

[6] **ALICE** 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** no. 14, (2019) 142301, arXiv:1903.01790 [nucl-ex].

[7] **CMS** Collaboration, V. Khachatryan *et al.*, "Strange Particle Production in *pp* Collisions at $\sqrt{s} = 0.9$ and 7 TeV," *JHEP* **05** (2011) 064, arXiv:1102.4282 [hep-ex].

[8] **ALICE** Collaboration, B. B. Abelev *et al.*, "Production of $\Sigma(1385)^{\pm}$ and $\Xi(1530)^{0}$ in proton-proton collisions at $\sqrt{s} = 7$ TeV," *Eur. Phys. J.* **C75** no. 1, (2015) 1, arXiv:1406.3206 [nucl-ex].

[9] J. Adolfsson *et al.*, "QCD Challenges from pp to A-A Collisions," 2020. arXiv:2003.10997 [hep-ph].

[10] P. Braun-Munzinger and J. Stachel, "The quest for the quark-gluon plasma," *Nature* **448** (2007) 302–309.

[11] W. Busza, K. Rajagopal, and W. van der Schee, "Heavy Ion Collisions: The Big Picture, and the Big Questions," *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 339–376, arXiv:1802.04801 [hep-ph].

[12] J. L. Nagle and W. A. Zajc, "Small System Collectivity in Relativistic Hadronic and Nuclear Collisions," *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 211–235, arXiv:1801.03477 [nucl-ex].

[13] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider," *Phys. Rev.* **C92** no. 3, (2015) 034906, arXiv:1306.0121 [hep-ph].

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

[15] C. Bierlich, G. Gustafson, and L. Lönnblad, "A shoving model for collectivity in hadronic collisions," arXiv:1612.05132 [hep-ph].

[16] N. Fischer and T. Sjöstrand, "Thermodynamical String Fragmentation," *JHEP* **01** (2017) 140, arXiv:1610.09818 [hep-ph].

[17] T. Sjöstrand and M. van Zijl, "A Multiple Interaction Model for the Event Structure in Hadron Collisions," *Phys. Rev.* **D36** (1987) 2019.

[18] T. Sjöstrand, "The Development of MPI Modeling in Pythia," *Adv. Ser. Direct. High Energy Phys.* **29** (2018) 191–225, arXiv:1706.02166 [hep-ph].

[19] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton Fragmentation and String Dynamics," *Phys. Rept.* **97** (1983) 31–145.

[20] S. Ferreres-Solé and T. Sjöstrand, "The space-time structure of hadronization in the Lund model," *Eur. Phys. J.* **C78** no. 11, (2018) 983, arXiv:1808.04619 [hep-ph].

[21] M. Bähr *et al.*, "Herwig++ Physics and Manual," *Eur. Phys. J.* **C58** (2008) 639–707, arXiv:0803.0883 [hep-ph].

[22] J. Bellm *et al.*, "Herwig 7.0/Herwig++ 3.0 release note," *Eur. Phys. J.* **C76** no. 4, (2016) 196, arXiv:1512.01178 [hep-ph].

[23] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].

[24] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An Introduction to PYTHIA 8.2," *Comput. Phys. Commun.* **191** (2015) 159–177, arXiv:1410.3012 [hep-ph].

[25] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *JHEP* **02** (2009) 007, arXiv:0811.4622 [hep-ph].

[26] E. Bothmann *et al.*, "Event Generation with Sherpa 2.2," *SciPost Phys.* **7** no. 3, (2019) 034, arXiv:1905.09127 [hep-ph].

[27] Y.-X. Zhang *et al.*, "Comparison of heavy-ion transport simulations: Collision integral in a box," *Phys. Rev.* **C97** no. 3, (2018) 034625, arXiv:1711.05950 [nucl-th].

[28] S. A. Bass *et al.*, "Microscopic models for ultrarelativistic heavy ion collisions," *Prog. Part. Nucl. Phys.* **41** (1998) 255–369, arXiv:nucl-th/9803035 [nucl-th]. [Prog. Part. Nucl. Phys.41,225(1998)].

[29] J. Weil *et al.*, "Particle production and equilibrium properties within a new hadron transport approach for heavy-ion collisions," *Phys. Rev.* **C94** no. 5, (2016) 054905, arXiv:1606.06642 [nucl-th].

[30] B. H. Sa and A. Tai, "An Event generator for the firecracker model and the rescattering in high-energy p A and A-A collisions: LUCIAE version 2.0," *Comput. Phys. Commun.* **90** (1995) 121–140.

[31] B.-H. Sa, D.-M. Zhou, Y.-L. Yan, X.-M. Li, S.-Q. Feng, B.-G. Dong, and X. Cai, "PACIAE 2.0: An Updated parton and hadron cascade model (program) for the relativistic nuclear collisions," *Comput. Phys. Commun.* **183** (2012) 333–346, arXiv:1104.1238 [nucl-th].

[32] A. V. da Silva, C. Bierlich, D. Dobrigkeit Chinellato, and J. Takahashi, "Studying the effect of the hadronic phase in nuclear collisions with PYTHIA and UrQMD," in *18th International Conference on Strangeness in Quark Matter (SQM 2019) Bari, Italy, June 10-15, 2019*. 2019. arXiv:1911.12824 [nucl-th].

[33] G. 't Hooft, "A Planar Diagram Theory for Strong Interactions," *Nucl. Phys.* **B72** (1974) 461.

[34] T. Sjöstrand, "Jet Fragmentation of Nearby Partons," *Nucl. Phys.* **B248** (1984) 469–502.

[35] K. Gallmeister and T. Falter, "Space-time picture of fragmentation in PYTHIA/JETSET for HERMES and RHIC," *Phys. Lett.* **B630** (2005) 40–48, arXiv:nucl-th/0502015 [nucl-th].

[36] C. Bierlich and C. O. Rasmussen, "Dipole evolution: perspectives for collectivity and γ^*A collisions," *JHEP* **10** (2019) 026, arXiv:1907.12871 [hep-ph].

[37] A. H. Mueller, "Soft gluons in the infinite momentum wave function and the BFKL pomeron," *Nucl. Phys.* **B415** (1994) 373–385.

[38] A. H. Mueller and B. Patel, "Single and double BFKL pomeron exchange and a dipole picture of high-energy hard processes," *Nucl. Phys.* **B425** (1994) 471–488, arXiv:hep-ph/9403256.

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

[40] Z. Xu and C. Greiner, "Thermalization of gluons in ultrarelativistic heavy ion collisions by including three-body interactions in a parton cascade," *Phys. Rev.* **C71** (2005) 064901, arXiv:hep-ph/0406278 [hep-ph].

[41] G. Peter, D. Behrens, and C. C. Noack, "Poincare covariant particle dynamics. 1: Intranuclear cascade model," *Phys. Rev.* **C49** (1994) 3253–3265.

[42] D. Behrens, G. Peter, and C. C. Noack, "Poincare covariant particle dynamics. 2: Fragmentation for ultrarelativistic reactions," *Phys. Rev.* **C49** (1994) 3266–3274.

[43] C. B. Duncan and P. Skands, "Fragmentation of Two Repelling QCD Strings," `arXiv:1912.09639 [hep-ph]`.

[44] E. M. Levin and L. L. Frankfurt, "The Quark hypothesis and relations between cross-sections at high-energies," *JETP Lett.* **2** (1965) 65–70.

[45] H. J. Lipkin, "Quarks for pedestrians," *Phys. Rept.* **8** (1973) 173–268.

[46] **Particle Data Group** Collaboration, M. Tanabashi *et al.*, "Review of Particle Physics," *Phys. Rev.* **D98** no. 3, (2018) 030001.

[47] R. Garcia-Martin, R. Kaminski, J. R. Pelaez, J. Ruiz de Elvira, and F. J. Yndurain, "The Pion-pion scattering amplitude. IV: Improved analysis with once subtracted Roy-like equations up to 1100 MeV," *Phys. Rev.* **D83** (2011) 074004, `arXiv:1102.2183 [hep-ph]`.

[48] J. R. Pelaez, A. Rodas, and J. Ruiz De Elvira, "Global parameterization of $\pi\pi$ scattering up to 2 GeV," *Eur. Phys. J.* **C79** no. 12, (2019) 1008, `arXiv:1907.13162 [hep-ph]`.

[49] J. Pelaez and A. Rodas, "Pion-kaon scattering amplitude constrained with forward dispersion relations up to 1.6 GeV," *Phys. Rev. D* **93** no. 7, (2016) 074025, `arXiv:1602.08404 [hep-ph]`.

[50] V. Srinivasan *et al.*, "Pi- pi+ –> pi- pi+ Interactions Below 0.7-GeV from pi- p –> pi- pi+ n Data at 5-GeV/c," *Phys. Rev.* **D12** (1975) 681.

[51] S. D. Protopopescu, M. Alston-Garnjost, A. Barbaro-Galtieri, S. M. Flatte, J. H. Friedman, T. A. Lasinski, G. R. Lynch, M. S. Rabin, and F. T. Solmitz, "Pi pi Partial Wave Analysis from Reactions pi+ p —> pi+ pi- Delta++ and pi+ p —> K+ K- Delta++ at 7.1-GeV/c," *Phys. Rev.* **D7** (1973) 1279.

[52] N. N. Biswas, N. M. Cason, I. Derado, V. P. Kenney, J. A. Poirier, and W. D. Shephard, "Total Pion-Pion Cross Sections for the 2-GeV Di-Pion Mass Region," *Phys. Rev. Lett.* **18** no. 7, (1967) 273–276.

[53] W. J. Robertson, W. D. Walker, and J. L. Davis, "High-energy pi pi collisions," *Phys. Rev.* **D7** (1973) 2554–2564.

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

[55] A. Donnachie and P. V. Landshoff, "Dynamics of Elastic Scattering," *Nucl. Phys.* **B267** (1986) 690–701.

[56] G. A. Schuler and T. Sjöstrand, "Hadronic diffractive cross-sections and the rise of the total cross-section," *Phys. Rev.* **D49** (1994) 2257–2267.

[57] P. Koch and C. B. Dover, "$K^{\pm}$, $\bar{p}$ and Ω^{-} Production in Relativistic Heavy Ion Collisions," *Phys. Rev.* **C40** (1989) 145–155.

[58] G. A. Schuler and T. Sjöstrand, "A Scenario for high-energy gamma gamma interactions," *Z. Phys.* **C73** (1997) 677–688, arXiv:hep-ph/9605240 [hep-ph].

[59] C. O. Rasmussen and T. Sjöstrand, "Models for total, elastic and diffractive cross sections," *Eur. Phys. J.* **C78** no. 6, (2018) 461, arXiv:1804.10373 [hep-ph].

[60] S. Navin, "Diffraction in Pythia," arXiv:1005.3894 [hep-ph].

[61] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune," *Eur. Phys. J.* **C74** no. 8, (2014) 3024, arXiv:1404.5630 [hep-ph].

[62] T. Sjöstrand and P. Z. Skands, "Baryon number violation and string topologies," *Nucl. Phys.* **B659** (2003) 243, arXiv:hep-ph/0212264 [hep-ph].

[63] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP* **04** (2008) 063, arXiv:0802.1189 [hep-ph].

[64] **ATLAS** Collaboration, G. Aad *et al.*, "Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC," *New J. Phys.* **13** (2011) 053033, arXiv:1012.5104 [hep-ex].

[65] C. Bierlich *et al.*, "Robust Independent Validation of Experiment and Theory: Rivet version 3," *SciPost Phys.* **8** (2020) 026, arXiv:1912.05451 [hep-ph].

[66] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Measurement of charged particle spectra in minimum-bias events from proton–proton collisions at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C* **78** no. 9, (2018) 697, arXiv:1806.11245 [hep-ex].

[67] **ATLAS** Collaboration, G. Aad *et al.*, "Charged-particle distributions in $\sqrt{s}$ = 13 TeV pp interactions measured with the ATLAS detector at the LHC," *Phys. Lett. B* **758** (2016) 67–88, arXiv:1602.01633 [hep-ex].

[68] **ATLAS** Collaboration, M. Aaboud *et al.*, "Charged-particle distributions at low transverse momentum in $\sqrt{s} = 13$ TeV pp interactions measured with the ATLAS detector at the LHC," *Eur. Phys. J. C* **76** no. 9, (2016) 502, arXiv:1606.01133 [hep-ex].

[69] R. Hagedorn, "Statistical thermodynamics of strong interactions at high-energies," *Nuovo Cim. Suppl.* **3** (1965) 147–186.

[70] R. Hagedorn, "Remarks on the thermodynamical model of strong interactions," *Nucl. Phys.* **B24** (1970) 93–139.

[71] R. Hagedorn, "Multiplicities, p_T Distributions and the Expected Hadron $\rightarrow$ Quark - Gluon Phase Transition," *Riv. Nuovo Cim.* **6N10** (1983) 1–50.

[72] A. Bialas, "Fluctuations of string tension and transverse mass distribution," *Phys. Lett. B* **466** (1999) 301–304, arXiv:hep-ph/9909417.

[73] A. Casher, H. Neuberger, and S. Nussinov, "Chromoelectric Flux Tube Model of Particle Production," *Phys. Rev. D* **20** (1979) 179–188.

[74] P. Skands, "Introduction to QCD," in *Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales*, pp. 63–124. 2017. arXiv:1207.2389 [hep-ph].

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

[76] **ALEPH, DELPHI, L3, OPAL, LEP Electroweak** Collaboration, S. Schael *et al.*, "Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP," *Phys.Rept.* **532** (2013) 119–244, arXiv:1302.3415 [hep-ex].

[77] T. Sjöstrand and V. A. Khoze, "On Color rearrangement in hadronic W^+W^- events," *Z. Phys.* **C62** (1994) 281–310, arXiv:hep-ph/9310242.

[78] L. Lönnblad and T. Sjöstrand, "Modeling Bose-Einstein correlations at LEP-2," *Eur. Phys. J.* **C2** (1998) 165–180, arXiv:hep-ph/9711460 [hep-ph].

Paper II

II

Hadronic Rescattering in pA and AA Collisions

Christian Bierlich, Torbjörn Sjöstrand and Marius Utheim

Eur. Phys. J. A **57**, 227 (2021)
e-Print: `arXiv:2103.09665` [hep-ph]
MCnet–21–03, LU TP 21–08

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 $\approx 0.1 - 1$ 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. The physics involved after this freeze-out is the main topic of this paper, though with the large difference to traditional approaches, that it happens much sooner.

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 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 $\mathrm{d}n_{\mathrm{charged}}/\mathrm{d}\eta$. 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. 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 the UrQMD [27] or SMASH [28] programs. What makes Angantyr/Pythia different is that there is no QGP phase, so that rescattering can start earlier, and therefore hypothetically can give larger effects. 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 [29]. 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 *eg.* with UrQMD, which was done by Ref. [30].

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 [31]. 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, 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 [31], 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. In this article the rescattering studies are extended to pA and AA collisions, 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, *eg.* a J/ψ depletion in central collisions. The most interesting result is a clear signal of elliptic flow induced by rescatterings, that even matches experimental PbPb numbers at large multiplicities, to be contrasted with the miniscule effects in pp.

The outline of the article is as follows. In Section 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 Section 3 effects in this model are tested on its own, while Section 4 shows comparisons with data. Some conclusions and an outlook are presented in Section 5. Finally, technical aspects related to computation time for rescattering are discussed in the appendix.

Natural units are assumed throughout the article, *ie.* $c = \hbar = 1$. 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 [32, 33], 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 $\pm b_{\mathrm{AA}}/2$, *eg.* 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 *eg.* 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 ($T(\boldsymbol{b})$) is used. It allows for the calculation of the amplitude $T_{kl}(\boldsymbol{b})$ 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 [34] being the default choice. The parametrization of $T(\boldsymbol{b})$ 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 $AB_j \to AX_j$. 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 $\mathrm{d}M^2_{X_j}/M^2_{X_j}$, 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 $\mathrm{d}x_j/x_j$, 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 $\mathbb{P}\mathbb{P}$ 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.

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 [35, 36]. 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, *eg.* 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 *eg.* [37]. A preliminary study in [31] 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 [38]. 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 [29]. 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 $N_C \to \infty$ 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 [39]. In it, a linear confinement is assumed, *ie.* 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 $\mathrm{q}\bar{\mathrm{q}}$ pair, the linearity leads to a straightforward relationship between the energy–momentum and the space–time pictures:

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

If there is enough energy, the string between an original $\mathrm{q}_0\bar{\mathrm{q}}_0$ pair may break by producing new $\mathrm{q}_i\bar{\mathrm{q}}_i$ pairs, where the intermediate q_i ($\bar{\mathrm{q}}_i$) are pulled towards the $\bar{\mathrm{q}}_0$ (q_0) end, such that the original colour field is screened. This way the system breaks up into a set of n colour singlets $\mathrm{q}_0\bar{\mathrm{q}}_1 - \mathrm{q}_1\bar{\mathrm{q}}_2 - \mathrm{q}_2\bar{\mathrm{q}}_3 - \ldots - \mathrm{q}_{n-1}\bar{\mathrm{q}}_0$, that we can associate with the primary hadrons. By (II.1) the location of the breakup vertices in space–time is linearly related to the energy–momentum of the hadrons produced between such vertices [29].

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 *eg.* 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 $N_C \to \infty$ approximation [40], 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 $\mathrm{q}\bar{\mathrm{q}}$ 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. [41], and was then extended to reconstruct matching space–time production vertices in [29]. 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 $\mathrm{q}\bar{\mathrm{q}}$ 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 $\bar{\mathrm{q}}$ 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 $\mathrm{q}\bar{\mathrm{q}}$ string, where the relevant length of each hadron string piece is related to its energy. For a given hadron, define E_{hq} ($E_{\mathrm{h}\bar{\mathrm{q}}}$) as half the energy of the hadron plus the full energy of all hadrons lying between it and the q ($\bar{\mathrm{q}}$) end, and use this as a measure of how closely associated a hadron is with the respective endpoint. Also let $\mathbf{r}_{\perp\mathrm{q}}$ ($\mathbf{r}_{\perp\bar{\mathrm{q}}}$) be the (anti)quark transverse

production coordinates. Then define the hadron production vertex offset to be

$$\Delta\mathbf{r}_{\perp \mathrm{h}} = \frac{E_{\mathrm{h}\bar{\mathrm{q}}}\,\mathbf{r}_{\perp \mathrm{q}} + E_{\mathrm{hq}}\,\mathbf{r}_{\perp \bar{\mathrm{q}}}}{E_{\mathrm{hq}} + E_{\mathrm{h}\bar{\mathrm{q}}}} = \frac{(E_{\mathrm{tot}} - E_{\mathrm{hq}})\,\mathbf{r}_{\perp \mathrm{q}} + E_{\mathrm{hq}}\,\mathbf{r}_{\perp \bar{\mathrm{q}}}}{E_{\mathrm{tot}}} \,, \tag{II.2}$$

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 $\int P(b)\,\mathrm{d}^2b = \sigma_{AB}(E_{\mathrm{CM}})$. 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. 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 an effective delay before the new hadrons are formed and can begin to interact. For the other processes, such as elastic scattering or an intermediate resonance decay, there is the option to have effective formation times before new interactions are allowed.

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 $\tau_0 = 1/\Gamma$ is the inverse of the 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, 42], 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, *ie.* 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. [31], 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, *ie.* without MPIs. We will see in Section 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, *ie.* $AB \to AB$. In our implementation, these are considered different from elastic scattering through a resonance, *eg.* $\pi^+\pi^- \to \rho^0 \to \pi^+\pi^-$, although the two could be linked by interference terms. In experiments, usually all $AB \to AB$ events are called elastic because it is not possible to tell which underlying mechanism is involved.

- Resonance formation typically can be written as $AB \to R \to CD$, 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 $AB \to XB$ or $AB \to AX$ and double $AB \to X_1X_2$. 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 $AB \to A^*B$, $AB \to AB^*$ or $AB \to A^*B^*$. 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 $K\pi$ cross sections are found using the calculations of Peláez et al. [43–45], which partly are based on Chiral Perturbation Theory.

- The neutral Kaon system is nontrivial, with strong interactions described by the $K^0/\bar{K}^0$ 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 $\bar{K}^0$ with that hadron. When a collision occurs, the $K_{S,L}$ is converted into either K^0 or $\bar{K}^0$, where the probability 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 [46], consisting of one fixed term, one "pomeron" $\ln^2 s$ ($s = E^2_{CM}$) and two "reggeon" $s^{-\eta}$ ones.

- NN and $N\pi$ elastic cross sections are partly covered by the CERN/HERA data parameterizations [47].

- 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 [48].
- Differential diffractive cross sections are described by the SaS (Schuler and Sjöstrand) ansatz [34, 49], 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) [50, 51] 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

$$n_{\mathrm{q,AQM}} = n_{\mathrm{u}} + n_{\mathrm{d}} + 0.6\, n_{\mathrm{s}} + 0.2\, n_{\mathrm{c}} + 0.07\, n_{\mathrm{b}} \; . \tag{II.3}$$

  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 $\bar{\mathrm{q}}\bar{\mathrm{q}}$, and an anticolour ditto, $\bar{\mathrm{q}}$ 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 $m_A^* + m_B^* < E_{\mathrm{CM}}$ 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 $\mathrm{u}\bar{\mathrm{u}}$ or $\mathrm{d}\bar{\mathrm{d}}$ 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

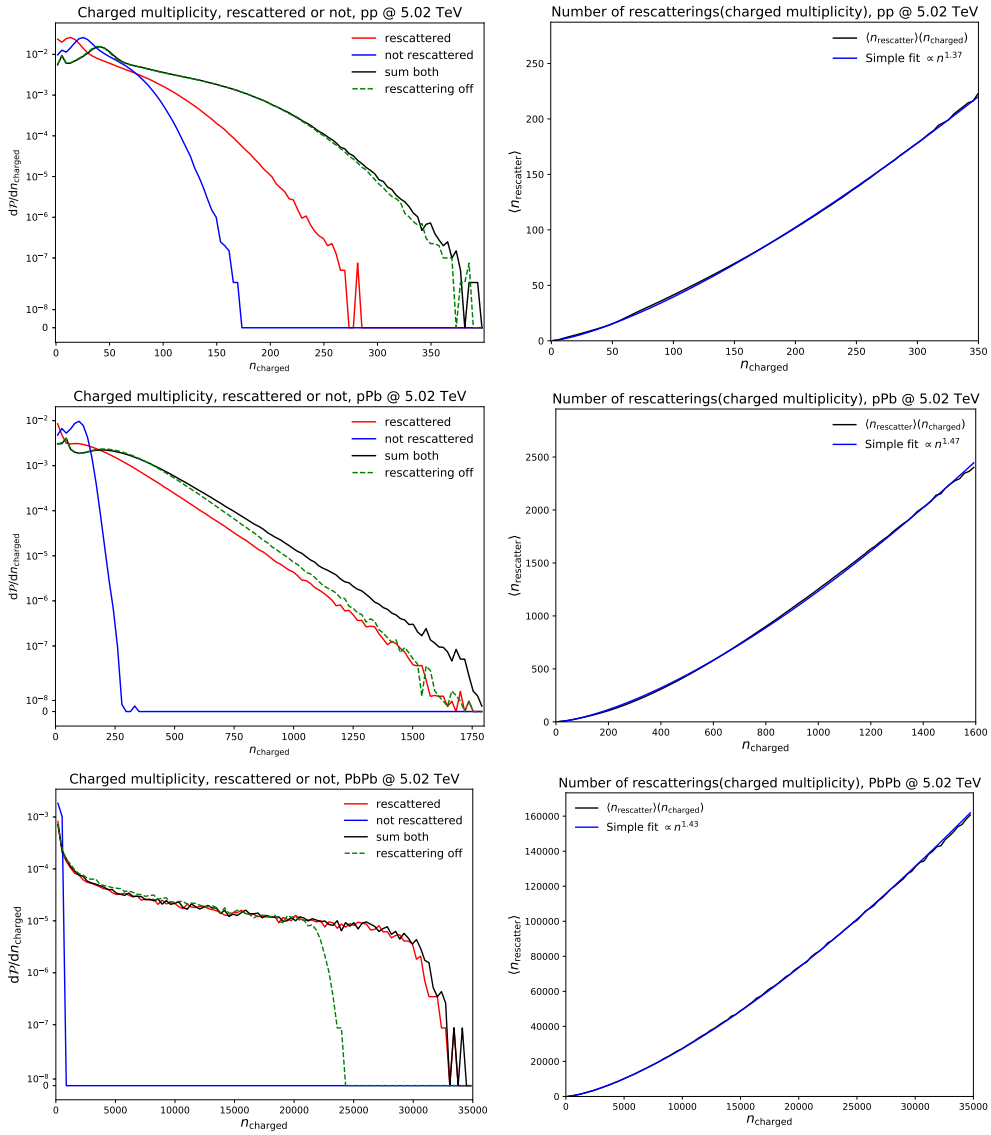
In this section we will study the rescattering model in pp, pPb and PbPb collisions. All collision energies are set to 5.02 TeV per nucleon-nucleon system. This includes pp, for comparison reasons; results at the more standard 13 TeV pp energy have already been presented elsewhere [31].

3.1 Multiplicities

The current lack of $3 \to 2$ processes in our model, to partly balance the $2 \to 3$ 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. 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 should rather wait until $3 \to 2$ has been included. For now we accept some mismatch.

Charged multiplicity distributions are shown in Figure II.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, 52] thus has to be explained by other mechanisms, such as the rope model [26] or other approaches that give an increased string tension [53].



(a) (b)

Figure II.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}.

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 Figure II.1b. The number of potential interactions at the beginning of rescattering is proportional to n^2_{primary}, where the number of primary hadrons $n_{\mathrm{primary}} \simeq n_{\mathrm{ch}}$. 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^p_{ch} for some power p. As seen in Figure II.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 Figure II.2a by incoming particle kind and in Figure II.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 Figure II.3a,b, and the ratios with/without are shown in Figure II.3c,d. The effects 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 [52]. 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.

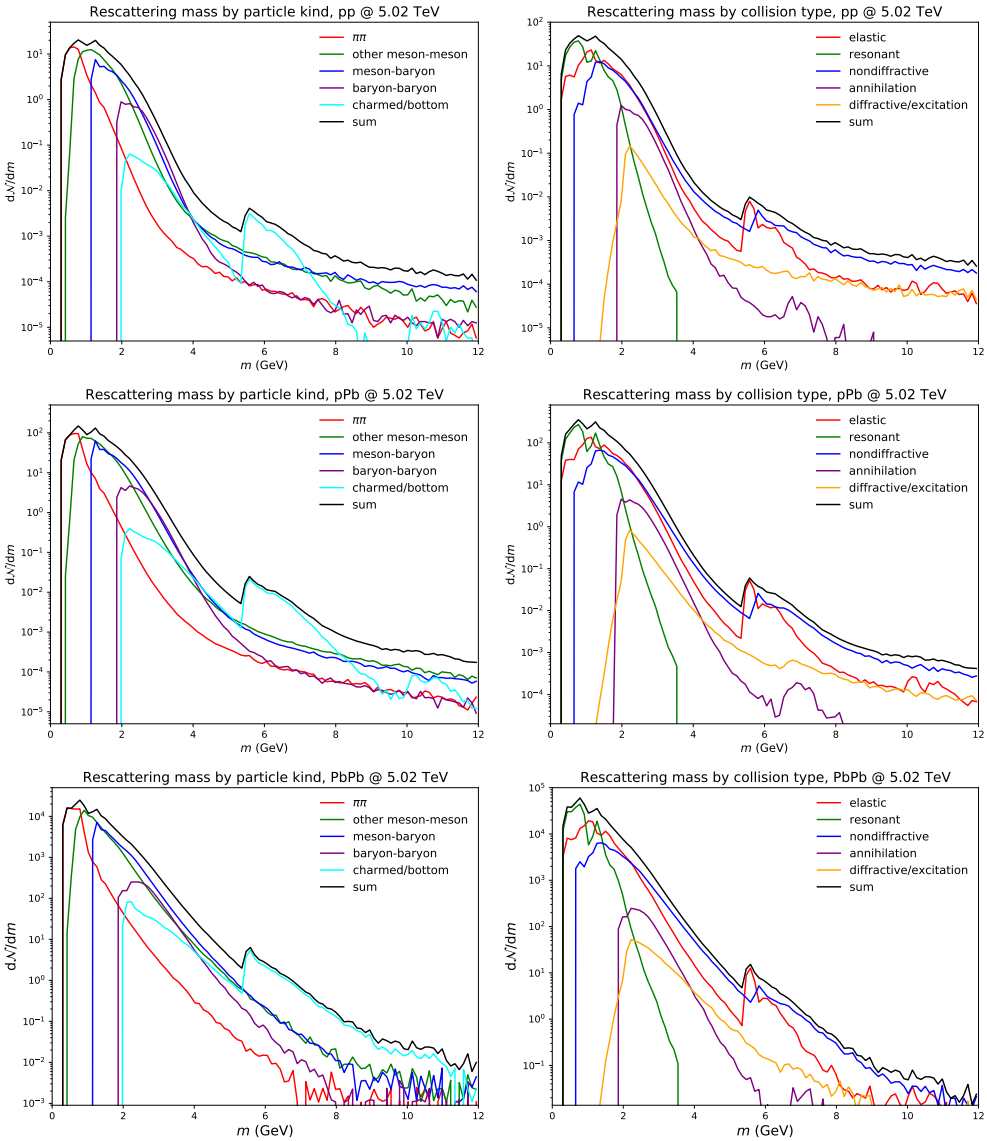


Figure II.2: Invariant masses for rescatterings, (a) by particle kind and (b) by rescattering process type.

We believe this effect can be ameliorated by implementing $3 \to 2$ 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.

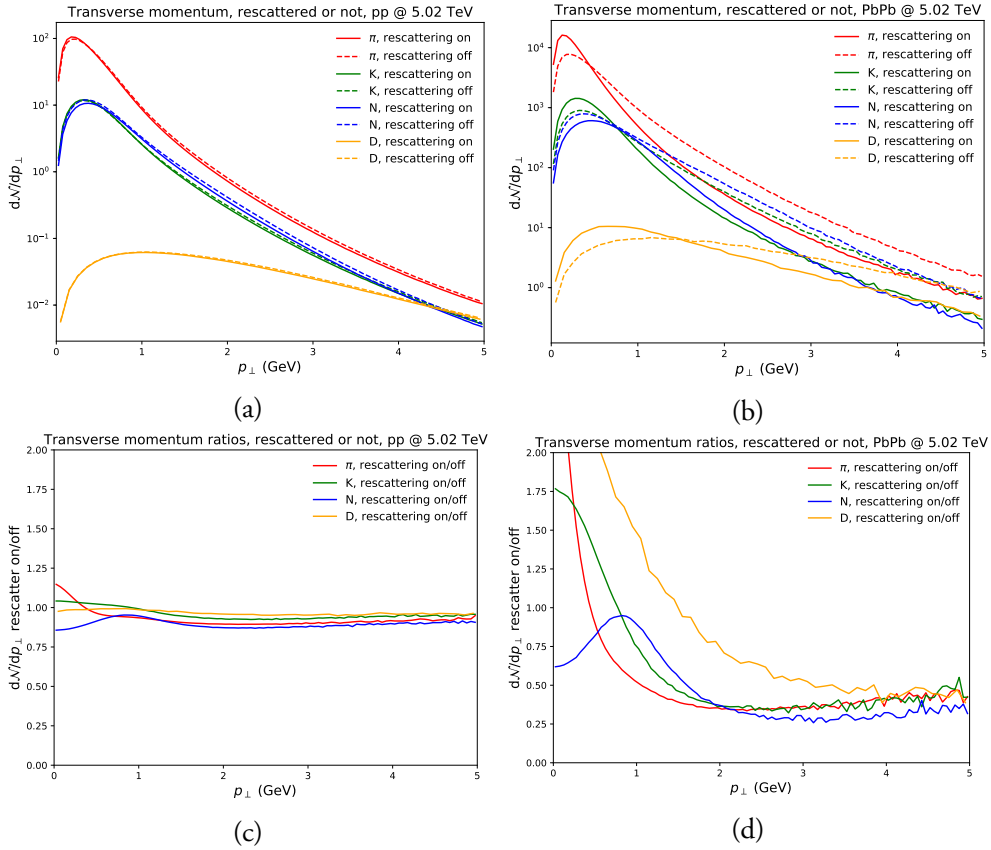


Figure II.3: $p_\perp$ spectra for pions, kaons, nucleons and D mesons, for (a) pp and (b) PbPb, together with ratios between the spectra with to without rescattering, for (c) pp and (d) PbPb.

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. Especially 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 Figure II.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 [54, 55]. 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 Figure II.4.

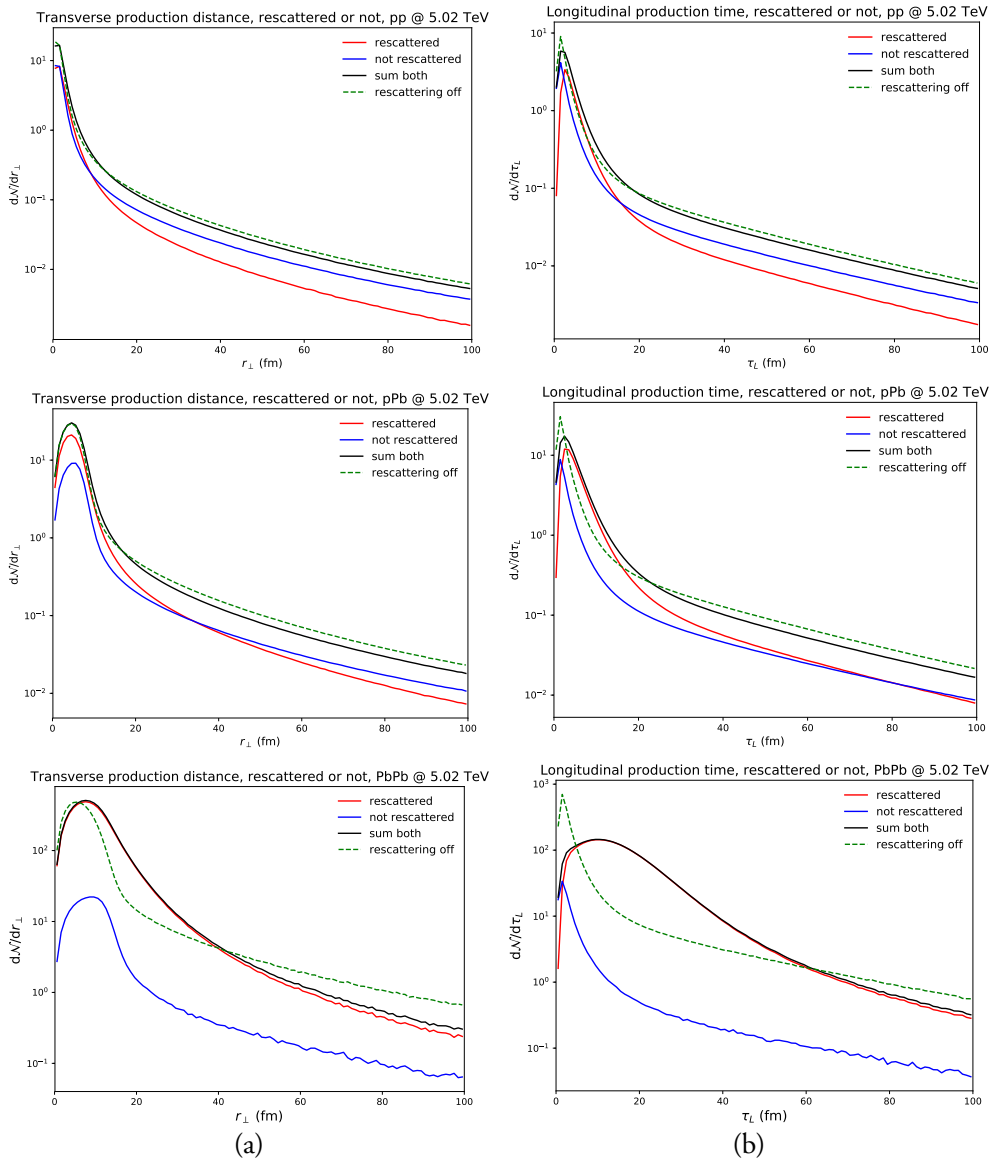


Figure II.4: (a) $r_\perp$ and (b) τ_L spectra. Note that rescattered also refers to hadrons produced in decays of rescattered particles, even though they themselves were not directly involved in rescattering.

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* mul-

tiplicity 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 $\mathrm{K}\pi \to \mathrm{K}^* \to \mathrm{K}\pi$. That is, some of the combinatorial background gets to be reclassified as K^*, without any change of the overall $\mathrm{K}\pi$ mass spectrum. To find the more subtle effects of nontrivial processes requires a detailed fitting of the $\mathrm{K}\pi$ 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

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

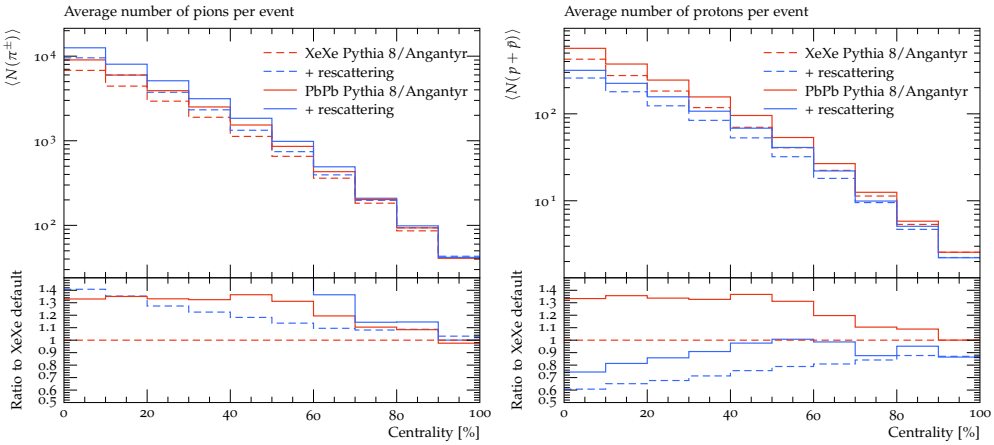
We note, however, that the results presented for nucleus-nucleus collisions can be transferred directly to experimental centrality measures, as the Angantyr model provides a good description of *eg.* forward energy, which correlates directly with the theoretical impact parameter.

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

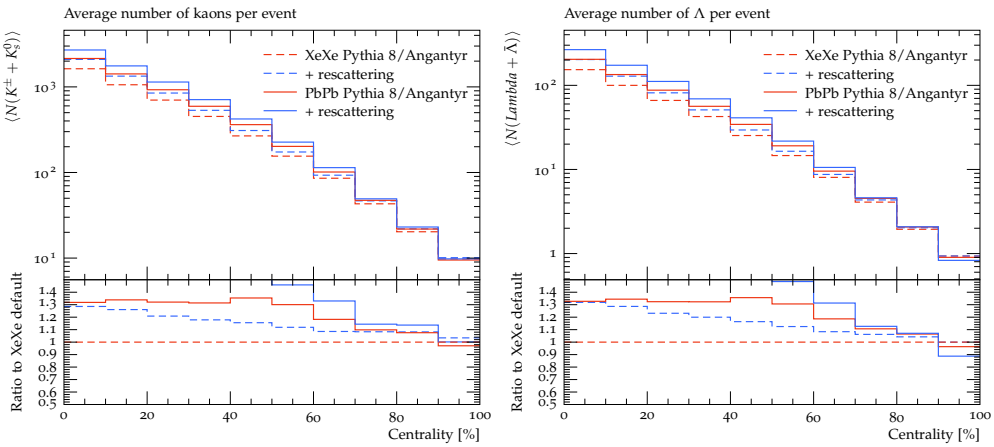
While the effect for pions is negligible in peripheral collisions, it grows to about 40% in central collisions. The effect on protons is also largest in central collisions, while in peripheral collisions it is still at a 10% level. This is particularly interesting in the context of recent years' introduction of microscopic models to explain the increase of strange baryon yields with increasing multiplicity, which overestimate the amount of protons [56].

In Figure II.6 we move to strange mesons and baryons, with the total kaon ($\mathrm{K}^{\pm}$ and $\mathrm{K}^0_{\mathrm{L,S}}$) and Λ multiplicity, (a) and (b) respectively. While there is a large effect on the direct yields



(a) (b)

Figure II.5: Average per-event yields of (a) pions ($\pi^{\pm}$) and (b) protons ($\mathrm{p}, \bar{\mathrm{p}}$) in PbPb and XeXe collisions at $\sqrt{s_{\mathrm{NN}}} = 2.74$ and 5.44 TeV respectively, as function of collision centrality.

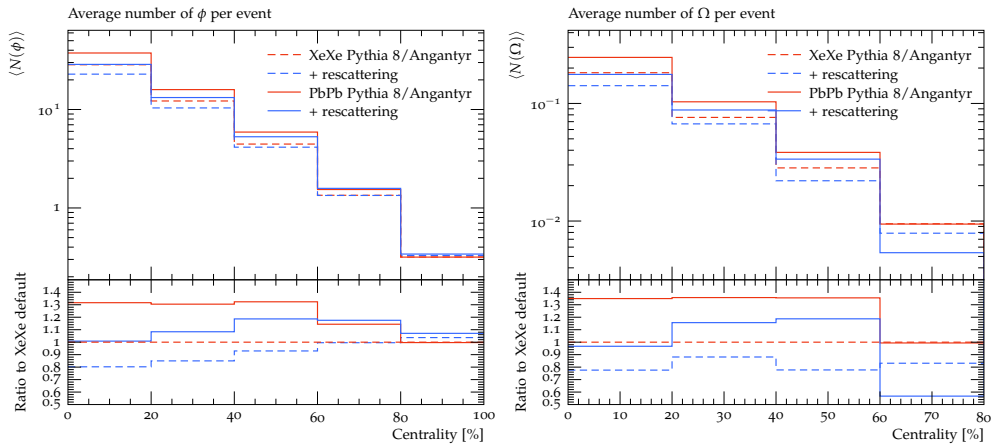


(a) (b)

Figure II.6: Average per-event yields of (a) kaons ($\mathrm{K}^{\pm}, \mathrm{K}^{0}_{\mathrm{L,S}}$) and (b) Λ ($\Lambda, \overline{\Lambda}$) in PbPb and XeXe collisions at $\sqrt{s_{\mathrm{NN}}} = 2.74$ and 5.44 TeV respectively, as function of collision centrality.

of both species, it is almost identical to the change in $\pi^{\pm}$ in Figure II.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 Figure II.7a and b respectively. The ϕ multiplicity decreases by about 20% in



(a) (b)

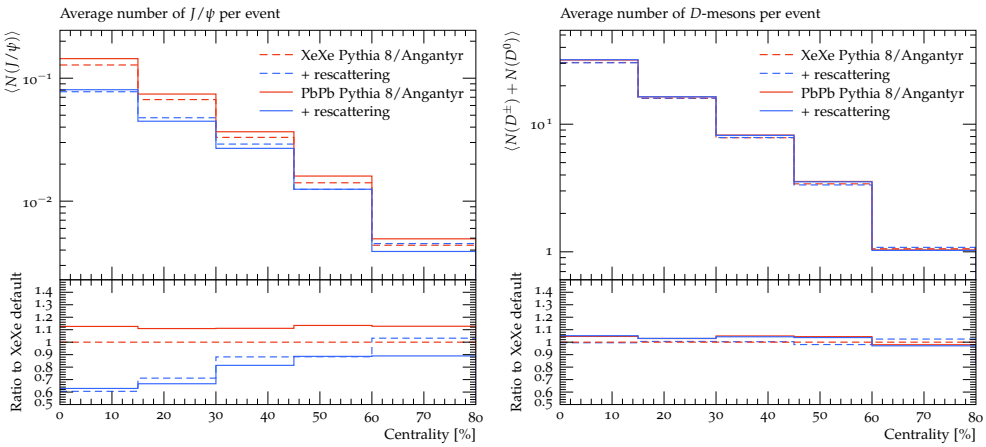
Figure II.7: Average per-event yields of (a) ϕ and (b) Ω^- in PbPb and XeXe collisions at $\sqrt{s_{\mathrm{NN}}} = 2.74$ and 5.44 TeV respectively, as function of collision centrality.

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.

Notably (and as opposed to *eg.* UrQMD), the rescattering framework implemented in Pythia, includes cross sections for heavy flavour mesons and baryons. In Figure II.8 we show the effect on (a) J/ψ and (b) D mesons ($\mathrm{D}^{\pm}$ 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 [57] 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 Figure II.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.



(a) (b)

Figure II.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.

Elliptic flow

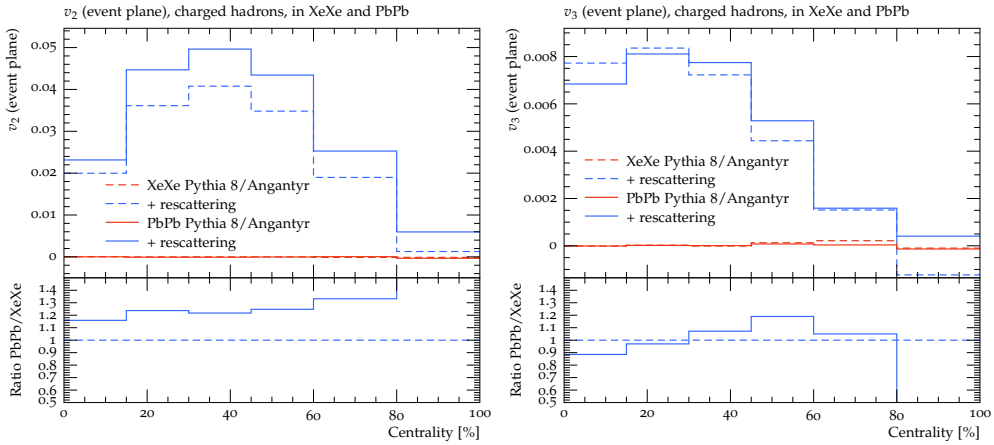
One of the most common ways to characterize heavy ion collisions is by the measurement of flow coefficients (v_n's), defined as the coefficients of a Fourier expansion of the single particle azimuthal yield, with respect to the event plane Ψ_n [58, 59]:

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

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 Section 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

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

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



(a) (b)

Figure II.9: Flow coefficients (a) v_2 and (b) v_3 in PbPb and XeXe collisions at $\sqrt{s_{\mathrm{NN}}} = 2.74$ 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.

coefficients can then simply be calculated as

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

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

In Figure II.9 we show (a) v_2 and (b) v_3 as functions of collision centrality for charged particles for XeXe collisions at $\sqrt{s_{\mathrm{NN}}} = 5.44$ TeV and PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.74$ 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 (II.7) ensures that no non-flow contributions enter the results.

For v_3 (Figure II.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 $\mathrm{pp}(\bar{\mathrm{p}}\mathrm{p})$ 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 $v_2(\mathrm{p}) > v_2(\pi) > v_2(\mathrm{K}) \approx v_2(\Lambda) > v_2(\Omega) > v_2(\phi)$, with the latter reaching its maximum for v_2 about an order of magnitude less than for protons.

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 $\mathrm{c\bar{c}}$ 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 Figure II.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 [60]. A clear difference is 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 Figure II.10b, v_2 for PbPb and XeXe are compatible within the statistical error. More importantly, the result is also compatible with experimental data [61]. Together with the result from Figure II.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 $\mathrm{c\bar{c}}$ states. A foreseen improvement of this treatment would be to consider differences with input taken *eg.* from lattice calculations.

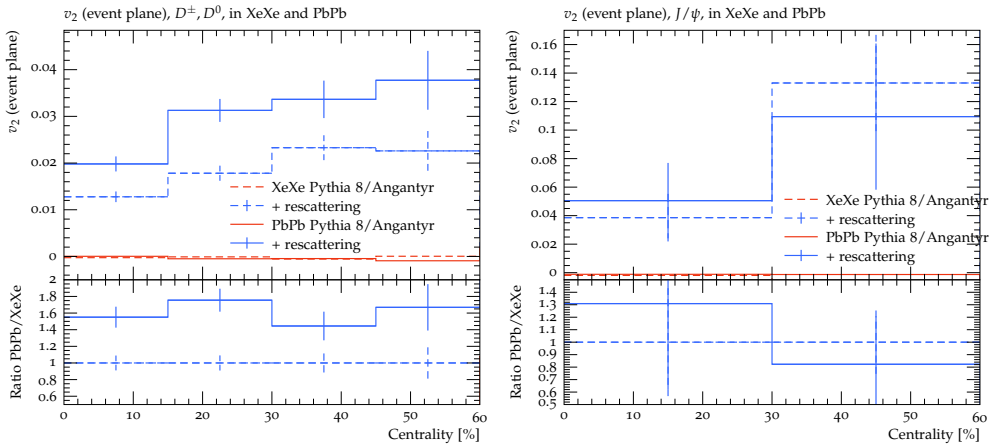
4 Comparison with data

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 [62] 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 (II.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 section 3.1 we described how the current lack of $3 \to 2$ processes in the rescattering framework increases the total multiplicities. In Figure II.11 Angantyr with and without



(a) (b)

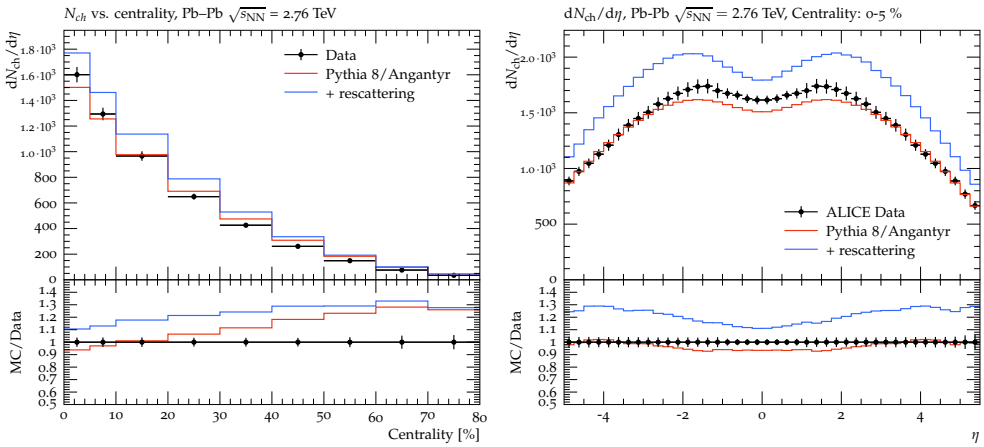
Figure II.10: The v_2 flow coefficient for (a) D mesons ($D^{\pm}, D^0$) and (b) J/ψ as a function of centrality in PbPb and XeXe collisions at $\sqrt{s_{NN}} = 2.74$ and 5.44 TeV respectively. Error bars are statistical errors. Results shown with and without rescattering, and are calculated with respect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

rescattering is compared to experimental data [63, 64].

In Figure II.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 Figure II.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 $|\eta| < 0.8$ [65] in Figure II.12, with 0-20% centrality shown in Figure II.12a, and 40-60% in Figure II.12b. It is seen that particles at intermediate $p_\perp \approx 1 - 6$ GeV are pushed down to very low $p_\perp$ (pion wind) in rescatterings which, due to the lack of $3 \to 2$ processes, 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 physics effects which must be taken into account, this clearly points to the need of further model improvement. Beyond re-tuning and adding $3 \to 2$ processes, the addition of string-string interactions before hadronic rescattering will change the over-



(a) (b)

Figure II.11: Charged multiplicities in PbPb collisions $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV. At mid-rapidity as (a) function of centrality, and (b) differential in η in centrality 0-5%. Data from ALICE [63, 64].

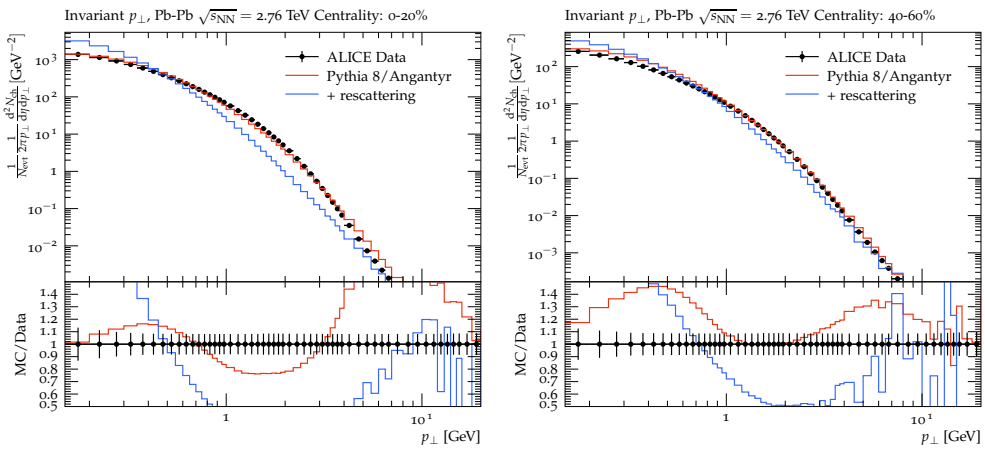
all soft kinematics. This is an important next step, which will be taken in a forthcoming publication.

4.2 Flow coefficients

As indicated in section 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 [66].

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, equations (II.5) and (II.7) cannot be applied directly. Instead the flow coefficients are calculated using two- and multi-particle azimuthal correlations using the so-called generic framework [67], implemented in the Rivet framework [68], including the use of sub-events [69].



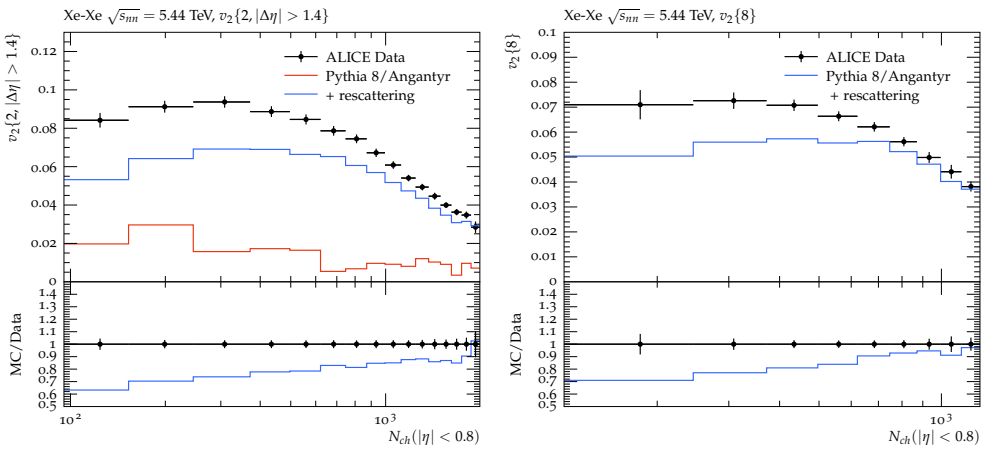
(a) (b)

Figure II.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 [65].

In figure II.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 section 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 figure II.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 that the agreement at high multiplicities is slightly better. As it was also observed in section 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. This is particularly interesting seen in the connection with recent results, that interactions between strings before



(a) (b)

Figure II.13: Elliptic flow in XeXe collisions at $\sqrt{s_{\mathrm{NN}}} = 5.44$ 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].

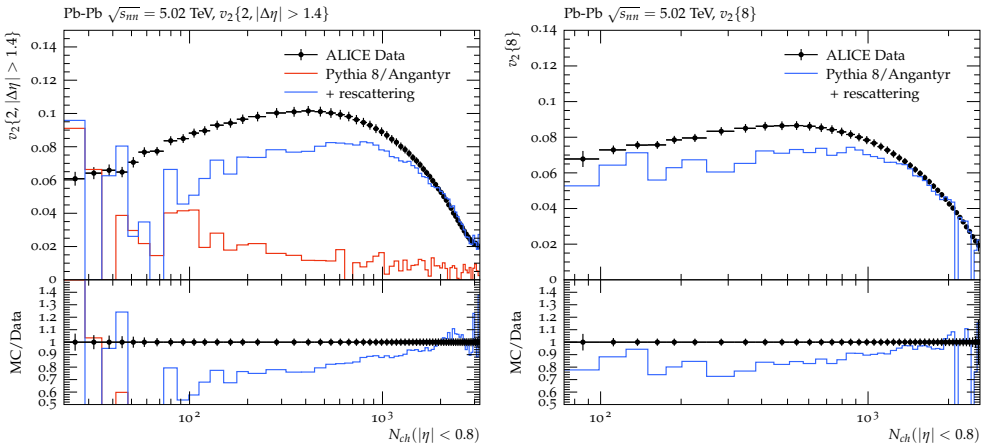
hadronization in the string shoving model [25] will also give a sizeable contribution to flow coefficients in heavy ion collisions, without fully describing data. The combination of the two frameworks, to test whether the combined effect is compatible with data, will be a topic for a future paper. It should be mentioned that the contributions from different models, acting one after the other, does not add linearly [30, 70].

4.3 Jet modifications from rescattering

As shown, both in Figure II.3 and Figure II.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. [71] 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-



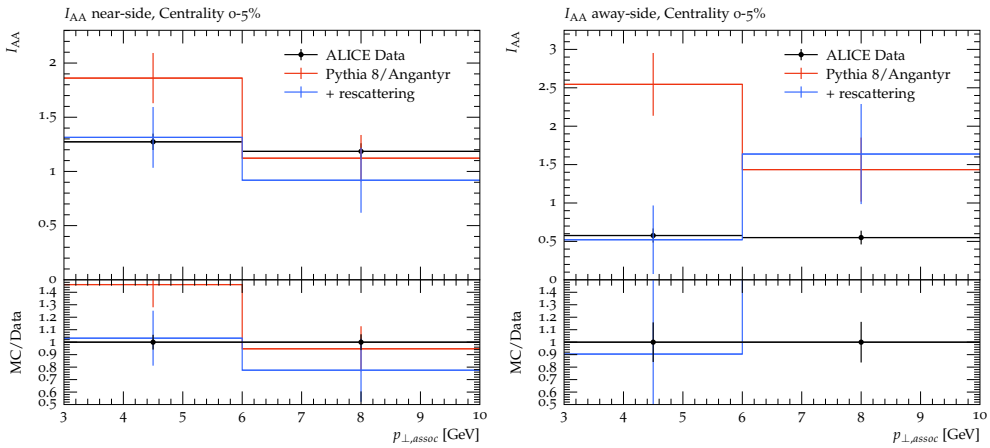
(a) (b)

Figure II.14: Elliptic flow in PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ 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].

to-back high-$p_\perp$ hadron correlations in central AuAu collisions at RHIC [72]. 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 $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV [73]. In this study, trigger particles of 8 GeV $< p_{\perp,\mathrm{trig}} <$ 15 GeV are correlated in φ with associated particles of 4 GeV $< p_{\perp,\mathrm{assoc}} < p_{\perp,\mathrm{trig}}$. 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 Figure II.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 [73].

It is seen that, by default, Pythia/Angantyr overestimates the away-side I_{AA} in the whole $p_{\perp,\mathrm{assoc}}$ 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 rescattering 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



(a) (b)

Figure II.15: Modification of high-$p_\perp$ azimuthal correlations, I_{AA}, in PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ 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 [73].

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 [31], which focused on validating it in the context of pp collisions. In this paper, the main focus has been physics studies in pPb and PbPb collisions.

Before going into finer details, it is worth to consider how rescattering changes the bulk properties of events. Notably it increases the charged multiplicity by about 20 % in PbPb events. One key reason is that we have implemented $2 \to n, n > 2$ processes but not $n \to 2$. For a system in thermal equilibrium the two kinds should balance each other. In the modelling of PbPb collisions, however, our original production is not a thermal process, and the subsequent rescattering is occurring in an expanding out-of-equilibrium

system. As an example, minijet production gives a larger rate of higher-$p_\perp$ particles than a thermal spectrum would predict, and therefore a fraction of higher invariant rescattering masses with more $2 \to n$ processes. While the inclusion of $n \to 2$ therefore is high on the priority list of future model developments, it is not likely to fully restore no-rescattering multiplicities. For pp, a slight retuning of the $p_{\perp 0}$ parameter for multiparton interactions helped restore approximate agreement with data, but for pPb and PbPb also some tweaks may be needed in the Angantyr framework. Until that is done, users need to be aware of such shortcomings.

The most obvious effect of rescattering is the changed shape of $p_\perp$ spectra, where pions lose momentum. Owing to their low mass they tend to be produced with higher velocity, which in collisions then can be used to speed up heavier particles. The above-mentioned $2 \to n$ processes act to reduce the overall $\langle p_\perp \rangle$ values, however, and for kaons and D mesons the result is a net slowdown. Here it should be remembered that the D's are produced only in perturbative processes, and so can have large velocities to begin with. For protons indeed a speedup can be observed, but here the significant rate of baryon–antibaryon annihilation clouds the picture.

Still at the basic level, the detailed event record allows us to map out both the space–time evolution, the nature of rescatterings and the change of particle composition. An example is resonance suppression, which was discussed briefly, but where an analysis paralleling the experimental one is outside the scope of the current article. Also of interest is the converse, the formation of particular particles in rescattering, such as $f_2(1270)$ resonances [74] or exotic hadrons. Care must be taken in such studies however, as 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.

Amongst the physics results presented, the most remarkable is the observation of a sizeable elliptic flow in AA collisions, where data is described particularly well at high multiplicities. 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, *ie.* when moving to larger systems. This should be contrasted with the pp results [31], where the rescattering flow effects were tiny and far below data. Thus rescattering may be one source of flow, but apparently not the only one. The Pythia/Angantyr framework also includes other effects that contribute to the flow, notably shoving [23], where an improved modelling [25] will soon be part of the standard code.

Another interesting observation is the suppression of J/ψ production in central collisions, by the breakup into D mesons in rescattering. The D meson rate is hardly affected, since it is more than two orders of magnitude larger to begin with. One should note that the handling of charm collisions largely is based on the Additive Quark Model, which does not

distinguish between J/ψ and ψ', so again there is room for improvement.

There are also examples where rescattering, as currently implemented, is going in the wrong direction. The $p_\perp$ spectrum in PbPb is reasonably well described without rescattering, but becomes way too soft with it. Hyperon production rates also drop, where data wants more such production [14].

The most important follow-up project in a not too distant future is to combine all the features that have been introduced on top of the basic Pythia/Angantyr model, notably ropes, shoving and rescattering, and 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.

Another possible application is (anti)deuteron production, and even heavier (anti)nuclei. In the past this has often been modelled using coalescence of particles close in momentum space, on the assumption that such particles also have been produced near to each other in space–time. (One such model is even included in Pythia [75].) This usually is not a bad approximation in e^+e^- annihilation or pp, at least as modelled by string fragmentation. But the much larger volume of particle production and rescattering in AA obviously requires due consideration to the space–time proximity of (anti)nucleons, and also that deuterons can break up by rescattering processes.

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.

Acknowledgements

Work supported in part by the Swedish Research Council, contract numbers 2016-05996 and 2017-003, in part by the MCnetITN3 H2020 Marie Curie Innovative Training Network, grant agreement 722104, and in part by the Knut and Alice Wallenberg foundation, contract number 2017.0036. This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No 668679.

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

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

Appendix - Algorithmic complexity

The rescattering algorithm needs to compare each hadron pair. This has an asymptotic complexity of $\mathcal{O}(n_{\text{record}}^2)$, 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, *eg.* if one of the compared particles 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 Figure .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 $\sim$ 2 for pp, $\sim$ 20 for pPb and $\sim$ 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.

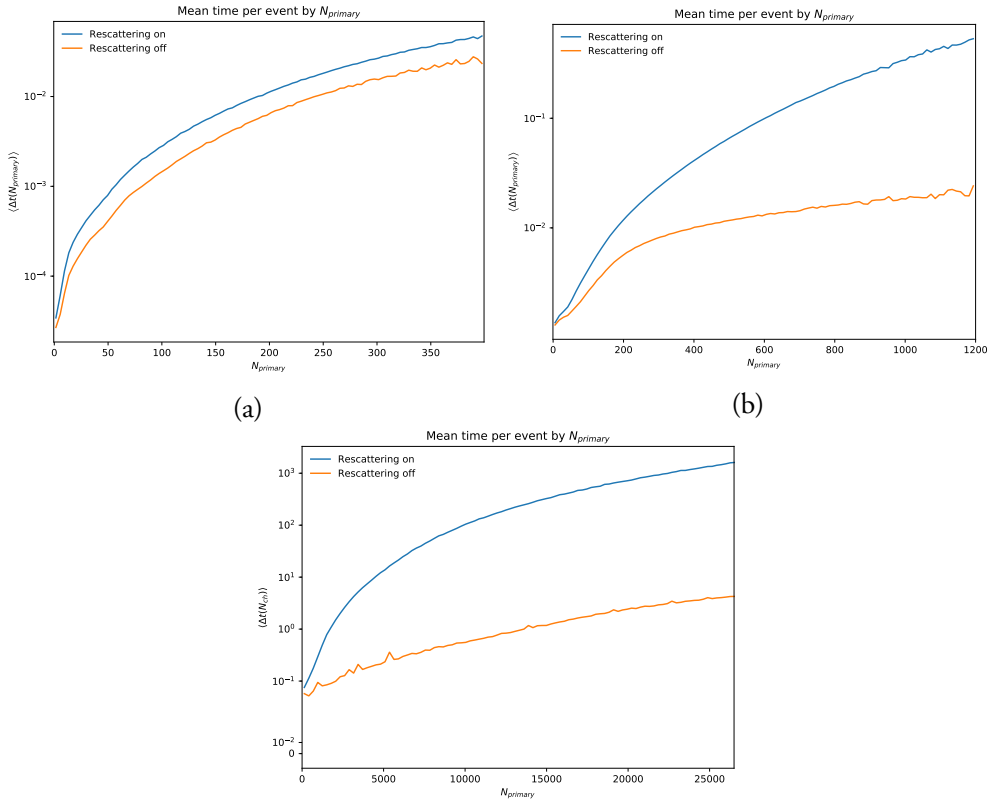


Figure .16: The average generation time of each event with a specified primary hadron multiplicity, for (a) pp, (b) pPb and (c) PbPb, with $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV.

1 Bibliography

[1] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, "Glauber modeling in high energy nuclear collisions," *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 205–243, nucl-ex/0701025.

[2] P. Bożek, W. Broniowski, and M. Rybczyński, "Wounded quarks in A+A, p+A, and p+p collisions," *Phys. Rev. C* **94** (2016), no. 1 014902, 1604.07697.

[3] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, "The Color Glass Condensate," *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 463–489, 1002.0333.

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

[5] A. Kurkela, A. Mazeliauskas, J.-F. Paquet, S. Schlichting, and D. Teaney, "Matching the Nonequilibrium Initial Stage of Heavy Ion Collisions to Hydrodynamics with QCD Kinetic Theory," *Phys. Rev. Lett.* **122** (2019), no. 12 122302, 1805.01604.

[6] U. Heinz and R. Snellings, "Collective flow and viscosity in relativistic heavy-ion collisions," *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 123–151, 1301.2826.

[7] M. Luzum and H. Petersen, "Initial State Fluctuations and Final State Correlations in Relativistic Heavy-Ion Collisions," *J. Phys. G* **41** (2014) 063102, 1312.5503.

[8] C. Gale, S. Jeon, and B. Schenke, "Hydrodynamic Modeling of Heavy-Ion Collisions," *Int. J. Mod. Phys. A* **28** (2013) 1340011, 1301.5893.

[9] R. D. Weller and P. Romatschke, "One fluid to rule them all: viscous hydrodynamic description of event-by-event central p+p, p+Pb and Pb+Pb collisions at $\sqrt{s} = 5.02$ TeV," *Phys. Lett. B* **774** (2017) 351–356, 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** (2010) 091, 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** (2016), no. 17 172301, 1509.04776.

[12] **CMS** Collaboration, V. Khachatryan *et. al.*, "Evidence for collectivity in pp collisions at the LHC," *Phys. Lett. B* **765** (2017) 193–220, 1606.06198.

[13] **ALICE** 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** (2019), no. 14 142301, 1903.01790.

[14] **ALICE** Collaboration, J. Adam *et. al.*, "Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions," *Nature Phys.* **13** (2017) 535–539, 1606.07424.

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

[16] T. Pierog, I. Karpenko, J. Katzy, E. Yatsenko, and K. Werner, "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider," *Phys. Rev. C* **92** (2015), no. 3 034906, 1306.0121.

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

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

[19] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP* **05** (2006) 026, hep-ph/0603175.

[20] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.* **191** (2015) 159–177, 1410.3012.

[21] B. Andersson, G. Gustafson, and B. Nilsson-Almqvist, "A Model for Low p(t) Hadronic Reactions, with Generalizations to Hadron - Nucleus and Nucleus-Nucleus Collisions," *Nucl. Phys. B* **281** (1987) 289–309.

[22] B. Andersson, G. Gustafson, and H. Pi, "The FRITIOF model for very high-energy hadronic collisions," *Z. Phys. C* **57** (1993) 485–494.

[23] C. Bierlich, G. Gustafson, and L. Lönnblad, "A shoving model for collectivity in hadronic collisions," 1612.05132.

[24] C. Bierlich, G. Gustafson, and L. Lönnblad, "Collectivity without plasma in hadronic collisions," *Phys. Lett. B* **779** (2018) 58–63, 1710.09725.

[25] C. Bierlich, S. Chakraborty, G. Gustafson, and L. Lönnblad, "Setting the string shoving picture in a new frame," 2010.07595.

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

[27] S. Bass *et. al.*, "Microscopic models for ultrarelativistic heavy ion collisions," *Prog. Part. Nucl. Phys.* **41** (1998) 255–369, 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** (2016), no. 5 054905, 1606.06642.

[29] S. Ferreres-Solé and T. Sjöstrand, "The space–time structure of hadronization in the Lund model," *Eur. Phys. J. C* **78** (2018), no. 11 983, 1808.04619.

[30] A. V. da Silva, W. M. Serenone, D. Dobrigkeit Chinellato, J. Takahashi, and C. Bierlich, "Studies of heavy-ion collisions using PYTHIA Angantyr and UrQMD," 2002.10236.

[31] T. Sjöstrand and M. Utheim, "A Framework for Hadronic Rescattering in pp Collisions," *Eur. Phys. J. C* **80** (2020), no. 10 907, 2005.05658.

[32] W. Broniowski, M. Rybczynski, and P. Bozek, "GLISSANDO: Glauber initial-state simulation and more..," *Comput. Phys. Commun.* **180** (2009) 69–83, 0710.5731.

[33] M. Rybczynski, G. Stefanek, W. Broniowski, and P. Bozek, "GLISSANDO 2 : GLauber Initial-State Simulation AND mOre…, ver. 2," *Comput. Phys. Commun.* **185** (2014) 1759–1772, 1310.5475.

[34] G. A. Schuler and T. Sjöstrand, "Hadronic diffractive cross-sections and the rise of the total cross-section," *Phys. Rev. D* **49** (1994) 2257–2267.

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

[36] T. Sjöstrand, "The Development of MPI Modeling in Pythia," *Adv. Ser. Direct. High Energy Phys.* **29** (2018) 191–225, 1706.02166.

[37] C. Bierlich and C. O. Rasmussen, "Dipole evolution: perspectives for collectivity and γ^*A collisions," *JHEP* **10** (2019) 026, 1907.12871.

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

[39] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton Fragmentation and String Dynamics," *Phys. Rept.* **97** (1983) 31–145.

[40] G. 't Hooft, "A Planar Diagram Theory for Strong Interactions," *Nucl. Phys. B* **72** (1974) 461.

[41] T. Sjöstrand, "Jet Fragmentation of Nearby Partons," *Nucl. Phys. B* **248** (1984) 469–502.

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

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

[44] J. Pelaez, A. Rodas, and J. Ruiz De Elvira, "Global parameterization of $\pi\pi$ scattering up to 2 GeV," *Eur. Phys. J. C* **79** (2019), no. 12 1008, 1907.13162.

[45] J. Pelaez and A. Rodas, "Pion-kaon scattering amplitude constrained with forward dispersion relations up to 1.6 GeV," *Phys. Rev. D* **93** (2016), no. 7 074025, 1602.08404.

[46] **Particle Data Group** Collaboration, M. Tanabashi *et. al.*, "Review of Particle Physics," *Phys. Rev. D* **98** (2018), no. 3 030001.

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

[48] P. Koch and C. Dover, "$K^{\pm}$, $\bar{p}$ and Ω^{-} Production in Relativistic Heavy Ion Collisions," *Phys. Rev. C* **40** (1989) 145–155.

[49] G. A. Schuler and T. Sjöstrand, "A Scenario for high-energy gamma gamma interactions," *Z. Phys. C* **73** (1997) 677–688, hep-ph/9605240.

[50] E. Levin and L. Frankfurt, "The Quark hypothesis and relations between cross-sections at high-energies," *JETP Lett.* **2** (1965) 65–70.

[51] H. J. Lipkin, "Quarks for pedestrians," *Phys. Rept.* **8** (1973) 173–268.

[52] **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** (2015), no. 5 226, 1504.00024.

[53] N. Fischer and T. Sjöstrand, "Thermodynamical String Fragmentation," *JHEP* **01** (2017) 140, 1610.09818.

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

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

[56] C. Bierlich and J. R. Christiansen, "Effects of color reconnection on hadron flavor observables," *Phys. Rev. D* **92** (2015), no. 9 094010, 1507.02091.

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

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

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

[60] **ALICE** Collaboration, B. B. Abelev *et. al.*, "Azimuthal anisotropy of D meson production in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV," *Phys. Rev. C* **90** (2014), no. 3 034904, 1405.2001.

[61] **ALICE** Collaboration, E. Abbas *et. al.*, "J/Psi Elliptic Flow in Pb-Pb Collisions at $\sqrt{s_{\rm NN}}$ = 2.76 TeV," *Phys. Rev. Lett.* **111** (2013) 162301, 1303.5880.

[62] C. Bierlich *et. al.*, "Robust Independent Validation of Experiment and Theory: Rivet version 3," *SciPost Phys.* **8** (2020) 026, 1912.05451.

[63] **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** (2011) 032301, 1012.1657.

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

[65] **ALICE** Collaboration, B. Abelev *et. al.*, "Centrality Dependence of Charged Particle Production at Large Transverse Momentum in Pb–Pb Collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV," *Phys. Lett. B* **720** (2013) 52–62, 1208.2711.

[66] D. Teaney and L. Yan, "Triangularity and Dipole Asymmetry in Heavy Ion Collisions," *Phys. Rev. C* **83** (2011) 064904, 1010.1876.

[67] A. Bilandzic, C. H. Christensen, K. Gulbrandsen, A. Hansen, and Y. Zhou, "Generic framework for anisotropic flow analyses with multiparticle azimuthal correlations," *Phys. Rev. C* **89** (2014), no. 6 064904, 1312.3572.

[68] C. Bierlich *et. al.*, "Confronting experimental data with heavy-ion models: RIVET for heavy ions," *Eur. Phys. J. C* **80** (2020), no. 5 485, 2001.10737.

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

[70] J. Auvinen and H. Petersen, "Evolution of elliptic and triangular flow as a function of $\sqrt{s_{NN}}$ in a hybrid model," *Phys. Rev. C* **88** (2013), no. 6 064908, 1310.1764.

[71] P. Dorau, J.-B. Rose, D. Pablos, and H. Elfner, "Jet Quenching in the Hadron Gas: An Exploratory Study," *Phys. Rev. C* **101** (2020), no. 3 035208, 1910.07027.

[72] **STAR** 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** (2003) 082302, nucl-ex/0210033.

[73] **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** (2012) 092301, 1110.0121.

[74] P. Lebiedowicz and 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** (2020) 135816, 2007.12485.

[75] L. A. Dal and A. R. Raklev, "Alternative formation model for antideuterons from dark matter," *Phys. Rev. D* **91** (2015), no. 12 123536, 1504.07242. [Erratum: Phys.Rev.D 92, 069903 (2015), Erratum: Phys.Rev.D 92, 089901 (2015)].

Paper III

III

Forming Molecular States with Hadronic Rescattering

Philip Ilten and Marius Utheim

To be submitted to *Eur. Phys. J. A*,
e-Print: `arXiv:2108.03479` [hep-ph]
MCnet–21–13, LU TP 21–31

Abstract

A method for modelling the prompt production of molecular states using the hadronic rescattering framework of the general-purpose PYTHIA event generator is introduced. Production cross sections of possible exotic hadronic molecules via hadronic rescattering at the LHC are calculated for the $\chi_{c1}(3872)$ resonance, a possible tetraquark state, as well as three possible pentaquark states, $P_c^+(4312)$, $P_c^+(4440)$, and $P_c^+(4457)$. For the P_c^+ states, the expected cross section from Λ_b^0 decays is compared to the hadronic-rescattering production. The $\chi_{c1}(3872)$ cross section is compared to the fiducial $\chi_{c1}(3872)$ cross-section measurement by LHCb and found to contribute at a level of $\mathcal{O}(1\%)$. Finally, the expected yields of P_c^+ production from hadronic rescattering during Run 3 of LHCb are estimated. The prompt background is found to be significantly larger than the prompt P_c^+ signal from hadronic rescattering.

1 Introduction

While exotic bound quark states beyond the minimal $q\bar{q}$ meson and the qqq baryon structure have been proposed for some time [1–5], most experimentally observed hadrons fit these minimal bound quark configurations. Prior to the Large Hadron Collider (LHC), a number of observations for both exotic tetraquark, $qq\bar{q}\bar{q}$, and pentaquark states, $qqqq\bar{q}$, were claimed. However, many of these exotic states could not be verified by later experiments [6], excepting the Z(4430) [7] and $\chi_{c1}(3872)$ resonances [8] which remain as possible tetraquark candidates. Now with the LHC, more than 60 new hadrons have been observed with over 10 new candidates for exotic tetraquark and pentaquark states.

In 2015, the LHCb experiment discovered two resonances in the $J/\psi\,p$ mass spectrum, which were identified as possible pentaquark candidates, $P_c^+(4380)$ and $P_c^+(4450)$ [9].[1] A subsequent 2019 LHCb analysis with a larger data sample observed a possible additional resonance, $P_c^+(4312)$, and resolved the $P_c^+(4450)$ pentaquark structure as two resonances, $P_c^+(4440)$ and $P_c^+(4457)$ [10]. With this new observed mass structure, a more in-depth amplitude study of the observed $P_c^+(4380)$ resonance must be performed, leaving the existence of the $P_c^+(4380)$ state ambiguous. The three viable P_c^+ candidates, $P_c^+(4312)$, $P_c^+(4440)$, and $P_c^+(4457)$, were not observed via prompt production from the $p\,p$ collision point, but rather from the decay of Λ_b^0 baryons

Similarly, the $\chi_{c1}(3872)$ state (also known as X(3872)) was first observed through B-meson decays by Belle in 2003 [8], which was later confirmed by BaBar [11]. More recently, LHCb measured the quantum numbers of the $\chi_{c1}(3872)$ to be $J^{PC} = 1^{++}$ [12, 13]. Lying within 0.2 GeV of the $D^0\,\bar{D}^{*0}$ threshold, the $\chi_{c1}(3872)$ resonance is oftentimes interpreted as such a molecular state [14]. The alternative interpretation of the $\chi_{c1}(3872)$ as a four-quark state is also a possibility, but the molecular interpretation remains as the preferred model [15–21]. The $P_c^+(4312)$, $P_c^+(4440)$ and $P_c^+(4457)$ pentaquarks observed by LHCb may also be described by a molecular state of $\Sigma_c^*\,\bar{D}^0$ or $\Sigma_c^*\,\bar{D}^{*0}$ [22–30]. Other models have also been proposed for the P_c^+ states such as hadro-charmonium, a compact charmonium state bound in light hadronic matter [31].

The experimental observations above of exotic hadrons all consider production from heavy hadron decays, B-meson decays for the $\chi_{c1}(3872)$ state and Λ_b^0 for the P_c^+ states. Already, some predictions for prompt production have been made using a coalescence type model where free constituents of the molecular state may combine into a bound molecular state if close in momentum space [32, 33]. These types of models have been successful in modelling deuteron production at the LHC, including a full implementation in Pythia [34], a general-purpose event generator which allows for parameterised cross sections differential

[1] Inclusion of charge conjugate states and processes are implied throughout this work, unless explicitly noted in the text.

in momentum space for multiple initial states. However, these models do not consider the complete space-time picture of LHC events, and require tuning of coalescence parameters to data, whether cut-offs or overall normalisations.

Recent developments in Pythia now allow hadronic resonances to be formed from hadronic rescattering in a full space-time picture [35, 36], where only the partial widths of the hadronic resonance being formed are required to fully specify the model. In this paper, this hadronic rescattering framework has been modified to predict prompt exotic hadron production for both the $\chi_{c1}(3872)$ and P_c^+ states at the LHC. The details of the models used to describe the exotic hadrons are introduced in Section 2, while results are given in Section 3 and conclusions are given in Section 4.

2 Models for exotic hadron production

The hadronic rescattering framework of Pythia can perform $2 \to 1$ scattering where the initial state hadrons combine to form a resonance hadron. The cross section of this process depends on the mass and total width of the resonance, as well as the partial width of the given channel for the resonance. In the default Pythia framework, only pre-defined resonances and rescattering channels can be used for rescattering. In this work, the framework has been expanded to allow the addition of any arbitrary hadron resonance production from rescattering. Specific configurations for $\chi_{c1}(3872)$, $P_c^+(4312)$, $P_c^+(4440)$, and $P_c^+(4457)$ are then defined, given model assumptions, to determine the hadronic rescattering cross sections.

2.1 Exotic hadron properties

The $\chi_{c1}(3872)$ mass is well measured to be 3871.69 ± 0.17 MeV [37]. However, the $\chi_{c1}(3872)$ widths, both total and partial, are not as well known. In this work, the total width of the $\chi_{c1}(3872)$ is set to the world average, 1.19 ± 0.21 MeV, which is a combination of a dedicated inclusive LHCb line-shape analysis [38] and a measurement by LHCb of $\chi_{c1}(3872)$ production from B-decays [39]. The partial widths of the $\chi_{c1}(3872)$ are set by normalising the central branching ratios, $\mathcal{B}$, reported by the PDG, and multiplying these by the total width. The experimental uncertainty on these branching ratios is large, but the $D^0\,\bar{D}^{*0}$ and $D^0\,\pi^0\,\bar{D}^0$ channels dominate. Because the latter is not a two-body decay, the $\chi_{c1}(3872)$ cannot be produced in rescattering through this channel, but it still gives a significant contribution to the total cross section. Likewise, the $J/\psi\gamma$ and $\psi(2S)\gamma$ are not used for resonance formation since photons are ignored by the rescattering framework, but their contributions to the total cross section are still included. The branching ratios and partial widths used for the $\chi_{c1}(3872)$ are given in Table III.1.

Table III.1: Experimentally measured branching ratios for the $\chi_{c1}(3872)$, as taken from the PDG [37]. The partial widths, in MeV, for hadronic rescattering are given for each channel and are calculated as the product of the normalised branching ratio and the experimentally measured $\chi_{c1}(3872)$ width of 1.19 ± 0.21 MeV [37].

	PDG $\mathcal{B}$	width [MeV]
$D^0\,\bar{D}^{*0}$	$(3.7 \pm 0.9)\times10^{-1}$	4.3×10^{-1}
$J/\psi\,\omega$	$(4.3 \pm 2.1)\times10^{-2}$	5.0×10^{-2}
$J/\psi\,\rho^0$	$(3.8 \pm 1.2)\times10^{-2}$	4.4×10^{-2}
$\chi_{c1}\,\pi^0$	$(3.4 \pm 1.6)\times10^{-2}$	4.0×10^{-2}
$J/\psi\,\gamma$	$(8.0 \pm 4.0)\times10^{-3}$	9.3×10^{-3}
$\psi(2S)\,\gamma$	$(4.5 \pm 2.0)\times10^{-2}$	5.3×10^{-2}
$D^0\,\pi^0\,\bar{D}^0$	$(4.9\,^{+\,1.8}_{-\,2.0})\times10^{-1}$	5.7×10^{-1}

The masses of the P_c^+ resonances are set to the central values of the LHCb measurements. Experimental observations of the P_c^+ candidates are limited to the $p\,J/\psi$ decay channel, and so theory predictions based on a molecular model from Ref. [40] are used instead to define the total and partial widths. There, pentaquarks are treated as $\Sigma_c^+\,\bar{D}^0/\Sigma_c^+\,\bar{D}^{*0}$ molecular states, with the $P_c^+(4312)$ resonance considered as a spin-1/2 $\Sigma_c^+\,\bar{D}^0$ state. The $P_c^+(4440)$ and $P_c^+(4457)$ resonances are treated as $\Sigma_c^+\,\bar{D}^{*0}$ states, and two possible spin assignments are considered, either spin-1/2 or spin-3/2. For all spin configurations, the predicted total widths for the P_c^+ states are consistent with the observed widths, although these widths have large experimental uncertainty. Ref. [40] suggests that the $P_c^+(4440)$ resonance is most likely spin-1/2 and the $P_c^+(4457)$ resonance is spin-3/2, but notes that the opposite assignment cannot be excluded. Different form factors can also be used, and so two different models are considered in this study. Model 1 uses the (f_1, f_3) form factor set of Ref. [40], while model 2 uses the (f_2, f_3) set. Both models assume the $P_c^+(4440)$ resonance is spin-1/2 and the $P_c^+(4457)$ resonance is spin-3/2. The pentaquark partial widths used in this paper are summarised in Table III.2.

2.2 Exotic hadron production from rescattering

In the hadronic rescattering framework of Pythia, two hadrons will interact if they pass each other in their centre-of-mass (CM) frame with an impact parameter $b < \sqrt{\sigma/\pi}$, where σ is the total cross section depending on the particle species and the CM energy. When two particles do interact, the specific process to simulate is chosen with a probability proportional to the partial cross section of that process. Since rescattering in Pythia increases the charged particle multiplicity, the recommendation of Ref. [35] is followed, where setting $p_{\perp 0}$ parameter of the multi-parton interaction (MPI) framework to 2.345 GeV compensates for this effect. This reduces the event multiplicity before rescattering, with respect to the default Pythia MPI tune.

Table III.2: Partial widths in MeV for each pentaquark state of the two models considered from Ref. [40]. For both models the $P_c^+(4440)$ is chosen to be spin-1/2 while the $P_c^+(4457)$ is chosen to be spin-3/2.

	model 1 width [MeV]			model 2 width [MeV]		
	$P_c^+(4312)$	$P_c^+(4440)$	$P_c^+(4457)$	$P_c^+(4312)$	$P_c^+(4440)$	$P_c^+(4457)$
$\Lambda_c^+ \bar{D}^0$	6.0×10^{-2}	5.6	1.5	3.0×10^{-1}	2.7	1.2
$\Lambda_c^+ \bar{D}^{*0}$	3.8	1.4×10^{1}	6.1	1.1×10^{1}	1.2×10^{1}	6.9
$\Sigma_c^+ \bar{D}^0$	–	3.4	1.0	–	3.4	9.0×10^{-1}
$\Sigma_c^{*+} \bar{D}^0$	–	8.0×10^{-1}	6.2	–	9.0×10^{-1}	7.2
$n\,\pi^+$	2.0×10^{-3}	1.0×10^{-3}	5.0×10^{-5}	8.5×10^{-1}	1.0×10^{-1}	3.0×10^{-1}
$n\,\rho^+$	2.0×10^{-5}	1.5×10^{-4}	1.0×10^{-5}	4.0×10^{-4}	2.0×10^{-1}	5.0×10^{-2}
$p\,\pi^0$	2.0×10^{-3}	1.0×10^{-3}	5.0×10^{-5}	8.5×10^{-1}	1.0×10^{-1}	3.0×10^{-1}
$p\,\rho^0$	2.0×10^{-5}	1.5×10^{-4}	1.0×10^{-5}	4.0×10^{-4}	2.0×10^{-1}	5.0×10^{-2}
$p\,\omega$	1.0×10^{-4}	1.0×10^{-4}	9.0×10^{-5}	3.0×10^{-3}	1.5	4.0×10^{-1}
$p\,\eta_c$	1.0×10^{-2}	3.0×10^{-4}	6.0×10^{-5}	4.0×10^{-1}	7.0×10^{-2}	3.0×10^{-3}
$p\,J/\psi$	1.0×10^{-3}	3.0×10^{-2}	1.0×10^{-2}	1.0×10^{-1}	6.0×10^{-1}	6.0×10^{-1}
$p\,\chi_{c0}$	–	8.0×10^{-4}	3.0×10^{-5}	–	1.0×10^{-1}	3.0×10^{-3}

For processes involving charm hadrons, the total cross section in Pythia is calculated using the additive quark model (AQM) [41, 42]. In this model, the total cross section for two initial hadrons A and B is given by

$$\sigma_{\mathrm{AQM,tot}} = (40\ \mathrm{mb})\frac{n_{\mathrm{eff},A}}{3}\frac{n_{\mathrm{eff},B}}{3}, \tag{III.1}$$

where n_{eff} is the effective number of quarks in each hadron. In Pythia, this number is defined from the quark numbers of the hadron, and by default is

$$n_{\mathrm{eff}} = n_{\mathrm{u}} + n_{\mathrm{d}} + 0.6n_{\mathrm{s}} + 0.2n_{\mathrm{c}} + 0.07n_{\mathrm{b}}. \tag{III.2}$$

As an example, the total $p\,J/\psi$ cross section determined by the AQM is $\sigma_{\mathrm{AQM,tot}} = 5.33$ mb. The cross section for elastic scattering (in mb) is also determined with AQM,

$$\sigma_{\mathrm{AQM,el}} = 0.039\sigma_{\mathrm{AQM}}^{3/2}. \tag{III.3}$$

The difference between the total AQM cross section and the elastic AQM cross section gives an inelastic AQM cross section of

$$\sigma_{\mathrm{AQM,inel}} = \sigma_{\mathrm{AQM,tot}} - \sigma_{\mathrm{AQM,el}}, \tag{III.4}$$

which in default Pythia corresponds to diffractive and non-diffractive interactions.

In the model of Ref. [40], pentaquarks can in principle also form in $n\,\rho^0$, $p\,\rho^0$, or $p\,\omega$ interactions. Pythia also uses the AQM model for the total cross section of these processes, but pentaquark formation through these processes is so rare that the contribution will be

negligible. Finally, pentaquarks can be produced in nucleon–pion interactions, *i.e.* the $\mathrm{n}\,\pi^+$ and $\mathrm{p}\,\pi^0$ channels. The formation probability is also very small here, but now the contribution may be non-negligible due to the abundance of these particles in LHC collisions. For these processes, the total cross sections at energies near the pentaquark masses are given by the $\mathrm{HPR_1R_2}$ parameterisation [43].

The partial cross section for a resonance formation process $AB \to R$ is given by a nonrelativistic Breit–Wigner [43],

$$\sigma_{\mathrm{res}} = \frac{\pi}{p_{\mathrm{CM}}^2} \frac{(2S_R+1)}{(2S_A+1)(2S_B+1)} \frac{\Gamma_{R\to AB}\Gamma_R}{(m_R - E_{\mathrm{CM}})^2 + \frac{1}{4}\Gamma_R^2}, \tag{III.5}$$

where p_{CM} and E_{CM} are the momentum and energy of the incoming particles in their CM frame, S is the spin of each particle, and m_R and Γ_R, $\Gamma_{R\to AB}$ are the mass, total width, and partial width of the resonance, respectively. These widths are mass dependent, as described in Ref. [44], giving mass distributions as shown in Figure III.1. Using these widths can give mass distributions with longer tails than are physically reasonable, and therefore explicit mass bounds are required. These explicit cut-offs can give discontinuities in the mass distribution, but this is not expected to significantly affect any relevant physical observables.

It is important to keep in mind that resonance formation does not change the mass spectrum when the decay products are the same as the incoming particles. For instance, the process $\Lambda_{\mathrm{c}}^+\bar{\mathrm{D}} \to \mathrm{P}_{\mathrm{c}}^+(4440) \to \Lambda_{\mathrm{c}}^+\bar{\mathrm{D}}$ will not change the $\Lambda_{\mathrm{c}}^+\bar{\mathrm{D}}$ mass spectrum, as they must already be correlated in order to form the resonance. However, in a system out of equilibrium, resonances can change the relative composition of particles. If resonance production receives a significant contribution from a particular channel, *e.g.* $\mathrm{p}\,\pi^0$ where the flux is large, but their decays are dominated by a different channel, *e.g.* $\mathrm{P}_{\mathrm{c}}^+ \to \Lambda_{\mathrm{c}}^+\bar{\mathrm{D}}$, then a peak structure would be appear in that decay channel.

When resonance formation is possible, the total cross section is fixed, and $\sigma_{\mathrm{AQM,inel}}$ is reduced by the resonance cross section. If σ_{res} is greater than $\sigma_{\mathrm{AQM,inel}}$, the total cross section is increased to $\sigma_{\mathrm{res}} + \sigma_{\mathrm{AQM,el}}$, which as an example, occurs for the $\Lambda_{\mathrm{c}}^+\,\bar{\mathrm{D}}^{*0} \to \mathrm{P}_{\mathrm{c}}^+(4312)$ process. The cross sections for tetraquark and pentaquark resonances, as a function of the rescattering centre-of-mass energy, are shown in Figure III.2 for the primary rescattering channels. Some of these cross sections grow very large near the kinematic threshold of the channel, which is particularly visible for $\mathrm{D}^0\,\bar{\mathrm{D}}^{*0} \to \chi_{\mathrm{c1}}(3872)$. The technical reason for this is that the lower mass bound for the particle lies below the threshold so the width does not vanish, hence the factor $1/p_{\mathrm{CM}}^2$ in (III.5) dominates. Physically, this can be motivated by the fact that slow-moving particles spend more time near each other, and have a larger chance of interacting. For the $\chi_{\mathrm{c1}}(3872)$ resonance, however, the cross section grows larger than what might be considered reasonable considering the range of strong interactions. In the rescattering framework, the range of interactions is capped at a generous 5 fm, corresponding to a cross section of roughly 785 mb, which limits the $\chi_{\mathrm{c1}}(3872)$ cross section.

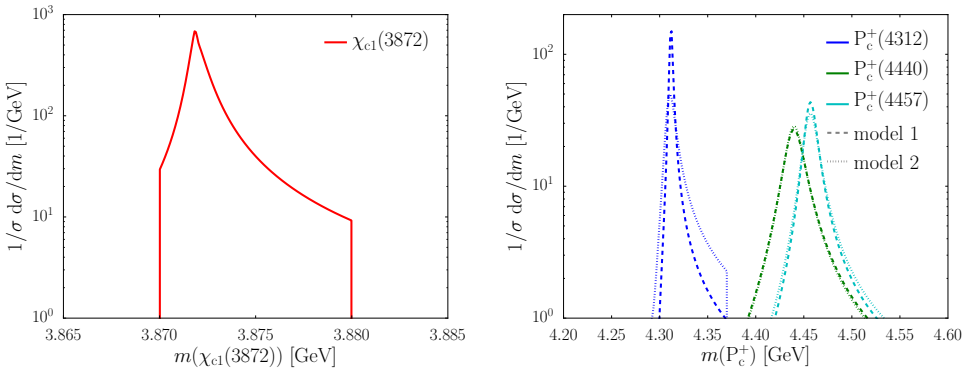


Figure III.1: Mass distributions for the (left) $\chi_{c1}(3872)$ tetraquark and (right) P_c^+ pentaquarks.

The interpretation of such extremely large cross sections is not clear, but a more detailed handling is outside the scope of this study.

2.3 Pentaquark production from Λ_b^0 decays

While the focus of this study is exotic hadron production from hadronic rescattering, it is useful in the context of the P_c^+ states to compare these prompt production cross sections with the expected cross sections from Λ_b^0 decays. The branching ratios of the Λ_b^0 into P_c^+ states have not been experimentally measured, but can be fully determined given the partial widths of Table III.2 and the results of Ref. [10]. Here the contribution ratio is defined as

$$\mathcal{R} = \frac{\mathcal{B}(\Lambda_b^0 \to \mathrm{P}_c^+ \mathrm{K}^-)\,\mathcal{B}(\mathrm{P}_c^+ \to \mathrm{p}\,\mathrm{J}/\psi)}{\mathcal{B}(\Lambda_b^0 \to \mathrm{p}\,\mathrm{J}/\psi\,\mathrm{K}^-)}, \tag{III.6}$$

and has been measured for each P_c^+ resonance. The branching ratio in the denominator is set as the experimentally measured value of $\mathcal{B}(\Lambda_b^0 \to \mathrm{p}\,\mathrm{J}/\psi\,\mathrm{K}^-) = (3.2^{+0.6}_{-0.5}) \times 10^{-4}$ [45]. By considering $\mathcal{B}(\mathrm{P}_c^+ \to \mathrm{p}\,\mathrm{J}/\psi)$ for each pentaquark state as set by Table III.2, the relevant Λ_b^0 branching ratios can be determined by

$$\mathcal{B}(\Lambda_b^0 \to \mathrm{P}_c^+ \mathrm{K}^-) = \mathcal{B}(\Lambda_b^0 \to \mathrm{p}\,\mathrm{J}/\psi\,\mathrm{K}^-)\frac{\mathcal{R}}{\mathcal{B}(\mathrm{P}_c^+ \to \mathrm{p}\,\mathrm{J}/\psi)}. \tag{III.7}$$

Values of $\mathcal{R}$ for each pentaquark state are also taken from experiment [10] and are given in Table III.3, where the Λ_b^0 branching ratios are also provided. Note that in model 1, the branching ratios for $\mathrm{P}_c^+ \to \mathrm{p}\,\mathrm{J}/\psi$ are lower than for model 2 by an order of magnitude or more, which gives a much smaller $\Lambda_b^0 \to \mathrm{P}_c^+ \mathrm{K}^-$ branching ratio for the former.

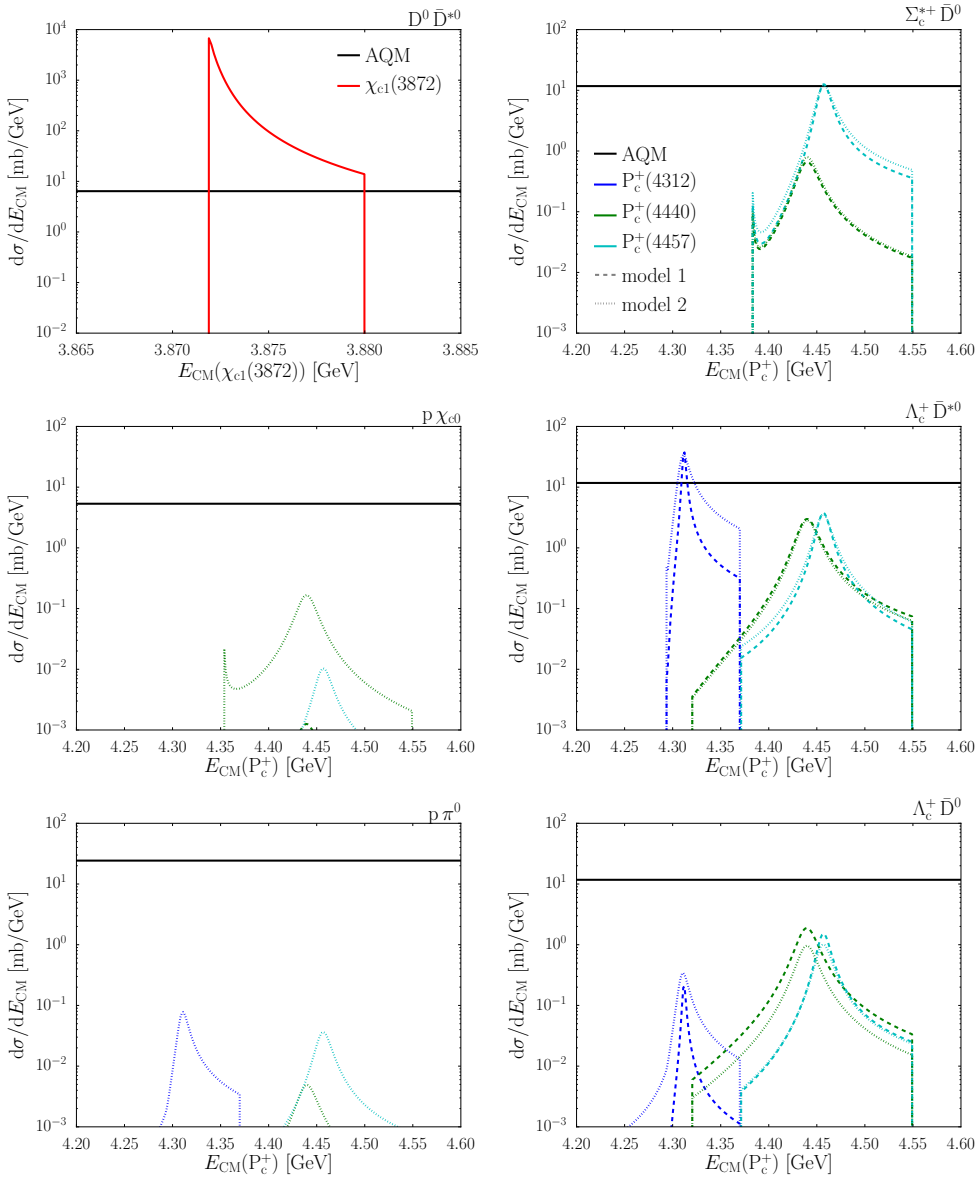


Figure III.2: Cross sections of the tetraquark and pentaquark resonance formation for relevant two-particle rescattering channels. The (solid black) total AQM cross section corresponds to the Pythia default in the absence of exotic hadrons, and while drawn across the entire E_{CM} range is only available above mass threshold for each resonance. For the pentaquarks, cross sections are calculated using (dashed) model 1 and (dotted) model 2 given Table III.2.

Table III.3: Branching ratios for $\Lambda_b^0 \to P_c^+ K^-$, determined using $\mathcal{R}$ [10] and $\mathcal{B}(\Lambda_b^0 \to p J/\psi K^-)$ [45] from data, and $\mathcal{B}(P_c^+ \to p J/\psi)$ calculated from Table III.2. Both $\mathcal{R}$ and $\mathcal{B}(P_c^+ \to p J/\psi)$ are listed for each pentaquark state, where the individual uncertainties on $\mathcal{R}$ have been combined in quadrature.

	$P_c^+(4312)$	$P_c^+(4440)$	$P_c^+(4457)$
$\mathcal{R}$	$(3.0\,^{+\,3.5}_{-\,1.1})\times10^{-3}$	$(1.1\,^{+\,0.4}_{-\,0.3})\times10^{-2}$	$(5.3\,^{+\,2.2}_{-\,2.1})\times10^{-3}$
model 1 $\mathcal{B}(\Lambda_b^0 \to P_c^+ K^-)$	3.7×10^{-3}	2.8×10^{-3}	2.5×10^{-3}
$\mathcal{B}(P_c^+ \to p J/\psi)$	2.6×10^{-4}	1.3×10^{-3}	6.8×10^{-4}
model 2 $\mathcal{B}(\Lambda_b^0 \to P_c^+ K^-)$	1.3×10^{-4}	1.3×10^{-4}	5.1×10^{-5}
$\mathcal{B}(P_c^+ \to p J/\psi)$	7.6×10^{-3}	2.7×10^{-2}	3.4×10^{-2}

3 Results

The cross sections of Figure III.2 are not production cross sections, but instead must be combined with the relevant flux of initial state particles which can rescatter to produce molecular states. High multiplicity environments are necessary to provide a sufficiently large initial state flux, and so high energy hadronic collisions, such as those produced at the LHC, are an ideal laboratory to study possible molecular state formation from hadronic rescattering. During Run 1, the LHC collided proton-proton beams at $\sqrt{s} = 7$ and 8 TeV, while during Run 2 $\sqrt{s} = 13$ TeV, which corresponds to the majority of the LHC data set. During Run 3, the LHC is expected to run at a CM energy of 14 TeV, although 13 TeV may also be used, depending upon the performance of the collider. Consequently, a configuration with $\sqrt{s} = 13$ TeV $p\,p$ collisions is conservatively chosen for this study, since particle flux also increases as $\sqrt{s}$ is increased.

Using the default Pythia parameter tune [46] and a modified version of Pythia 8.306, the average visible final state particle multiplicity for inelastic LHC events at $\sqrt{s} = 13$ TeV is expected to be $\mathcal{O}(200)$, with an inelastic cross-section of 78 mb. This predicted inelastic cross-section is in good agreement with LHC measurements [47, 48], including forward measurements from LHCb [49] and TOTEM [50]. The predicted particle density and energy flow distributions also describe experimental LHC data well [51–54], across a number of experimental event categorisations intended to separate elastic, diffractive, and inelastic scattering. Individual particle species are also typically described well [55, 56], including open-charm meson production [57], although experimental measurements for many rare mesons and baryons are not available for direct comparison.

The light pseudo-scalar mesons, π^0 and $\pi^\pm$, each have an average multiplicity of $\mathcal{O}(50)$, while the average multiplicity for light meson and baryons is at the level of $\mathcal{O}(10)$ per species. This includes the ρ^0, $\rho^\pm$, and ω vector mesons, and the $p/\bar{p}$ and $n/\bar{n}$ light baryons. The $D^0/\bar{D}^0$ and $D^{*0}/\bar{D}^{*0}$ mesons each contribute at an average multiplicity of $\mathcal{O}(10^{-1})$,

while the $\Lambda_c^{\pm}$ contributes at an average multiplicity of $\mathcal{O}(10^{-2})$. The relevant quarkonia states, J/ψ and χ_{c0}, have average multiplicities of $\mathcal{O}(10^{-3})$, where both colour octet and singlet contributions are included. For the J/ψ meson, feed-down production from both χ_c-meson and B-hadron decays is also included. The η_c contribution is significantly less at $\mathcal{O}(10^{-4})$, but this is a known underestimation by Pythia, since production is included only through hadronization and not through direct nonrelativistic QCD (NRQCD) calculations.

The production of both $\Sigma_c^{\pm}$ and $\Sigma_c^{*\pm}$ baryons is relatively rare, with average multiplicities of $\mathcal{O}(10^{-4})$ and $\mathcal{O}(10^{-3})$, respectively. The production of $\Lambda_b^0/\bar{\Lambda}_b^0$, relevant for displaced production of pentaquark states, is also rare with an average multiplicity of $\mathcal{O}(10^{-3})$. Given these average multiplicities, the $D^0\,\bar{D}^{*0}$ rescattering channel is expected to dominate $\chi_{c1}(3872)$ production, while the rescattering channels with Λ_c^+ baryons are expected to dominate P_c^+ production. If the rescattering probability for the Λ_c^+ channels, which depends on the kinematics of the scattering hadrons, is similar to the $\Lambda_b^0 \to P_c^+\,K^-$ branching ratio, then P_c^+ production via hadronic rescattering and Λ_b^0 decays is expected to have roughly similar rates.

3.1 Differential cross-sections

The cross sections for tetraquark and pentaquark production at the LHC with $\sqrt{s} = 13$ TeV are provided differentially in $p_\perp$ and rapidity by Figure III.3 and Figure III.4, respectively. The cross sections are separated by rescattering channel, where the first four channels of Figure III.2 are shown. For the $\chi_{c1}(3872)$, only the dominant $D^0\,\bar{D}^{*0}$ channel is given although both $J/\psi\,\omega$ and $J/\psi\,\rho^0$ channels do contribute, but at $\mathcal{O}(10^{-2})$ the rate of $D^0\,\bar{D}^{*0}$ production. For the $P_c^+(4312)$ and $P_c^+(4440)$ resonances the $\Lambda_c^+\,\bar{D}^{*0}$ channel is the leading production mechanism, while the $\Sigma_c^*\,\bar{D}^0$ channel is the dominant channel for $P_c^+(4457)$ production. The $\Sigma_c^+\,\bar{D}^0$ channel contributes to $P_c^+(4440)$ and $P_c^+(4457)$ production for both models, but is always subleading. The channels with light baryons do not significantly contribute to pentaquark production except the $p\,\chi_{c0}$ channel for the model 2 $P_c^+(4440)$ state. Note that the $p\,J/\psi$ discovery channel is not relevant for P_c^+ production.

The $p_\perp$ distributions of both the $\chi_{c1}(3872)$ and P_c^+ states peak near 2 GeV. However, the $p\,\chi_{c0}$ channel is significantly softer than the other channels, peaking near 1 GeV. It is important to note that the production of the χ_{c0} in Pythia is via a hard NRQCD matrix element, *e.g.* $g\,g \to g\,\chi_{c0}$, where the low $p_\perp$ divergence has been regulated with a $p_\perp$ damping term. This is in contrast to the other hadrons, which are produced directly from the hadronization process. The rapidity distribution for the $\chi_{c1}(3872)$ peaks centrally as do the P_c^+ rapidity distributions, although the P_c^+ distributions are slightly broader.

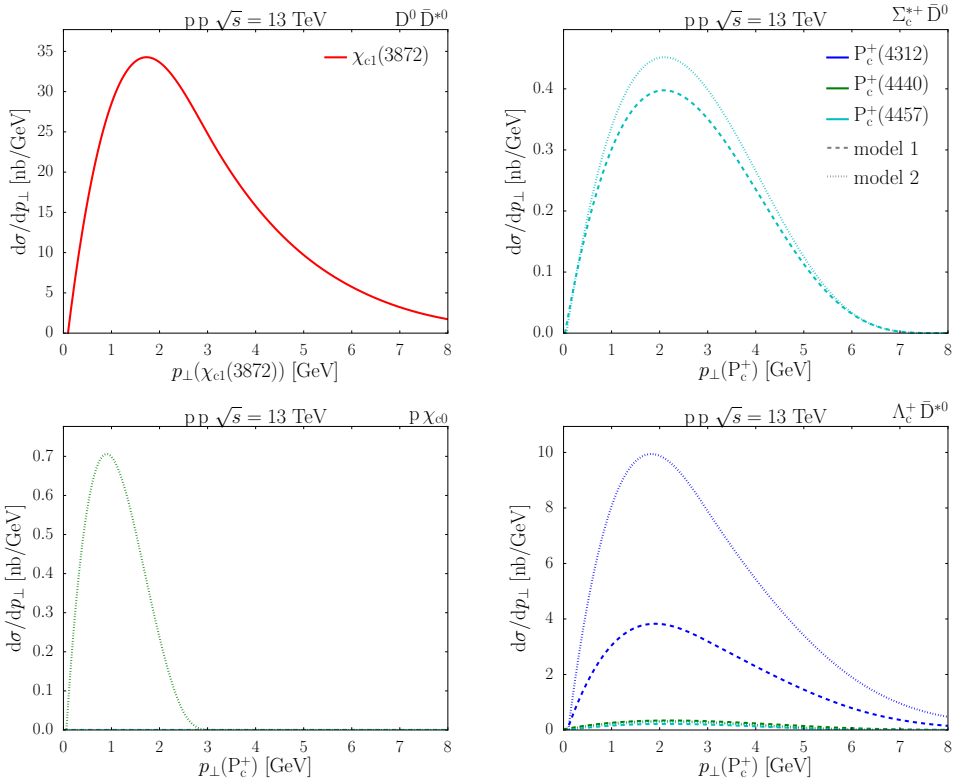


Figure III.3: Tetraquark and pentaquark resonance hadronic-rescattering cross sections, differential in resonance $p_\perp$, produced in p p collisions at $\sqrt{s} = 13$ TeV.

The total hadronic-rescattering cross sections for the pentaquarks are compared to the expected cross section from Λ_b^0 decays in Figure III.5. In general, the $p_\perp$ distributions between hadronic rescattering and Λ_b^0 decays are similar, and again tend to peak around 2 GeV. For the model 2 $\mathrm{P}_c^+(4440)$ state, the $\mathrm{p}\,\chi_{c0}$ channel contributes at the same level as the $\Lambda_c^+\,\bar{\mathrm{D}}^{*0}$ channel. This results in the softer $\mathrm{P}_c^+(4440)$ $p_\perp$ spectrum of model 2 in comparison to model 1. The rapidity distributions for P_c^+ states produced from Λ_b^0 decays are more central than for those produced in rescattering. For the former, this primarily depends upon the rapidity of the parent Λ_b^0 baryons, which in turn depend upon the jets from which they are produced. For the latter, the distributions are broader, and there is a rapidity dip around zero for the $\mathrm{P}_c^+(4457)$ state.

This dip is particularly pronounced in channels with lighter hadrons, *e.g.* $\mathrm{p}\,\pi^0$, whose total contribution to pentaquark formation is very small overall. Similar dips also appear in the rapidity spectrum of $\mathrm{p}\,\pi^0$ rescatterings with invariant masses around the pentaquark masses, regardless of whether pentaquarks are actually formed. Hence, this is a general

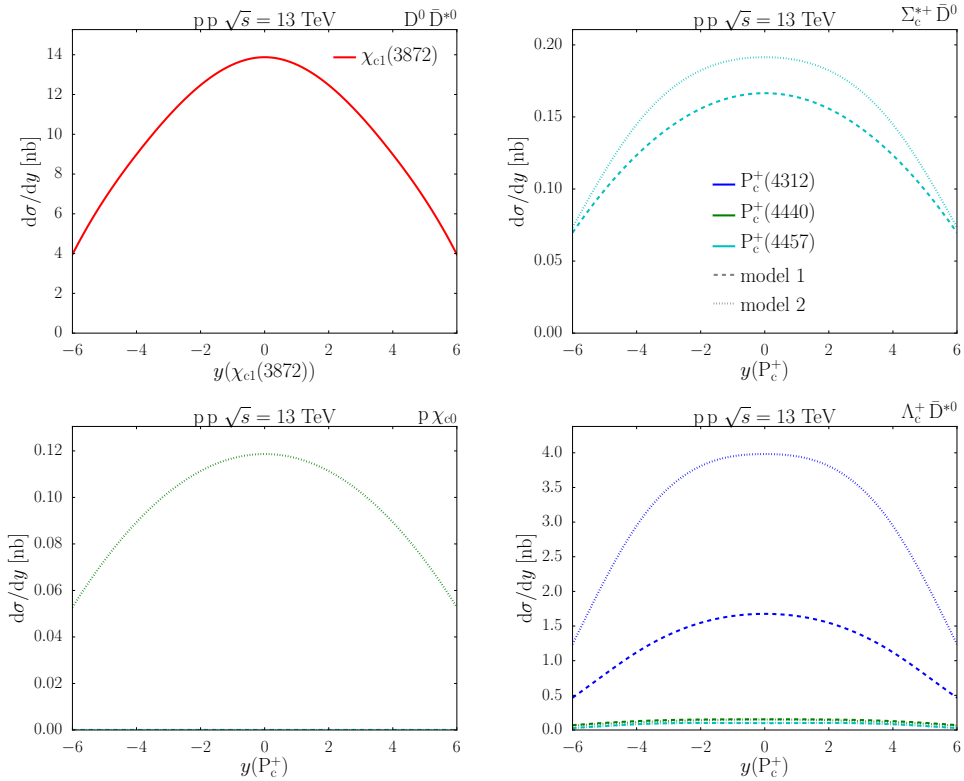


Figure III.4: Tetraquark and pentaquark resonance hadronic-rescattering cross sections, differential in resonance y, produced in p p collisions at $\sqrt{s} = 13$ TeV.

feature of the rescattering framework, and not specific to pentaquarks. Light hadrons must have a more pronounced difference in momenta in order to reach the invariant masses necessary for pentaquark formation. Consequently, the rapidity difference between the rescattering particles should be non-zero, although this still does not fully explain why the total momentum also has non-zero rapidity. Since this is a matter of rescattering in general, and has only a small effect on exotic hadrons, a more in-depth study is left for the future.

The cross sections for P_c^+ from rescattering are generally similar between the two models, which is expected since the dominant partial widths remain similar between the models. The cross section from Λ_b^0 decays differs significantly between the two models, however. This is because the J/ψ partial widths, used in determining $\mathcal{B}(\Lambda_b^0 \to \mathrm{P}_c^+\,\mathrm{K}^-)$, differ by $\mathcal{O}(10)$ to $\mathcal{O}(100)$. This has important experimental implications. Measuring the Λ_b^0 cross section in conjunction with the $\Lambda_b^0 \to \mathrm{P}_c^+\,\mathrm{K}^-$ production can help separate molecular state models, with expected differences as large as $\mathcal{O}(100)$.

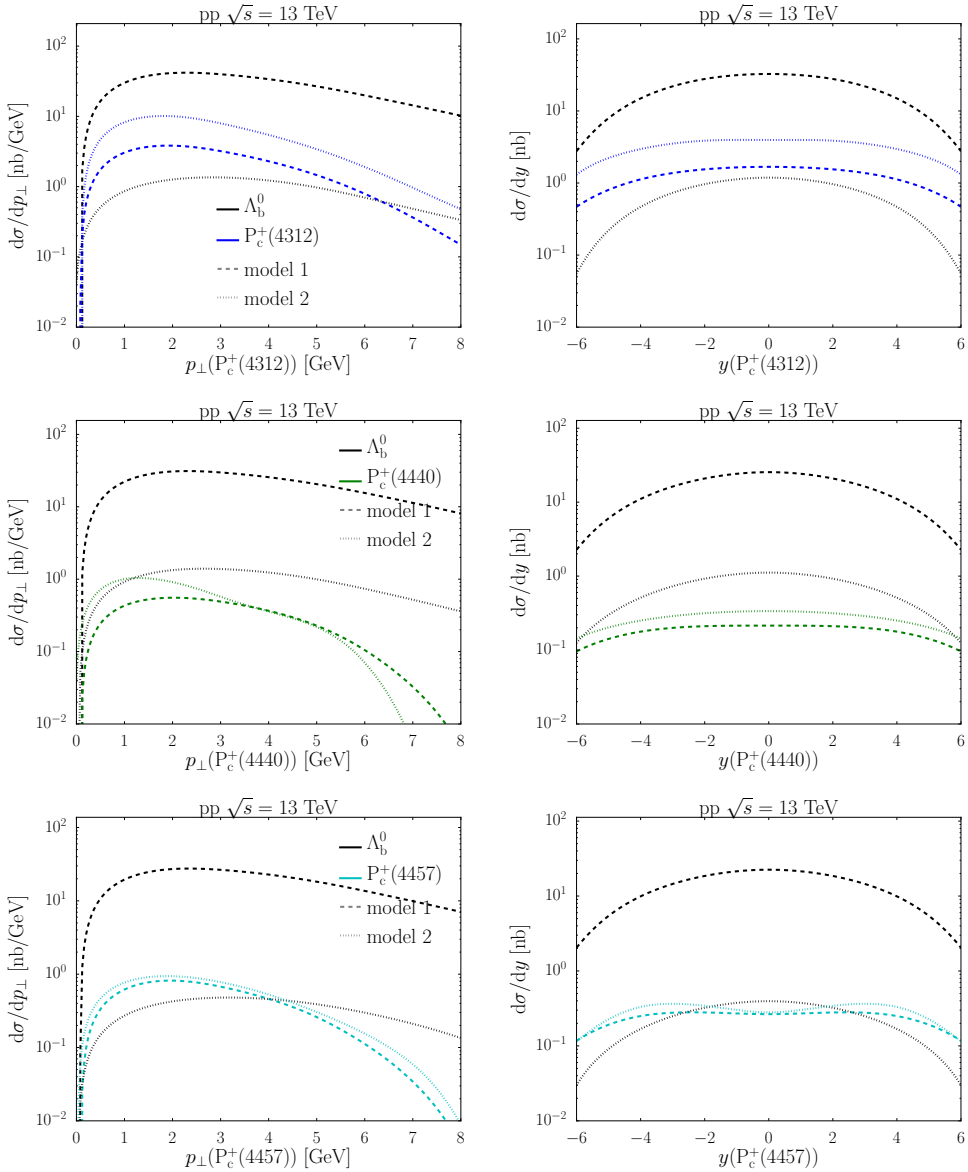


Figure III.5: Total hadronic-rescattering cross-section for (dashed) model 1 and (dotted) model 2 pentaquarks, compared to pentaquark production from (black) Λ_b^0 decays.

3.2 Tetraquarks with LHCb

The fiducial cross section for $\chi_{c1}(3872)$ production in p p collisions at a centre-of-mass energy of $\sqrt{s} = 7$ TeV has been measured by LHCb [58] to be,

$$\sigma_{\text{LHCb}}(\text{p p} \to \chi_{c1}(3872)[\to \text{J}/\psi\, \pi^+\, \pi^-] + X) = 5.4 \pm 1.3 \pm 0.8 \text{ nb}$$

where the $\chi_{c1}(3872) \to J/\psi\,\rho^0$ final state with a $\rho^0 \to \pi^+\,\pi^-$ decay has been used. For this fiducial cross section, the pseudorapidity of the $\chi_{c1}(3872)$ is required to be in the range $2.0 \leq \eta \leq 4.5$, and its momentum must be in the range $5 \leq p_\perp \leq 20$ GeV. These requirements ensure efficient detection of the final state and help minimise systematic uncertainties due to particle reconstruction inefficiencies. This cross section does not separate the prompt $\chi_{c1}(3872)$ production from feed-down production, where the $\chi_{c1}(3872)$ is produced from the decay of a heavier hadron. Indeed, $\chi_{c1}(3872)$ production from B-hadron decays is expected to be sizeable.

The predicted cross section for $\chi_{c1}(3872)$ production from hadronic rescattering in $p\,p$ collisions at $\sqrt{s} = 7$ TeV is,

$$\sigma_{\text{rescatter}}(p\,p \to \chi_{c1}(3872)[\to J/\psi\,\pi^+\,\pi^-] + X) = 0.04 \text{ nb}$$

where the uncertainty on this cross section will have contributions from the uncertainty of the $D^0/\bar{D}^0$ and $D^{*0}/\bar{D}^{*0}$ meson fluxes, and the uncertainty of the estimated hadronic rescattering cross sections. The former uncertainty depends upon the tuning of Pythia used, while the latter depends not only on the modelling of the $\chi_{c1}(3872)$ line-shape but also the partial widths for each rescattering channel. Both of these uncertainties are difficult to quantify and so the hadronic rescattering cross section is quoted here without uncertainty, with the explicit understanding that the uncertainty may be large.

The predicted hadronic-rescattering cross section is not larger than the measured total cross section, which lends some credence to this hadronic-rescattering model. However, the hadronic-rescattering cross section is $\mathcal{O}(100)$ times less than the measured cross section, indicating that hadronic rescattering is expected to provide a negligible contribution to $\chi_{c1}(3872)$ production. Utilising lifetime measurement capabilities, future LHCb measurements could separate the $\chi_{c1}(3872)$ cross-section into prompt and feed-down production, allowing for a direct comparison with this prediction. This could help determine how best to model prompt $\chi_{c1}(3872)$ formation, whether from direct NRQCD calculations, parton showers, hadronization, hadronic rescattering, *etc.*

3.3 Pentaquarks with LHCb

While pentaquark production has been unambiguously observed by LHCb using the exclusive $\Lambda_b^0 \to p\,J/\psi\,K^-$ decay, no P_c^+ cross sections from Λ_b^0 decays or otherwise, have been measured. To fully understand the nature of the observed pentaquark states, these measurements are necessary, including separate cross-section measurements of prompt and feed-down pentaquark production. Depending upon the expected rate of prompt pentaquark production, this may be challenging, as there can be large combinatorial backgrounds when considering a prompt pentaquark signal. The displaced vertex, K^- in the final state, and

Λ_b^0 mass constraint are all no longer available when searching for prompt pentaquark candidates. Consequently, a clean and fully reconstructed final state of charged particles is preferred to reduce these backgrounds.

For the model 1 $P_c^+(4312)$, the largest branching fraction for an exclusive charged final state is $\mathcal{O}(10^{-4})$ with at least five charged particles, including two or more $\pi^\pm$. The combinatorial backgrounds for such a final state will be large, and the high multiplicity will also result in relatively low momentum charged particles that are difficult to reconstruct. The $p\,\eta_c[\to K^+\,K^-]$ final state is expected to have a branching ratio of $\mathcal{O}(10^{-5})$, while the $p\,J/\psi[\to \mu^+\,\mu^-]$ branching ratio will be $\mathcal{O}(10^{-6})$. For the model 1 $P_c^+(4440)$, the branching ratio for the $p\,J/\psi[\to \mu^+\,\mu^-]$ final state is $\mathcal{O}(10^{-4})$, where all fully charged final states with higher branching ratios up to $\mathcal{O}(10^{-3})$ have multiplicities of five or larger. The situation for the $P_c^+(4457)$ is similar, but with the relevant branching ratios reduced by an order of magnitude. Consequently, without detailed background and detector simulation, the $P_c^+ \to p\,J/\psi[\to \mu^+\,\mu^-]$ decay still provides a reasonable final state for the model 1 pentaquarks.

The discovery channel of $p\,J/\psi$ is enhanced in model 2 by $\mathcal{O}(10)$ to $\mathcal{O}(100)$ with respect to model 1. For all model 2 pentaquarks, the $p\,J/\psi[\to \mu^+\,\mu^-]$ final state branching ratios are the same order of magnitude as the leading branching ratios, which for the $P_c^+(4440)$ and $P_c^+(4457)$ are in the $p\,\rho^0[\to \pi^+\,\pi^-]$ channel. The $p\,J/\psi[\to \mu^+\,\mu^-]$ branching ratio is $\mathcal{O}(10^{-4})$ for the $P_c^+(4312)$ and $\mathcal{O}(10^{-3})$ for the $P_c^+(4440)$ and $P_c^+(4457)$ states. The $p\,J/\psi[\to e^+\,e^-]$ branching ratios are also the same as the $p\,J/\psi[\to \mu^+\,\mu^-]$ branching ratios, but in the context of LHCb, electron reconstruction and identification is significantly more challenging than for muons.

This study considers using the LHCb detector to measure prompt pentaquark production via hadronic rescattering. The Run 3 LHCb detector [59–64] is a forward arm spectrometer with full particle reconstruction between pseudorapidities of 2 and 5, including a precision vertex detector, a charged particle tracking system, Cherenkov detectors providing particle identification, an electromagnetic calorimeter, and a muon system. Additionally, during Run 3 the LHCb data acquisition system will employ a real-time analysis strategy [63], where the entire detector will be read out and calibrated in real time. This will enable the full reconstruction of pentaquark candidates during online data taking, and minimise possible inefficiencies of data acquisition. The target Run 3 integrated luminosity for LHCb of 15 fb^{-1} at $\sqrt{s} = 13$ TeV is assumed for this study [64].

Selecting the $p\,J/\psi[\to \mu^+\,\mu^-]$ final state, the following LHCb fiducial requirements are made: the muons and protons must have $2 < \eta < 5$; the muons must have $p_\perp > 0.5$ GeV; and the proton must have $p_\perp > 1$ GeV. Given this fiducial selection, and assuming similar performance to the Run 2 detector [65], the reconstruction efficiency is expected to be near 100%. The background is evaluated as combinations of prompt real $J/\psi[\to \mu^+\,\mu^-]$ decays

with prompt protons. Background from particle mis-identification will also contribute, but is expected to be subleading. Real displaced P_c^+ signals can be separated from the prompt P_c^+ signal with fits to lifetime observables.

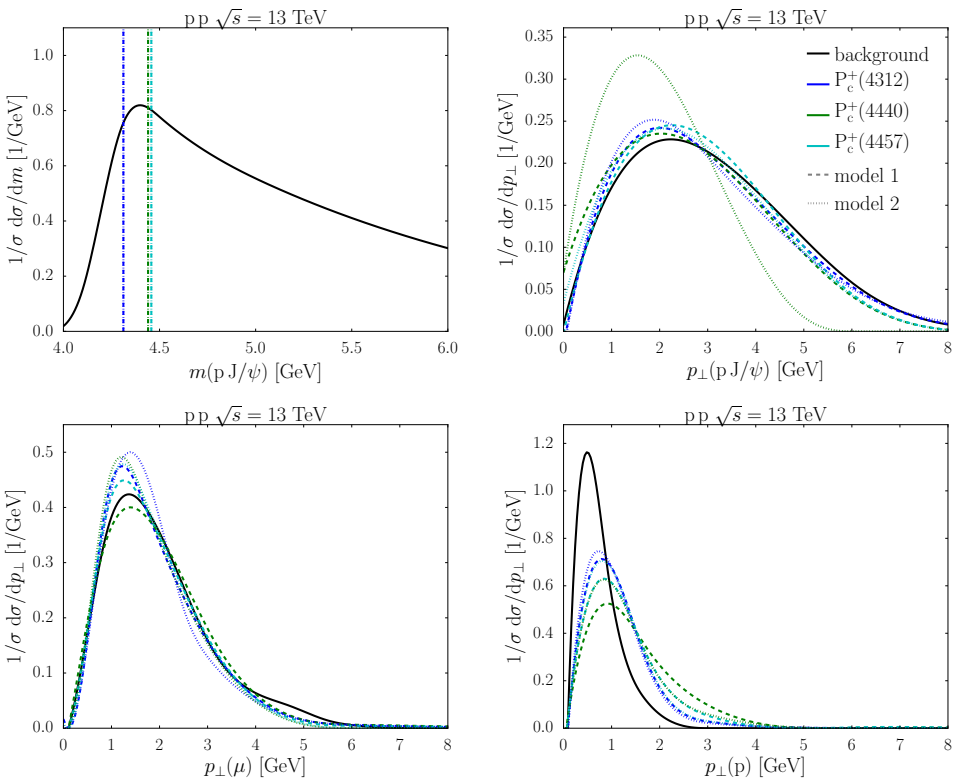


Figure III.6: Normalised cross-sections for $P_c^+ \to \mathrm{p\,J/\psi}$ decays and (solid black) combinatorial background, differential in (top left) final-state mass, (top right) final-state $p_\perp$, (bottom left) muon $p_\perp$, and (bottom right) proton $p_\perp$.

Comparisons of the background distributions with the signal distributions are shown in Figure III.6 with only the final state pseudorapidity requirements in place. All $p_\perp$ distributions are similar between the P_c^+ states, except for the model 2 $P_c^+(4440)$ state, due to the contribution of the $\mathrm{p}\,\chi_{c0}$ channel. The muon $p_\perp$ is slightly softer for the signal than for the background, while the proton $p_\perp$ is slightly harder for the signal than the background. However, these differences are not sufficiently pronounced to provide any kinematic discrimination. Further separation of signal and background may be possible, but would require detailed detector simulation more suitable within an experimental context. The P_c^+ candidate mass distribution peaks near the nominal P_c^+ masses, which complicates resonance fitting. Increasing the $p_\perp$ requirement on the proton could flatten this distribution, but will reduce the signal selection efficiency.

Table III.4: Expected number of reconstructed prompt background and signal P_c^+ candidates by the LHCb detector during Run 3.

	bkg	$P_c^+(4312)$	$P_c^+(4440)$	$P_c^+(4457)$
model 1	2×10^8	2×10^2	2×10^2	1×10^2
model 2		1×10^4	5×10^3	8×10^3

The LHCb $\mu^+\,\mu^-$ mass resolution is roughly 0.4% of the dimuon mass [65]. Even assuming twice the resolution for the three-body P_c^+ candidate with the inclusion of the proton, this is well below the natural width of the P_c^+ states. Consequently, a mass window of roughly six times the P_c^+ total width is used when determining the number of expected background candidates. The expected number of signal and background candidates is given in Table III.4. The mass window of $4.4 < m < 4.5$ GeV is considered for the background when comparing to all three P_c^+ states. The number of background candidates will vary slightly given different P_c^+ pole masses, but for the purposes of this study this approximation is sufficient. From Table III.4 it is clear that prompt P_c^+ production from hadronic rescattering, given the models considered here, will not be observable by LHCb. However, even limits on prompt P_c^+ production will still provide a valuable tool in understanding the nature of the observed pentaquark states.

4 Conclusion

The hadronic rescattering framework in Pythia has been modified to allow for the production of arbitrary hadronic resonances, with an emphasis placed on the production of exotic hadrons that may be molecular states. The relevant code will be published in an upcoming Pythia release. The production cross sections for the tetraquark candidate $\chi_{c1}(3872)$, and the pentaquark candidates $P_c^+(4312)$, $P_c^+(4440)$, and $P_c^+(4457)$, have been calculated for $\sqrt{s} = 13$ TeV p p collisions at the LHC. Using this implementation of Pythia, hadronic rescattering predictions could also be made for the additional exotic states discovered at the LHC. The hadronic rescattering cross section for $\chi_{c1}(3872)$ production at $\sqrt{s} = 7$ TeV in p p collisions is compared to the inclusive $\chi_{c1}(3872)$ cross-section measurement by LHCb and found to contribute at a 1% level. Finally, the expected number of prompt P_c^+ candidates from hadronic rescattering observed by LHCb during Run 3, using the exclusive final state of $p\,J/\psi[\to\ \mu^+\,\mu^-]$, is estimated and found to be significantly smaller than the estimated prompt background. However, cross-section measurements of the P_c^+ candidates, separated into prompt and displaced contributions, can still differentiate between predicted molecular models of these pentaquarks.

Acknowledgements

The authors would like to thank Stephen Mrenna for initial discussions about modelling pentaquark formation and Torbjörn Sjöstrand for further general discussion. PI is supported by the U.S. National Science Foundation grant OAC-2103889. MU is supported in part by the Swedish Research Council, contract number 2016-05996, and in part by the MCnetITN3 H2020 Marie Curie Innovative Training Network, grant agreement 722104.

References

[1] M. Gell-Mann, "A Schematic Model of Baryons and Mesons," *Phys. Lett.* **8** (1964) 214–215.

[2] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 2*. 2, 1964.

[3] R. L. Jaffe, "Multi-Quark Hadrons. 1. The Phenomenology of (2 Quark 2 anti-Quark) Mesons," *Phys. Rev. D* **15** (1977) 267.

[4] D. Strottman, "Multi - Quark Baryons and the MIT Bag Model," *Phys. Rev. D* **20** (1979) 748–767.

[5] H. J. Lipkin, "New Possibilities for Exotic Hadrons: Anticharmed Strange Baryons," *Phys. Lett. B* **195** (1987) 484–488.

[6] K. H. Hicks, "On the conundrum of the pentaquark," *Eur. Phys. J. H* **37** (2012) 1–31.

[7] **Belle** Collaboration, S. K. Choi *et. al.*, "Observation of a resonance-like structure in the $pi^{\pm}\psi'$ mass distribution in exclusive $B \to K\pi^{\pm}\psi'$ decays," *Phys. Rev. Lett.* **100** (2008) 142001, 0708.1790.

[8] **Belle** Collaboration, S. K. Choi *et. al.*, "Observation of a narrow charmonium - like state in exclusive B+- —> K+- pi+ pi- J / psi decays," *Phys. Rev. Lett.* **91** (2003) 262001, hep-ex/0309032.

[9] **LHCb** Collaboration, R. Aaij *et. al.*, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays," *Phys. Rev. Lett.* **115** (2015) 072001, 1507.03414.

[10] **LHCb** Collaboration, R. Aaij *et. al.*, "Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$," *Phys. Rev. Lett.* **122** (2019), no. 22 222001, 1904.03947.

[11] **BaBar** Collaboration, B. Aubert *et. al.*, "Study of the $B \to J/\psi K^- \pi^+ \pi^-$ decay and measurement of the $B \to X(3872) K^-$ branching fraction," *Phys. Rev. D* **71** (2005) 071103, hep-ex/0406022.

[12] **LHCb** Collaboration, R. Aaij *et. al.*, "Determination of the X(3872) meson quantum numbers," *Phys. Rev. Lett.* **110** (2013) 222001, 1302.6269.

[13] **LHCb** Collaboration, R. Aaij *et. al.*, "Quantum numbers of the $X(3872)$ state and orbital angular momentum in its $\rho^0 J\psi$ decay," *Phys. Rev. D* **92** (2015), no. 1 011102, 1504.06339.

[14] N. A. Tornqvist, "Isospin breaking of the narrow charmonium state of Belle at 3872-MeV as a deuson," *Phys. Lett. B* **590** (2004) 209–215, hep-ph/0402237.

[15] L. Maiani, F. Piccinini, A. D. Polosa, and V. Riquer, "Diquark-antidiquarks with hidden or open charm and the nature of X(3872)," *Phys. Rev. D* **71** (2005) 014028, hep-ph/0412098.

[16] **BaBar** Collaboration, B. Aubert *et. al.*, "Search for a charged partner of the X(3872) in the B meson decay $B \to X^- K$, $X^- \to J/\psi \pi^- \pi^0$," *Phys. Rev. D* **71** (2005) 031501, hep-ex/0412051.

[17] L. Maiani, A. D. Polosa, and V. Riquer, "A Theory of X and Z Multiquark Resonances," *Phys. Lett. B* **778** (2018) 247–251, 1712.05296.

[18] C. Bignamini, B. Grinstein, F. Piccinini, A. D. Polosa, and C. Sabelli, "Is the X(3872) Production Cross Section at Tevatron Compatible with a Hadron Molecule Interpretation?," *Phys. Rev. Lett.* **103** (2009) 162001, 0906.0882.

[19] P. Artoisenet and E. Braaten, "Production of the X(3872) at the Tevatron and the LHC," *Phys. Rev. D* **81** (2010) 114018, 0911.2016.

[20] F.-K. Guo, U.-G. Meißner, W. Wang, and Z. Yang, "Production of charm-strange hadronic molecules at the LHC," *JHEP* **05** (2014) 138, 1403.4032.

[21] M. Albaladejo, F.-K. Guo, C. Hanhart, U.-G. Meißner, J. Nieves, A. Nogga, and Z. Yang, "Note on X(3872) production at hadron colliders and its molecular structure," *Chin. Phys. C* **41** (2017), no. 12 121001, 1709.09101.

[22] R. Chen, X. Liu, X.-Q. Li, and S.-L. Zhu, "Identifying exotic hidden-charm pentaquarks," *Phys. Rev. Lett.* **115** (2015), no. 13 132002, 1507.03704.

[23] H.-X. Chen, W. Chen, X. Liu, T. G. Steele, and S.-L. Zhu, "Towards exotic hidden-charm pentaquarks in QCD," *Phys. Rev. Lett.* **115** (2015), no. 17 172001, 1507.03717.

[24] J. He, "$\bar{D}\Sigma_c^*$ and $\bar{D}^*\Sigma_c$ interactions and the LHCb hidden-charmed pentaquarks," *Phys. Lett. B* **753** (2016) 547–551, 1507.05200.

[25] H. Huang, C. Deng, J. Ping, and F. Wang, "Possible pentaquarks with heavy quarks," *Eur. Phys. J. C* **76** (2016), no. 11 624, 1510.04648.

[26] L. Roca and E. Oset, "On the hidden charm pentaquarks in $\Lambda_b \to J/\psi K^- p$ decay," *Eur. Phys. J. C* **76** (2016), no. 11 591, 1602.06791.

[27] Q.-F. Lü and Y.-B. Dong, "Strong decay mode $J/\psi p$ of hidden charm pentaquark states $P_c^+(4380)$ and $P_c^+(4450)$ in $\Sigma_c\bar{D}^*$ molecular scenario," *Phys. Rev. D* **93** (2016), no. 7 074020, 1603.00559.

[28] Y. Shimizu, D. Suenaga, and M. Harada, "Coupled channel analysis of molecule picture of $P_c(4380)$," *Phys. Rev. D* **93** (2016), no. 11 114003, 1603.02376.

[29] C.-W. Shen, F.-K. Guo, J.-J. Xie, and B.-S. Zou, "Disentangling the hadronic molecule nature of the $P_c(4380)$ pentaquark-like structure," *Nucl. Phys. A* **954** (2016) 393–405, 1603.04672.

[30] Y. Yamaguchi, A. Giachino, A. Hosaka, E. Santopinto, S. Takeuchi, and M. Takizawa, "Hidden-charm and bottom meson-baryon molecules coupled with five-quark states," *Phys. Rev. D* **96** (2017), no. 11 114031, 1709.00819.

[31] S. Dubynskiy and M. B. Voloshin, "Hadro-Charmonium," *Phys. Lett. B* **666** (2008) 344–346, 0803.2224.

[32] M. Karliner and B. R. Webber, "Coalescence model for Theta(c) pentaquark formation," *JHEP* **12** (2004) 045, hep-ph/0409121.

[33] P. Ling, X.-H. Dai, M.-L. Du, and Q. Wang, "Prompt production of the hidden charm pentaquarks in the LHC," 2104.11133.

[34] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.* **191** (2015) 159–177, 1410.3012.

[35] T. Sjöstrand and M. Utheim, "A Framework for Hadronic Rescattering in pp Collisions," *Eur. Phys. J. C* **80** (2020), no. 10 907, 2005.05658.

[36] C. Bierlich, T. Sjöstrand, and M. Utheim, "Hadronic rescattering in pA and AA collisions," *Eur. Phys. J. A* **57** (2021), no. 7 227, 2103.09665.

[37] **Particle Data Group** Collaboration, P. A. Zyla *et. al.*, "Review of Particle Physics," *PTEP* **2020** (2020), no. 8 083C01.

[38] **LHCb** Collaboration, R. Aaij *et. al.*, "Study of the lineshape of the $\chi_{c1}(3872)$ state," *Phys. Rev. D* **102** (2020), no. 9 092005, 2005.13419.

[39] **LHCb** Collaboration, R. Aaij *et. al.*, "Study of the $\psi_2(3823)$ and $\chi_{c1}(3872)$ states in $B^+ \to (J\psi\pi^+\pi^-)\, K^+$ decays," *JHEP* **08** (2020) 123, 2005.13422.

[40] Y.-H. Lin and B.-S. Zou, "Strong decays of the latest LHCb pentaquark candidates in hadronic molecule pictures," *Phys. Rev. D* **100** (2019), no. 5 056005, 1908.05309.

[41] E. M. Levin and L. L. Frankfurt, "The Quark hypothesis and relations between cross-sections at high-energies," *JETP Lett.* **2** (1965) 65–70.

[42] H. J. Lipkin, "Quarks for pedestrians," *Phys. Rept.* **8** (1973) 173–268.

[43] **Particle Data Group** Collaboration, M. Tanabashi *et. al.*, "Review of Particle Physics," *Phys. Rev. D* **98** (2018), no. 3 030001.

[44] S. A. Bass *et. al.*, "Microscopic models for ultrarelativistic heavy ion collisions," *Prog. Part. Nucl. Phys.* **41** (1998) 255–369, nucl-th/9803035.

[45] **LHCb** Collaboration, R. Aaij *et. al.*, "Study of the production of Λ_b^0 and $\overline{B}^0$ hadrons in pp collisions and first measurement of the $\Lambda_b^0 \to J/\psi p K^-$ branching fraction," *Chin. Phys. C* **40** (2016), no. 1 011001, 1509.00292.

[46] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes," *Phys. Rev. D* **82** (2010) 074018, 1005.3457.

[47] **ATLAS** Collaboration, M. Aaboud *et. al.*, "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC," *Phys. Rev. Lett.* **117** (2016), no. 18 182002, 1606.02625.

[48] **CMS** Collaboration, A. M. Sirunyan *et. al.*, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV," *JHEP* **07** (2018) 161, 1802.02613.

[49] **LHCb** Collaboration, R. Aaij *et. al.*, "Measurement of the inelastic pp cross-section at a centre-of-mass energy of 13 TeV," *JHEP* **06** (2018) 100, 1803.10974.

[50] **TOTEM** Collaboration, G. Antchev *et. al.*, "First measurement of elastic, inelastic and total cross-section at $\sqrt{s} = 13$ TeV by TOTEM and overview of cross-section data at LHC energies," *Eur. Phys. J. C* **79** (2019), no. 2 103, 1712.06153.

[51] **ATLAS** Collaboration, M. Aaboud *et. al.*, "Measurement of charged-particle distributions sensitive to the underlying event in $\sqrt{s} = 13$ TeV proton-proton collisions with the ATLAS detector at the LHC," *JHEP* **03** (2017) 157, 1701.05390.

[52] **CMS** Collaboration, V. Khachatryan *et. al.*, "Pseudorapidity distribution of charged hadrons in proton-proton collisions at $\sqrt{s}$ = 13 TeV," *Phys. Lett. B* **751** (2015) 143–163, 1507.05915.

[53] **LHCb** Collaboration, R. Aaij *et. al.*, "Measurement of charged particle multiplicities in *pp* collisions at $\sqrt{s} = 7$TeV in the forward region," *Eur. Phys. J. C* **72** (2012) 1947, 1112.4592.

[54] **LHCb** Collaboration, R. Aaij *et. al.*, "Measurement of the forward energy flow in *pp* collisions at $\sqrt{s} = 7$ TeV," *Eur. Phys. J. C* **73** (2013) 2421, 1212.4755.

[55] **LHCb** Collaboration, R. Aaij *et. al.*, "Measurement of prompt hadron production ratios in *pp* collisions at $\sqrt{s} = 0.9$ and 7 TeV," *Eur. Phys. J. C* **72** (2012) 2168, 1206.5160.

[56] **LHCb** Collaboration, R. Aaij *et. al.*, "Measurement of V^0 production ratios in *pp* collisions at $\sqrt{s} = 0.9$ and 7 TeV," *JHEP* **08** (2011) 034, 1107.0882.

[57] **LHCb** Collaboration, R. Aaij *et. al.*, "Measurements of prompt charm production cross-sections in *pp* collisions at $\sqrt{s} = 13$ TeV," *JHEP* **03** (2016) 159, 1510.01707. [Erratum: JHEP 09, 013 (2016), Erratum: JHEP 05, 074 (2017)].

[58] **LHCb** Collaboration, R. Aaij *et. al.*, "Observation of $X(3872)$ production in *pp* collisions at $\sqrt{s} = 7$ TeV," *Eur. Phys. J. C* **72** (2012) 1972, 1112.5310.

[59] **LHCb** Collaboration, I. Bediaga *et. al.*, "Framework TDR for the LHCb Upgrade: Technical Design Report,".

[60] **LHCb** Collaboration, I. Bediaga, "LHCb VELO Upgrade Technical Design Report,".

[61] **LHCb** Collaboration, "LHCb PID Upgrade Technical Design Report,".

[62] **LHCb** Collaboration, "LHCb Tracker Upgrade Technical Design Report,".

[63] "LHCb Trigger and Online Upgrade Technical Design Report,".

[64] **LHCb** Collaboration, R. Aaij *et. al.*, "Physics case for an LHCb Upgrade II - Opportunities in flavour physics, and beyond, in the HL-LHC era," 1808.08865.

[65] **LHCb** Collaboration, R. Aaij *et. al.*, "LHCb Detector Performance," *Int. J. Mod. Phys. A* **30** (2015), no. 07 1530022, 1412.6352.

Paper IV

IV

Hadron Interactions for Arbitrary Energies and Species, with Applications to Cosmic Rays

Torbjörn Sjöstrand and Marius Utheim

To be submitted to *Eur. Phys. J. C*,
e-Print: `arXiv:2108.03481` [hep-ph]
MCnet–21–14, LU TP 21–32

Abstract

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

1 Introduction

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

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

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

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

Two significant extensions are introduced to this end. One is a modelling of total, elastic, diffractive and nondiffractive cross sections for the various beam combinations, as needed to describe collision rates also at energies above 10 GeV. The other is parton distribution functions (PDFs) for a wide selection of mesons and baryons, as needed to describe the

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

Key to the understanding of atmospheric cascades is the model for hadronic interactions. Several different ones are used in the community, such as SIBYLL [9–12], QGSJET [13–15], DMPJET [16, 17], VENUS/EPOS [18–20], and HDPM (described in Ref. [21]). It is in this category that Pythia could offer an alternative model, constructed completely independently of either of the other ones, and therefore with the possibility to offer interesting cross-checks. In some respects it is likely to be more sophisticated than some of the models above, *eg.* by being able to handle a large range of beam particles almost from the threshold to the highest possible energies, with semi-perturbative interactions tailored to the incoming hadron type. In other respects it is not yet as developed, like a limited handling of nuclear effects and a lack of tuning to relevant data.

Neither of these programs can describe the important electromagnetic cascades, which instead typically are delegated to EGS [22]. At low energies GHEISHA [23] is often used for nuclear effects, with ISOBAR (described in Ref. [21]) and UrQMD [24] as alternatives. Generally a typical full simulation requires many components to be combined, under the control of a framework that does the propagation of particles through the atmosphere, taking into account *eg.* the atmospheric density variation and the bending of charged particles by the earth magnetic field. Two well-known examples of such codes are CORSIKA [21] and AIRES [25, 26]. Interestingly for us, the new CORSIKA 8 [27, 28] framework is written in C++, like Pythia 8 but unlike some of the other hadronic interaction models, and Pythia 8 is already interfaced to handle particle decays, so a further integration is a possibility.

One should also mention that an alternative to Monte Carlo simulation of cascades is to construct a numerical simulation from the cascade evolution equations, examples being SENECA [22] and MCeq [22]. Also in these cases the hadronic interaction models can provide valuable input. Angantyr has in fact already been used to this end, to describe $\mathrm{p}/\pi/\mathrm{K}$ interactions with nuclei [29].

Another application of hadronic cascades is in detector simulations with programs such as FLUKA [30] and GEANT [31–34], which have also been used for cascades in the atmosphere, see *eg.* [35–37]. GEANT4 depends on external frameworks for simulating collisions, like CORSIKA 8, and has been explicitly designed with an object-oriented architecture that allows users to insert their own physics implementations, one of the current possibilities

being Pythia 6 [1]. One central difference is that the medium is much denser in a detector, so particles propagate shorter distances before interacting. Hence, some particles that are too short-lived to interact in the atmosphere can do so in detector simulations, *eg.* D, B, Λ_c^+ and Λ_b^0. For the rest of this article we will focus on the atmospheric case, but we still implement all hadronic interactions relevant for either medium.

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

2 Cross sections and parton distributions

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

2.1 New total cross sections

The description of cross sections depends on the collision energy. At low energies various kinds of threshold phenomena and resonance contributions play a key role, and these can differ appreciably depending on the incoming hadron species. At high energies a more

smooth behaviour is expected, where the dominating mechanism of pomeron exchange should give common traits in all hadronic cross sections [38–41].

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

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

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

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

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

At the other extreme then lies models intended to describe high-energy cross sections. Here $\mathrm{pp}/\mathrm{p\bar{p}}$ collisions are central, given the access to data over a wide energy range, and the need to interpret this data. A few such models have been implemented in Pythia [42], giving the possibility of comparisons. Fewer models are available for diffractive topologies than for the total and elastic cross sections.

For the purposes of this study we will concentrate on the SaS/DL option, not necessarily because it is the best one for $\mathrm{pp}/\mathrm{p\bar{p}}$ but because we have the tools to extend it to the necessary range of collision processes in a reasonably consistent manner. The starting point is the Donnachie–Landshoff modelling of the total cross section [43]. In it, a common ansatz

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

is used for the collisions between any pair of hadrons A and B. Here s is the squared CM energy, divided by 1 GeV^2 to make it dimensionless. The terms s^{ϵ} and $s^{-\eta}$ are assumed to arise from pomeron and reggeon exchange, respectively, with tuned universal values $\epsilon = 0.0808$ and $\eta = 0.4525$. The X^{AB} and Y^{AB}, finally, are process-specific. $X^{\bar{A}B} = X^{AB}$ since the pomeron is charge-even, whereas generally $Y^{\bar{A}B} \neq Y^{AB}$, which can be viewed as a

Table IV.1: Coefficients X^{AB} and Y^{AB}, in units of mb, in eq. (IV.2) for various beam combinations. First section is from DL [43], second from SaS [47] and the rest are new for this study.

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

consequence of having one charge-even and one charge-odd reggeon. Recent experimental studies [44, 45] have shown that the high-energy picture should be complemented by a charge-odd odderon [46] contribution, but as of yet there is no evidence that such effects have a major impact on total cross sections.

In the context of γp and $\gamma\gamma$ studies, the set of possible beam hadrons was extended by Schuler and Sjöstrand (SaS) to cover vector meson collisions [47, 48]. Now we have further extended it to cover a range of additional processes on a p/n target, Table IV.1. The extensions have been based on simple considerations, notably the AQM, as outlined in the table. They have to be taken as educated guesses, where the seeming accuracy of numbers is not to be taken literally. For simplicity, collisions with protons and with neutrons are assumed to give the same cross sections, which is consistent with data, so only the former are shown. The reggeon term for ϕ^0p is essentially vanishing, consistent with the OZI rule [49–51], and we assume that this suppression of couplings between light u/d quarks and s quarks extends to c and b. Thus, for baryons, the reggeon Y^{AB} values are assumed pro-

portional to the number of light quarks only, while the AQM of eq. (IV.2) is still used for the pomeron term. Another simplification is that D/B and $\bar{\mathrm{D}}/\bar{\mathrm{B}}$ mesons are assigned the same cross section. Baryons with the same flavour content, or only differing by the relative composition of u and d quarks, are taken to be equivalent, *ie.* $\Lambda\mathrm{p} = \Sigma^{+}\mathrm{p} = \Sigma^{0}\mathrm{p} = \Sigma^{-}\mathrm{p}$.

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

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

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

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

2.2 New partial cross sections

The total cross section can be split into different components

$$\sigma_{\mathrm{tot}} = \sigma_{\mathrm{ND}} + \sigma_{\mathrm{el}} + \sigma_{\mathrm{SD}(XB)} + \sigma_{\mathrm{SD}(AX)} + \sigma_{\mathrm{DD}} + \sigma_{\mathrm{CD}} + \sigma_{\mathrm{exc}} + \sigma_{\mathrm{ann}} + \sigma_{\mathrm{res}} + \ldots\,. \tag{IV.5}$$

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

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

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

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

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

where $\alpha' \approx 0.25$ GeV^{-2} is the slope of the pomeron trajectory and $s_0 = 1/\alpha'$. In the final expression the SaS replacement is made to ensure that $\sigma_{\mathrm{el}}/\sigma_{\mathrm{tot}}$ goes to a constant

below unity at large energies, while offering a reasonable approximation to the logarithmic expression at low energies. The hadronic form factors $b_{A,B}$ are taken to be 1.4 for mesons and 2.3 for baryons, except that mesons made only out of c and b quarks are assumed to be more tightly bound and thus have lower values. As a final comment, note that a simple exponential in t is only a reasonable approximation at small $|t|$, but this is where the bulk of the elastic cross section is. For pp and $\mathrm{p\bar{p}}$ more sophisticated larger-$|t|$ descriptions are available [42].

Also diffractive cross sections are calculated using the SaS ansatz [48, 52]. The differential formulae are integrated numerically for each relevant collision process and the result suitably parametrized, including a special threshold-region ansatz [5]. Of note is that, if the hadronic form factor from pomeron-driven interactions is written as $\beta_{A\mathbb{P}}(t) = \beta_{A\mathbb{P}}(0) \exp(b_A t)$ then, with suitable normalization, $X^{AB} = \beta_{A\mathbb{P}}(0)\, \beta_{B\mathbb{P}}(0)$ in eq. (IV.2). Thus we can define $\beta_{\mathrm{p}\mathbb{P}}(0) = \sqrt{X^{\mathrm{pp}}}$ and other $\beta_{A\mathbb{P}}(0) = X^{A\mathrm{p}}/\beta_{\mathrm{p}\mathbb{P}}(0)$. These numbers enter in the prefactor of single diffractive cross sections, *eg.* $\sigma_{AB\to AX} \propto \beta^2_{A\mathbb{P}}(0)\, \beta_{B\mathbb{P}}(0) = X^{AB}\, \beta_{A\mathbb{P}}(0)$. This relation comes about since the A side scatters (semi)elastically while the B side description is an inclusive one, cf. the optical theorem. In double diffraction $AB \to X_1 X_2$ neither side is elastic and the rate is directly proportional to X^{AB}.

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

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

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

2.3 Hadronic collisions

At low energies the character of an event is driven entirely by nonperturbative processes. In a nondiffractive topology, this can be represented by the exchange of a single gluon, so soft that the momentum transfer can be neglected. The colour exchange leads to two colour octet hadron remnants, however. Each can be split into a colour triplet and a colour antitriplet part, q-$\bar{\mathrm{q}}$ for a meson and q-qq for a baryon. This leads to two (Lund [53]) strings being pulled out, each between the colour of one hadron and the anticolour of the other. In diffraction either a quark or a gluon is kicked out from the diffracted hadron,

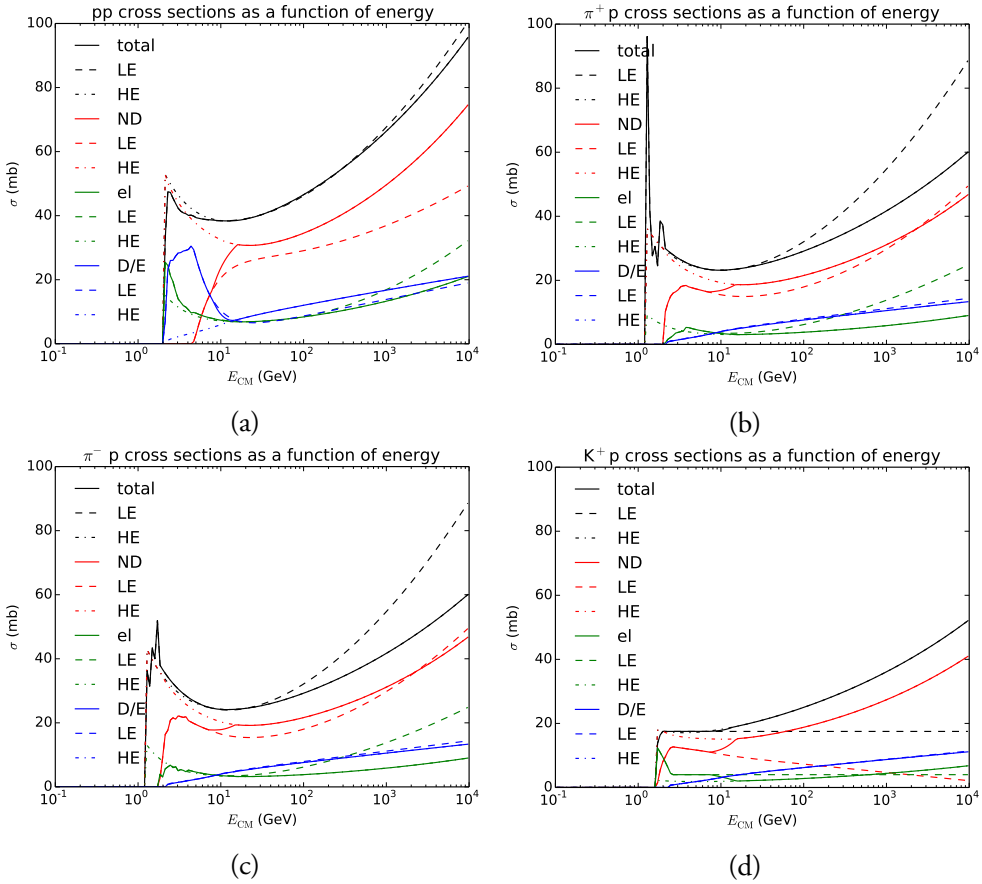


Figure IV.1: Total, nondiffractive (ND), elastic (el) and diffractive/excitation (D/E) cross sections for some common collision processes, (a) pp, (b) π^+p, (c) π^-p and (d) K$^+$p. Full lines show the cross sections actually used, while dashed show the low-energy (LE) and dash-dotted the high-energy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

giving either a straight string or one with a kink at the gluon. Other processes have their own descriptions [5].

At high energies, on the other hand, perturbative processes play a key role. A suitable framework is that of multiparton interactions, MPIs [54, 55]. In it, it is assumed that the composite nature of the hadrons leads to several separate parton–parton interactions, each dressed up with associated parton showers. At first glance the interactions occur independently, but at closer look they are connected by energy–momentum–flavour–colour conservation. Especially the last is nontrivial to model, and requires a special colour reconnection

step. There the total string length is reduced relative to a first assignment where the MPIs are largely decoupled from each other.

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

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

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

$$E_{\mathrm{min}} = E_{\mathrm{min},0} + 2\max(0., m_A - m_p) + 2\max(0., m_B - m_p) , \tag{IV.9}$$

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

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

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

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

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

$$f_{\mathrm{damp}}(p_\perp) = \left(\frac{\alpha_{\mathrm{s}}(p_{\perp 0}^2 + p_\perp^2)}{\alpha_{\mathrm{s}}(p_\perp^2)} \frac{p_\perp^2}{p_{\perp 0}^2 + p_\perp^2}\right)^2 . \tag{IV.11}$$

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

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

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

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

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

MPIs can also occur in high-mass diffraction, and is simulated in Pythia [56], but this is a smaller fraction.

The amount of MPIs in a collision directly impacts the event activity, *eg.* the average charged multiplicity. MPIs have almost exclusively been studied in pp and $\mathrm{p\bar{p}}$ collisions, however, so we have no data to go on when we now want to extend it to all the different collision types listed in Table IV.1. As a guiding principle we assume that $\langle n_{\mathrm{MPI}} \rangle$ should remain roughly constant, *ie.* plausibly hadronic collisions at a given (large) energy have a comparable event activity, irrespective of the hadron types. But we already assumed that total cross sections are lower for mesons than for baryons, and falling for hadrons with an increasing amount of strange, charm or bottom quarks, so naively then eq. (IV.13) would suggest a correspondingly rising $\langle n_{\mathrm{MPI}} \rangle$. There are (at least) two ways to reconcile this.

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

The alternative is to modify the PDFs and to let heavier quarks take a larger fraction of the respective total hadron momentum, such that there are fewer gluons and sea quarks at small x values and therefore a reduced collision rate. (A high-momentum quark will have an enhanced high-$p_\perp$ collision rate, but that is only one parton among many.) This is actually a well-established "folklore", that all long-lived constituents of a hadron must travel at approximately the same velocity for the hadron to stick together. It is a crucial aspect of the "intrinsic charm" hypothesis [57], where a long-lived $\mathrm{c\bar{c}}$ fluctuation in a proton takes a major fraction of the total momentum. In the inverse direction it has also been used to motivate heavy-flavour hadronization [58, 59]. This is the approach we will pursue in the following.

2.4 New parton distribution functions

Most PDF studies have concerned and still concern the proton, not least given the massive influx of HERA and LHC data. Several groups regularly produce steadily improved PDF sets [60–62]. The emphasis of these sets are on physics at high Q^2 and (reasonably) high

x to NLO or NNLO precision. In our study the emphasis instead is on inclusive events, dominated by MPIs at scales around $p_{\perp 0}$, *ie.* a few GeV, and stretching down to low x values. These are regions where NLO/NNLO calculations are notoriously unstable, and LO descriptions are better suited.

Moving away from protons, data is considerably more scarce. There is some for the pion, *eg.* [63–66], a very small amount for the Kaon [67], and nothing beyond that. There has also been some theoretical PDF analyses, based on data and/or model input, like [68–83] for the pion and [70, 73, 84–88] for the Kaon. But again nothing for hadrons beyond that, to the best of our knowledge, which prompts our own work on the topic.

In order to be internally consistent, we have chosen to take the work of Glück, Reya and coworkers as a starting point. The basic idea of their "dynamically generated" distributions is to start the evolution at a very low Q_0 scale, where originally the input was assumed purely valence-quark-like [89]. Over the years both gluon and u/d sea distributions have been introduced to allow reasonable fits to more precise data [90–93], but still with ansätze for the PDF shapes at Q_0 that involve a more manageable number of free parameters than modern high-precision (N)NLO ones do. Their LO fits also work well with the Pythia MPI framework. To be specific, we will use the GRS99 pion [71] as starting point for meson PDFs, and the GJR07 proton one [93] similarly for baryons. Also the GR97 Kaon one [70] will play some role.

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

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

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

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

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

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

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

In our work, we make the ansatz that hadron PDFs can be parameterized on the form given in eq. (IV.14) at the initial scale Q_0, but with $A = B = 0$ since there are no data or

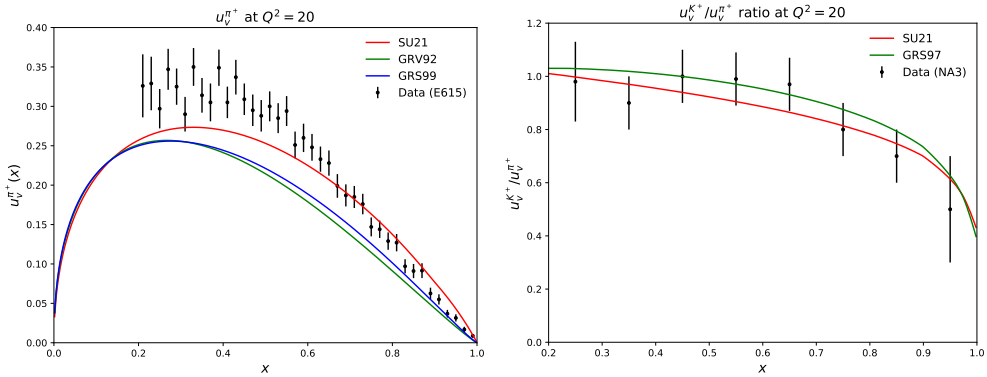


Figure IV.2: (a) Pion PDFs, comparing our simplified form to GRV92 and GRS99, and to data. (b) $\bar{u}_v^{K^-}/\bar{u}_v^{\pi^-}$ ratio, comparing our simplified parameterization on the form given in eq. (IV.14) to the slightly more detailed description of GRS97, and comparing to data. Note that both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

guiding principles to fix them in the generic case. The a and b parameters are allowed to vary with the particular parton and hadron in question, while N is fixed by eq. (IV.16) for valence quarks. The deviations introduced by the $A = B = 0$ assumption are illustrated in Figure IV.2. In Figure IV.2a E615 data [66] are compared with the π PDFs as given by GRV92 [68], by GRS99 [71], and by our simplified description (labeled SU21) where a has been adjusted to give the same $\langle x \rangle$ as for GRS99. In Figure IV.2b the $\bar{u}_v^{K^-}/\bar{u}_v^{\pi^-}$ ratio is compared between data [67], GRS97 [70] and our simplified model. In both cases the model differences are comparable with the uncertainty in data.

To further illustrate the changes introduced by setting $A = B = 0$, Figure IV.3 shows the number and transverse momentum of MPIs for different (a) proton and (b) pion PDFs, with average values as in Table IV.2. In both cases, our simplified SU21 ansatz leads to a shift that is comparable to the difference between the two standard PDFs. Thus we feel confident that our simplified ansatz is sufficient also for other hadrons, where there are neither data nor detailed theory calculations available. Nevertheless, for accuracy, we use the NNPDF2.3 QCD+QED LO distribution function for protons and GRS99 LO for pions in our studies, and the SU21 ansatz only for hadrons beyond that.

Given that there is no solid theory for heavy hadron PDFs, the specific choices of a and b necessarily are heuristic. Our guiding principle is that all quarks should have roughly the same velocity, as already mentioned, and thus heavier quarks must have a larger average momentum fraction $\langle x \rangle$, and a smaller b, while gluons and sea u/d must be softer. The $\langle x \rangle$ choices do not exactly agree with the assumed mass ratios in our AQM ansatz, eq. (IV.1), but are somewhat less uneven than that. This is supported by the Kaon data [67], and also

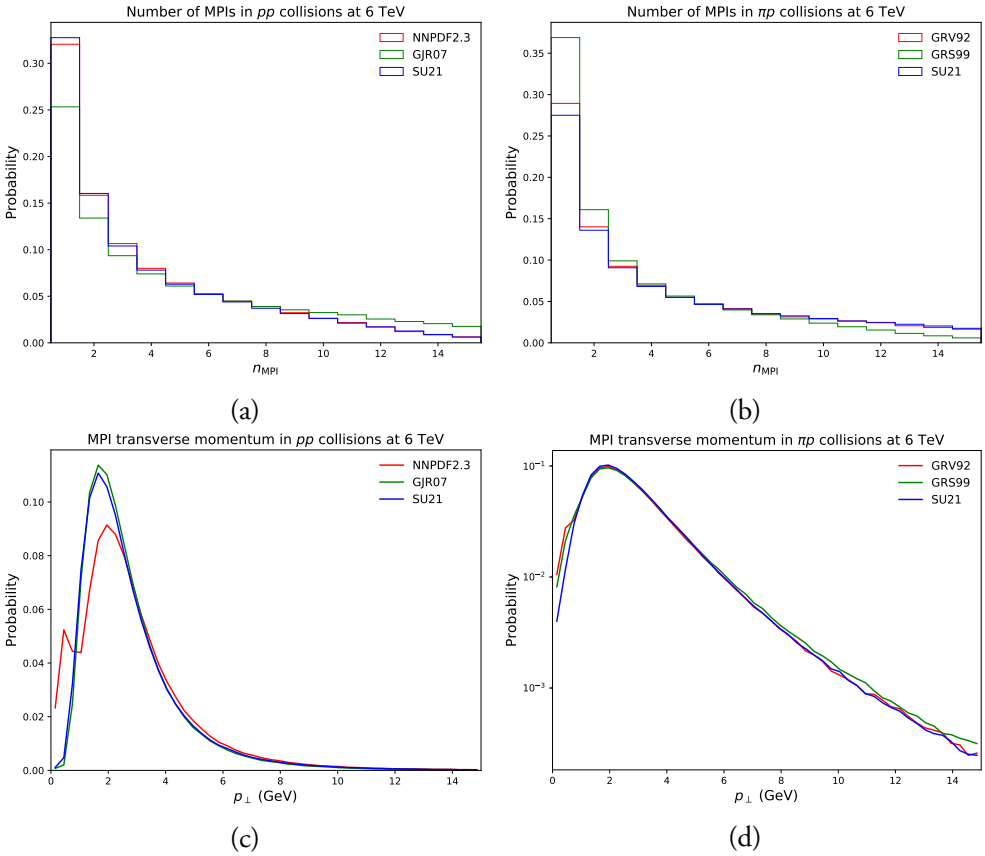


Figure IV.3: The (a,b) number and (c,d) transverse momentum spectrum of MPIs, for (a,c) protons and (b,d) pions. Each PDF has a cutoff and is considered constant below some Q_0, which leads to the bumps at low $p_\perp$, especially noticeable for NNPDF2.3 distribution in (c), whose cutoff is $Q = 0.5$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

by some modelling [57].

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

1. Let the valence quark distributions be given by $Nx^a(1-x)^b$, *ie.* put $A = B = 0$.
2. Choose sensible b and $\langle x \rangle$ values for each valence quark, based on the principles above.
3. Derive a from $\langle x \rangle = \dfrac{\int_0^1 dx\, x f(x)}{\int_0^1 dx\, f(x)} = (a+1)/(a+b+2)$.
4. Derive N to satisfy eq. (IV.16).

Table IV.2: Average number and transverse momentum spectrum for MPIs with different PDFs in pp and πp collisions. The default p PDF in Pythia is NNPDF2.3 QCD+QED LO with $\alpha_S(M_Z) = 0.130$ [94]. This default is used for the proton PDF in the πp collisions.

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

5. For sea and gluon distributions, pick a d and set $f(x) \propto x^d f^{\pi}(x)$ (here with A and B values as for the pion).

6. Rescale the gluon and u/d sea distributions by a common factor to satisfy the momentum sum relation
$$\int_0^1 dx \sum_q x f_q(x) = 1.$$

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

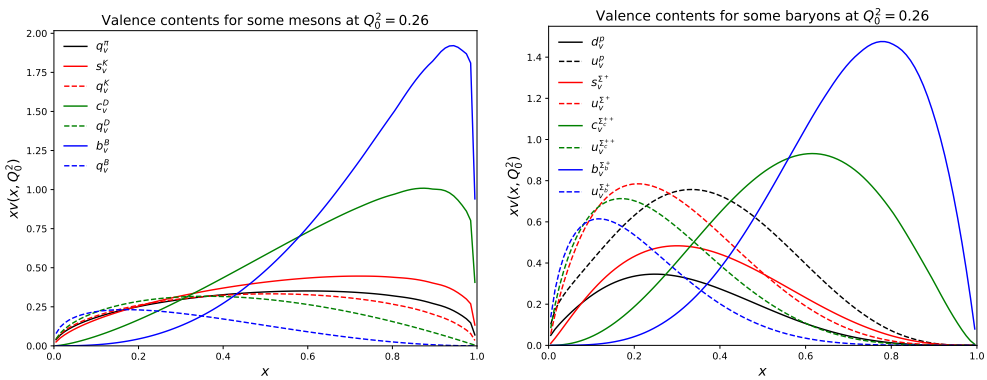


Figure IV.4: Different valence PDFs at the initial scale $Q_0^2 = 0.26$ GeV^2, for (a) π, K, D and B mesons, showing the flavoured valence and the q (= d/u) valence contents; and for (b) uu$q$ baryons for $q =$ d (proton), s (Σ^+), c (Σ_c^{++}) and b (Σ_b^+).

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

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

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

Once the initial state has been set up, the DGLAP equations [95–97] describe the evolution towards higher Q^2 scales. Any number of implementations of these equations exist, both private and public, such as QCDNUM [98], HERAFitter/xFitter [99] and APFEL [100], that in principle should be equivalent. We choose to use QCDNUM since we find it well documented and well suited for our purposes. Nevertheless there are some limitations that we had to circumvent.

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

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

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

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

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

After having studied the PDFs resulting from this procedure, we make one additional ad hoc adjustment for the J/ψ, $\mathrm{B_c}$ and Υ mesons, *ie.* the ones that have exclusively c and b valence content. Using only the procedure outlined so far, the average number of MPIs in interactions involving these particles is much higher than for other hadrons. This comes as no surprise in view of eq. (IV.13); the absence of light quarks makes for a small total and nondiffractive cross section, while the normal evolution allows a non-negligible gluon and sea to evolve right from the low Q_0 scale. But in real life one should expect heavy quarks to have a reduced emission rate of gluons below their mass scale. To compensate for this, increased Q_0^2 scales of 0.6 GeV^2, 0.75 GeV^2 and 1.75 GeV^2 are used for J/ψ, $\mathrm{B_c}$ and Υ, respectively. One could argue that similar shifts should be made for all hadrons containing a c or b quark, but if there are also light valence quarks then there should be some evolution already from small scales. Any mismatch in the emission can then more easily be absorbed in the overall uncertainty of the setup at the Q_0 starting scale.

2.5 The forward region

The fastest particles in the projectile region play a central role for the shower evolution in the atmosphere, so the modelling of this region is a topic of special interest. Traditionally Pythia is more aimed towards the modelling of the central region, and there are known issues in the forward region [101–103]. Briefly put, proton/neutron spectra are softer and pion spectra harder than data. In Sibyll, which normally uses Lund string fragmentation,

this has required a separate dedicated handling of leading-baryon formation [12].

We are not here able to report a final resolution of these issues, but a beginning has been made with two new options, both which modify the fragmentation of a diquark in the beam remnant. The first is to disallow popcorn handling [104], *ie.* the mechanism $q_1q_2 \to q_1\bar{q}_3 + q_2q_3$ whereby a meson can become the leading particle of a jet. (If two valence quarks are kicked out from the proton, which can happen in separate MPIs, the resulting junction topology is unaffected by the popcorn handling.) The second is to set the a and b parameters of the Lund symmetric fragmentation function $f(z) = (1/z)\, z^a \exp(-bm_\perp^2/z)$ separately from those in normal hadronization. There is some support for such a deviation in a few Lund studies [105–107], where it is argued that a drifting-apart of the two quarks of an original dipole indirectly leads to a hardening of the baryon spectrum.

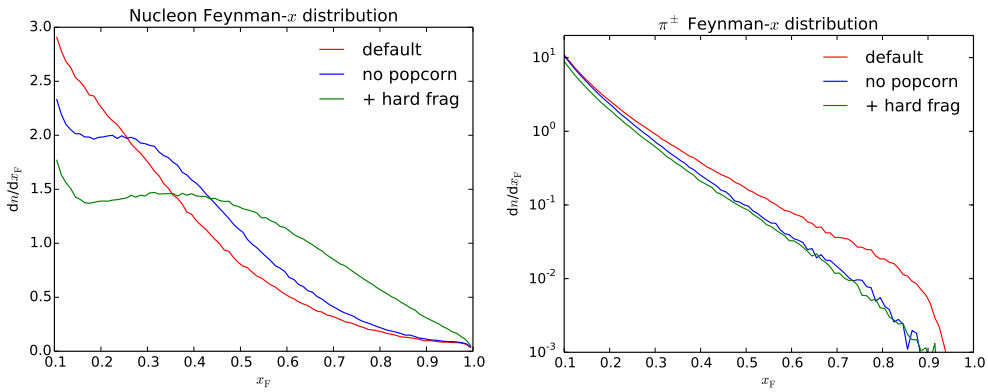


Figure IV.5: Feynman-x spectrum of (a) nucleons p/n and (b) $\pi^\pm$ for 6 TeV pp collisions with inelastic events, where the quasi-elastic side of single diffraction is not considered.

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

Further modifications are likely to be necessary, and tuning studies are in progress [108].

2.6 Technical details

The new "SU21" PDFs will be included in an upcoming release of PYTHIA as LHAPDF-compatible [109] files, using the `lhagrid1` format, as a central grid. The grids go down to $x = 10^{-9}$ and up to $Q = 10^4$ GeV.

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

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

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

3 Event properties and nuclear effects

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

3.1 Hadronic interaction properties

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

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

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

Studying the rapidity spectra closer, we find asymmetries around zero, depending on how different the PDFs of the projectile are from those of the target proton. That is, a projectile with harder PDFs should give a spectrum shifted towards positive rapidities, and vice versa. Harder valence quarks tend to be counteracted by softer gluons and sea quarks, however.

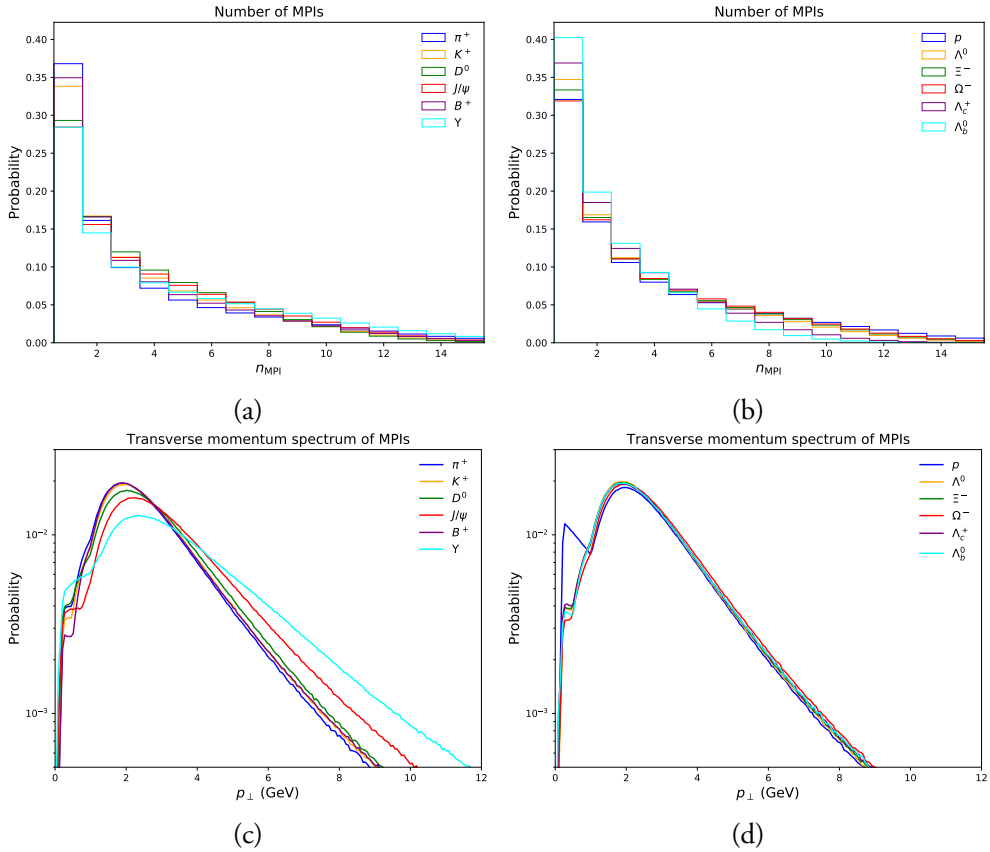


Figure IV.6: The (a,b) number and (c,d) transverse momentum spectrum of MPIs, for a selection of (a,c) meson–proton and (b,d) baryon–proton collisions at 6 TeV. Labels denote the respective hadron beam.

Possible effects also are partly masked by strings being stretched out to the beam remnants, no matter the rapidity of the perturbative subcollision. To better probe larger x values, we also study jet distributions, using the anti-$k_\perp$ algorithm [110, 111] with $R = 0.7$ and $p_{\perp\mathrm{jet}} > 50$ GeV. Some distributions are shown in Figure IV.8, with average values again given in Table IV.4. Indeed asymmetries now are quite visible. We also note that the high-$p_\perp$ jet rate, normalized relative to the total nondiffractive cross section, is enhanced in the cases with harder PDFs, even if this is barely noticeable in the average $p_\perp$ of all hadrons.

3.2 Nuclear collisions with Angantyr

Pythia comes with a built-in model for heavy-ion collisions, Angantyr [4], which describes much pA and AA data quite well. It contains a model for the selection of nuclear

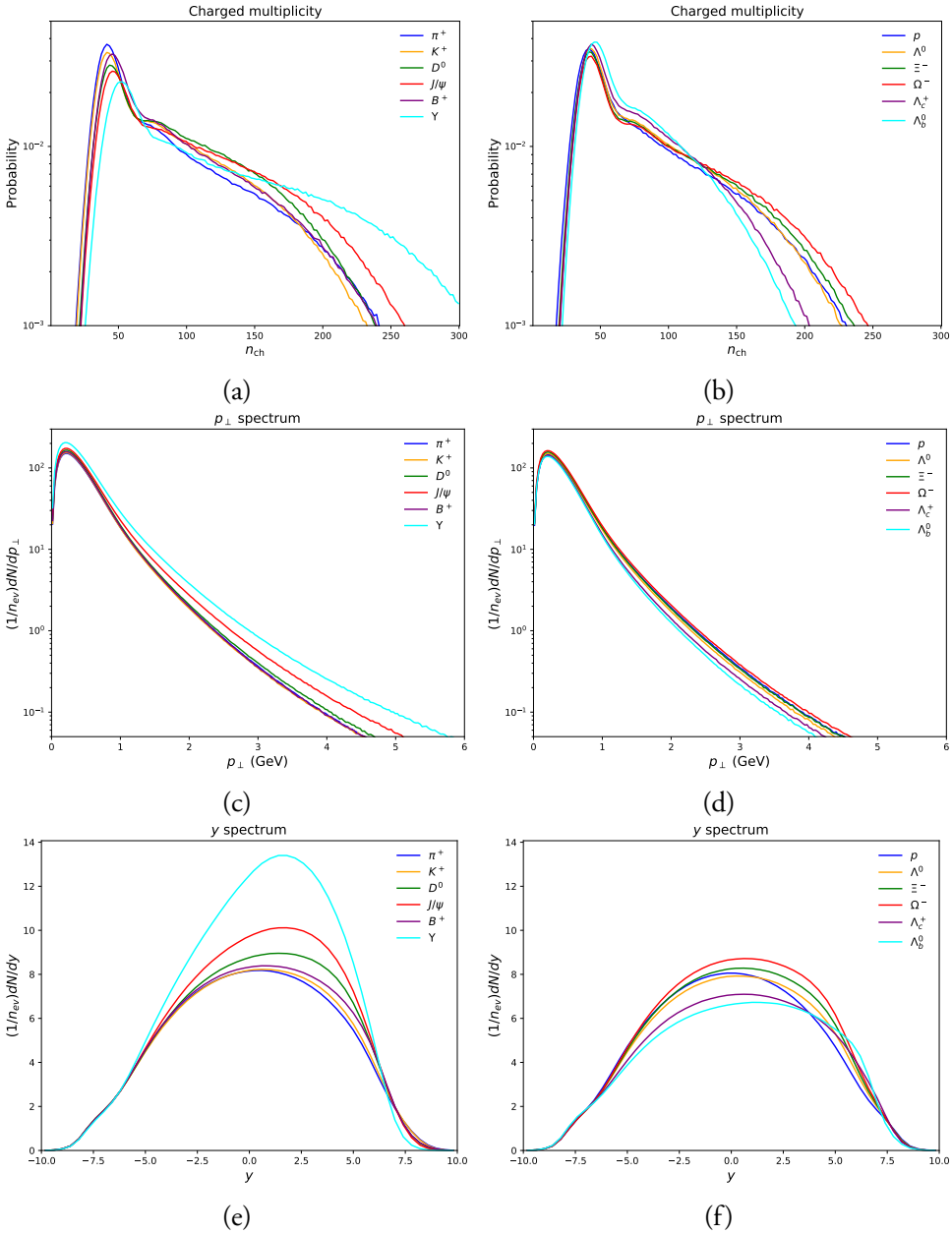


Figure IV.7: Charged-hadron (a,b) multiplicity distributions, (c,d) $p_\perp$ spectra and (e,f) rapidity distributions for a selection of (a,c,e) meson–proton and (b,d,f) baryon–proton collisions. All results at a 6 TeV collision energy, and for nondiffractive events only. Labels denote the respective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

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

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

geometry and impact parameter of collisions. In the Glauber formalism [112] the nucleons are assumed to travel along straight lines, and a binary nucleon–nucleon subcollision can result anytime two such lines pass close to each other. Any nucleon that undergoes at least one collision is called "wounded" [113]. In our case the projectile is a single hadron, so the number of subcollisions equals the number of wounded target nucleons.

In principle all of the components of the total cross section can contribute for each subcollision, but special consideration must be given to diffractive topologies. Notably diffractive excitation on the target side gives rapidity distributions tilted towards that side, a concept used already in the older Fritiof model [114] that partly has served as an inspiration for Angantyr. Alternatively, one can view such topologies as a consequence of the Pythia MPI machinery, wherein not all colour strings from several target nucleons are stretched all the way out to the projectile beam remnant, but some tend to get "short-circuited". If such colour connections occur flat in rapidity, then this is equivalent to a $\mathrm{d}M^2/M^2$ diffractive mass spectrum. To first approximation, an Angantyr hadron–nucleus collision can be viewed as one "normal" subcollision plus a variable number of diffractive events on the target side.

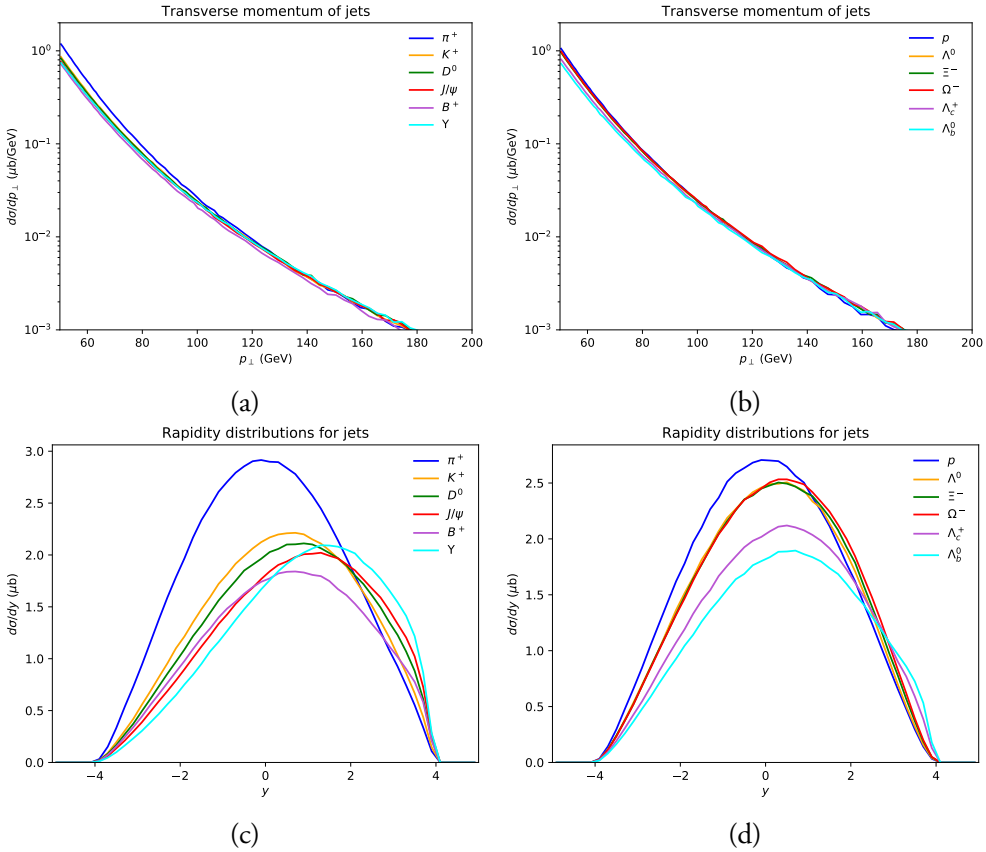


Figure IV.8: (a,b) Jet $p_\perp$ and (c,d) rapidity differential cross sections (in units of μb), with $p_{\perp,\min} = 50$ GeV. Average values are shown in Table IV.4. Labels denote the respective hadron beam.

In its basic form, Angantyr event generation is quite fast, about as fast as ordinary Pythia hadron–hadron collisions per hadron produced. That is, the overhead from nuclear geometry considerations and energy–momentum sharing between partly overlapping nucleon-nucleon collisions is negligible. The program becomes much slower if the more sophisticated features are switched on, such as ropes [115, 116], shove [117, 118], and hadronic rescattering [6]. These contribute aspects that only become apparent in a more detailed scrutiny of events, beyond what is needed for our purposes. With minor modifications to the Angantyr code itself, *ie.* on top of the Pythia-generic ones we have already introduced in this article, it is also possible to allow any hadron to collide with a nucleus.

There are two severe limitations, however. Firstly, a time-consuming recalculation of hadronic geometry parameters is required anytime the collision energy or incoming hadron species is to be changed in Angantyr. Potentially this could be fixed in the future, *eg.* by

interpolation in a grid of initializations at different energies, but it appears to be less simple than what we have introduced for the MPI framework. And secondly, Angantyr is only intended to be valid for nucleon–nucleon collision energies above roughly 100 GeV.

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

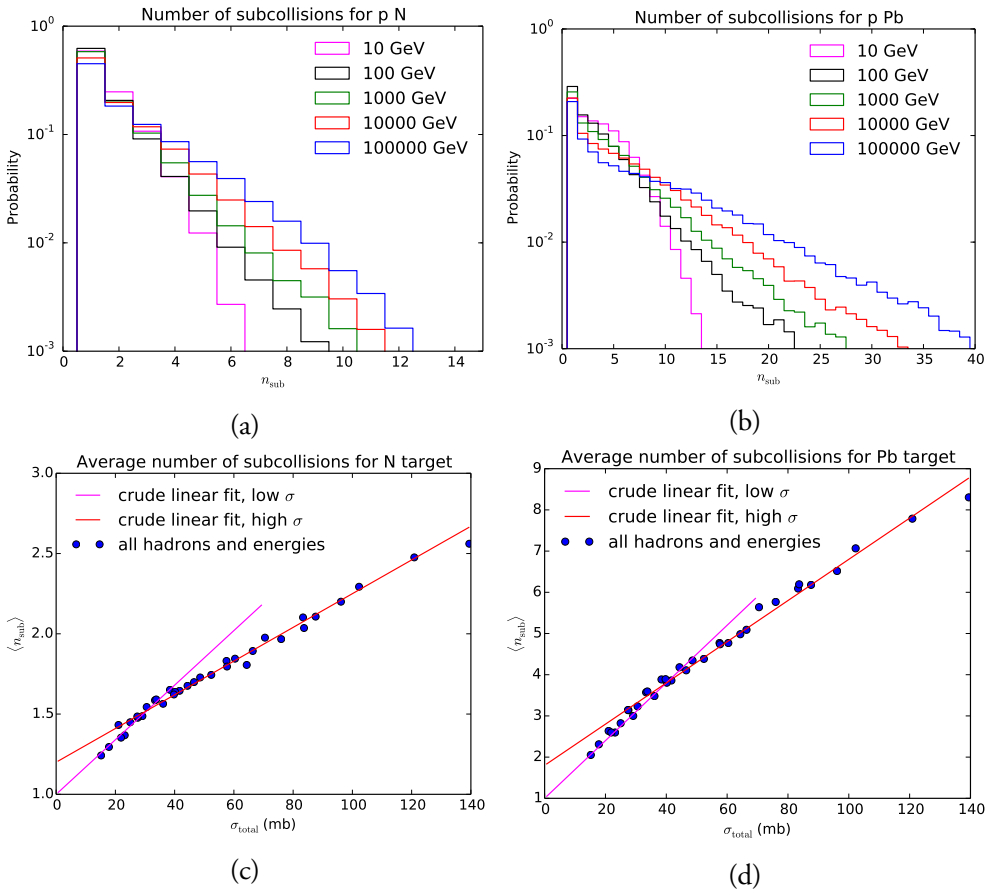


Figure IV.9: Number of subcollisions in (a) proton–nitrogen and (b) proton–lead collisions at five different subcollision energies. The average subcollision number in (c) hadron–nitrogen and (d) hadron–lead as a function of the total cross section, with some fits, see text for details.

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

An approximate geometric behaviour is also observed for other hN and hPb collisions. The average number of subcollisions depends on the hadron species and the collision energy, but mainly via the total cross section, as can be seen from Figure IV.9c,d. Here the results are shown for seven different incoming hadrons ($\mathrm{p}, \pi^+, \mathrm{K}^+, \phi^0, \Lambda^0, \Xi^0, \Omega^-$) at five different energies (10, 100, 1000, 10000, 100000 GeV). In the limit of $\sigma_{\mathrm{total}} \to 0$ one would never expect more than one subcollision. Given this constraint, a reasonable overall description is obtained as

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

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

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

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

The description of nuclear effects on hadronization is nontrivial. At Angantyr initialization the relative composition of different subprocesses is changed. This means *eg.* that the elastic rate in pN with one single subcollision is reduced relative to Pythia pp, Figure IV.10b,c. Each further subcollision in Angantyr involves the addition of a diffractive-like system on the target side, but the step from 1 to 2 also includes *eg.* a drop in the elastic fraction, Figure IV.10b, and the correlations arising from nucleons being assumed to have fluctuating sizes event-to-event. The diffractive systems have a higher activity and are more symmetric than corresponding Pythia ones, Figure IV.10c. As already mentioned

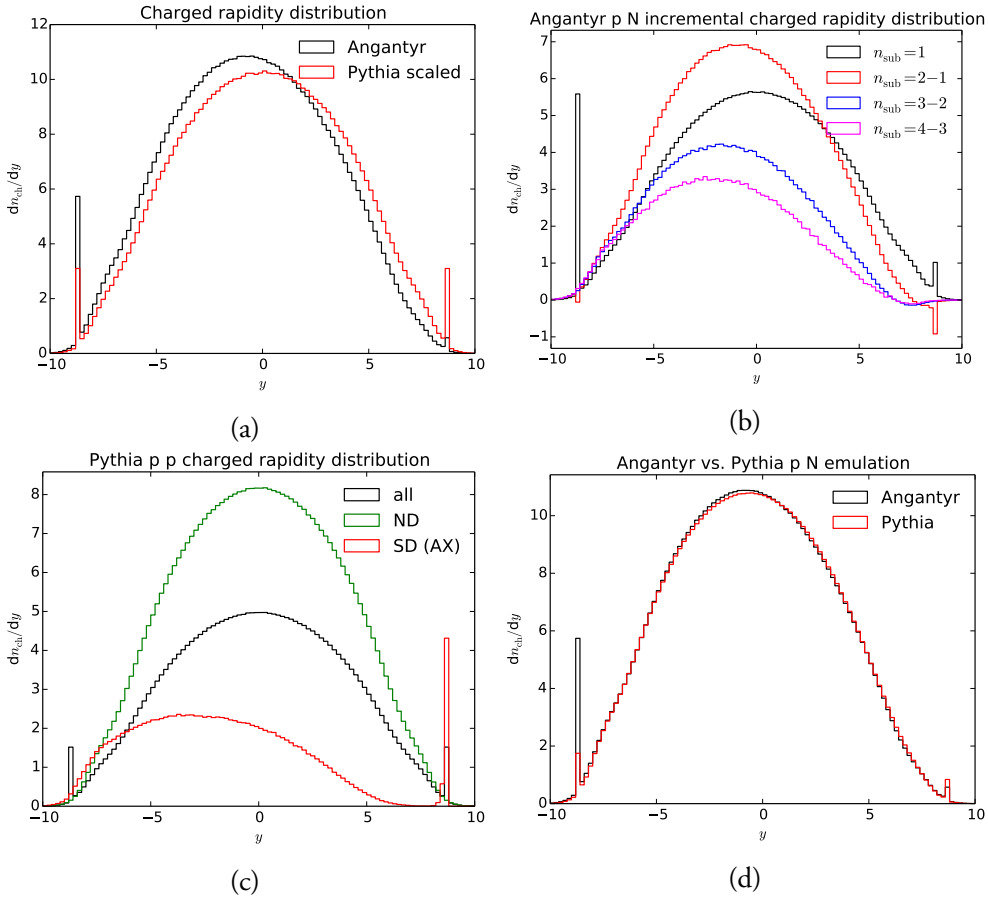


Figure IV.10: Charged rapidity distributions $\mathrm{d}n_{\mathrm{ch}}/\mathrm{d}y$ for a nucleon–nucleon collision energy of 6 TeV. (a) Inclusive Angantyr pN events relative to Pythia events scaled up with the average number of subcollisions in Angantyr. (b) Angantyr for different number of subcollisions,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

the Angantyr mechanism is not quite equivalent with that of ordinary diffraction, which is reflected in the choice of PDF for the "pomeron", as used for the MPI activity inside the diffractive system. In Angantyr a rescaled proton PDF is used, while by default the H1 2006 Fit B LO [119] pomeron PDF is used in Pythia. If MPIs (and parton showers) are switched off, the diffractive systems become quite similar, and have a marked triangular shape, as expected for a $\mathrm{d}M^2/M^2$ diffractive mass distribution. That is, without MPIs the asymmetry of Angantyr events is dramatically larger. It is possible to use the rescaled-

proton approach also in the Pythia simulation to obtain better agreement, but this is not quite good enough, so next we will present a slightly different solution.

3.3 Simplified nuclear collisions

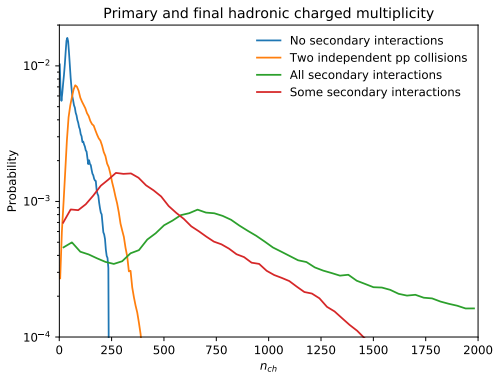


Figure IV.11: Two superimposed pp collisions compared with a single pp collision, as well as two collisions where the outgoing hadrons interact with a p medium. Either all the outgoing hadrons interact, or only some of them as outlined in the text.

pN events with two wounded nucleons in the target produce a rapidity distribution not so different from two separate pp collisions, Figure IV.10a. It would thus seem that a modelling with a naive atmosphere consisting of separate protons and neutrons would be a decent approximation. What one gets wrong in such a picture is the multiplication factor. That is, in the naive atmosphere, each of the hadrons produced in a first interaction can go on and interact in their turn. But in the correct one, with the incoming beams Lorentz-contracted pancakes, there is no time for any new hadrons to form in the passage of the proton through the nitrogen nucleus. To understand the effect of multiplication, we have studied the case where each of the products of a primary 6 TeV pp event, represented by a 19,200 TeV beam on a fixed target, can interact with one further p in the target, Figure IV.11. For this toy study the cascade stops after the second step. One should note that the primary pp collision at the maximal energy corresponds to the largest cross section in the cascade, while other hadrons at lower energies have smaller cross sections. To compensate for this we also show an option where the probability for a secondary interaction of each hadron from the primary collision is given by its cross section, normalized to the primary pp one. Also note that outgoing particles with low momenta may interact with protons in very soft elastic collisions. This introduces new slow-moving protons in the event record that presumably would not be detectable. We partly avoid this effect by only allowing hadrons to interact if the kinetic energy of the collision is larger than $E_{\mathrm{kin,min}}$, by default 0.2 GeV. This is just at the border where an inelastic collision with one additional outgoing pion is possible. Even

with such corrections it is clear that a lumping of individual nucleon collision into fewer but bigger nuclear ones makes a difference, reducing the average charged multiplicity by about a factor of five in this case.

Having noted the effect of excessive multiplication, we have introduced a simple model that allows us to keep it in check, and approximately emulate the main features of the Angantyr model for a hadron h impinging on a N target at rest.

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

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

loss is calculated before hadronization. The new procedure also leads to the central rapidity of each further subcollision being shifted towards the target region.

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

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

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

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

4 Modelling hadronic cascades

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

4.1 Medium density

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

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

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

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

where m is the target nucleus mass.

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

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

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

between z_0 and z:

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

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

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

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

In our simple study, we model the atmospheric density at altitude h starting from the ideal gas law, $\rho(h) = pM/RT$, where T and p are temperature and pressure at h, M is the molar mass of dry air, and R is the ideal gas constant. Assuming a linear drop-off for temperature (as is the case for the troposphere [120]), we have $T = T_0 - Lh$, where L is the temperature lapse rate. From the hydrostatic equation, $dp/dh = -g\rho$, the pressure is

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

Then

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

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

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

Using International Standard Atmosphere (ISO 2533:1975 [121]) values for the atmospheric parameters gives $\rho_0 = 1.225$ g/m^3 and $H = 10.4$ km. This approximation is good up until around $L = 18$ km (near the equator), but in our simplified framework, we assume the entire atmosphere follows this shape. In this case, it is possible to sample z according to eq. (IV.24), specifically,

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

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

4.2 Some atmospheric studies

In this section we study the cascade initiated by a proton hitting the model atmosphere above. In the future it would be useful to study also *eg.* an incoming iron nucleus, but here it is still not clear how to handle the nuclear remnants. The simulation includes hadronic cascades and decays, plus muon decays. But there is no simulation of electromagnetic cascades, nor electromagnetic energy loss of charged particles, nor bending in the earth magnetic field, nor a multitude of other effects. Photoproduction of hadronic states could be added, either for primary cosmic-ray photons or for secondary ones, but would require some more work, and is not considered important enough for now. The atmosphere is assumed to consist solely of nitrogen, which is a reasonable first approximation, given that *eg.* oxygen has almost the same atomic number.

Four atmospheric scenarios will be compared:

1. A constant-density atmosphere, like on earth surface, consisting of free protons and neutrons. It starts 10.4 km up, so has the same total interaction depth as the normal atmosphere at vanishing zenith angle.

2. Also a constant-density atmosphere of same height, but using the already described emulation of collisions with nitrogen.

3. An exponentially attenuated "nitrogen" atmosphere, with vanishing zenith angle. Upper cutoff at 100 km.

4. An exponentially attenuated "nitrogen" atmosphere, with 45° zenith angle. This means a factor $\sqrt{2}$ larger interaction depth than the other three scenarios. Again a 100 km upper cutoff.

The first option has already been rejected as providing too much (early) multiplication, but is kept here as a reference. What is expected to differ between the other atmospheres is the competition between decays and secondary interactions. An early interaction high up in an attenuated atmosphere gives the produced hadrons more time to decay before they can interact. For simplicity we assume an incoming proton energy of 10^8 GeV. This corresponds to a pp collision CM energy of 13.7 TeV, *ie.* just at around the maximal LHC energy. Only hadrons with a kinetic energy above $E_{\mathrm{kin,min}} = 0.2$ GeV are allowed to interact. Below that scale they can still decay, but those that do not are assumed to have dissipated and will not be counted in our studies below.

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

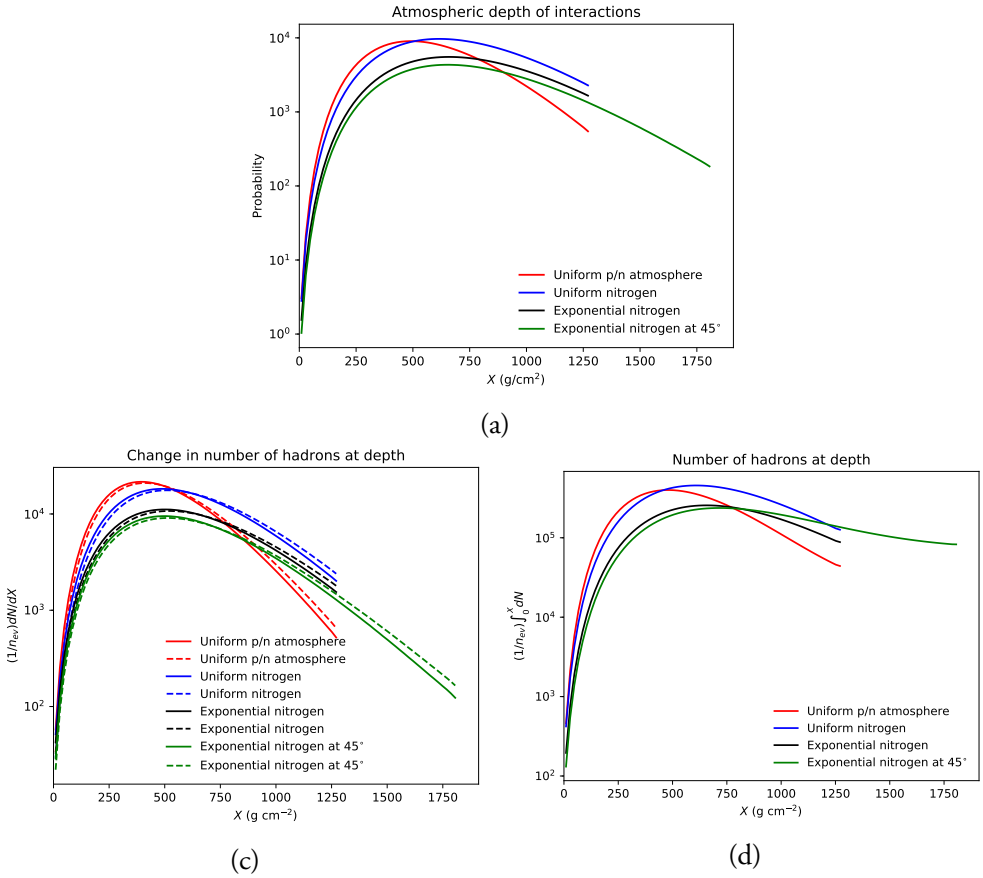


Figure IV.12: Evolution of a cascade initiated by a 10^8 GeV proton travelling through four different simple models of the atmosphere, as described in the text, as a function of atmospheric depth in g/cm^2. The models stop when reaching the surface. Shown is the number of (a) interactions, (b) hadrons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

Considering Figure IV.12a, the hadron–nucleon interaction rate as a function of atmospheric depth, we note that interactions begin earlier in the p/n atmosphere than in the N one, but then also peters out earlier, when most hadrons have low energies. Moving on to the exponential atmosphere, more hadrons (notably pions) decay before they can interact, which reduces the interaction rate. Even more so in the case of a 45° zenith angle, where more of the early evolution takes place in a thin atmosphere. The production rate of hadrons, Figure IV.12b, correlates rather well with the interaction rate, although the number of hadrons produced per collision would gradually decrease as each hadron gets to have a

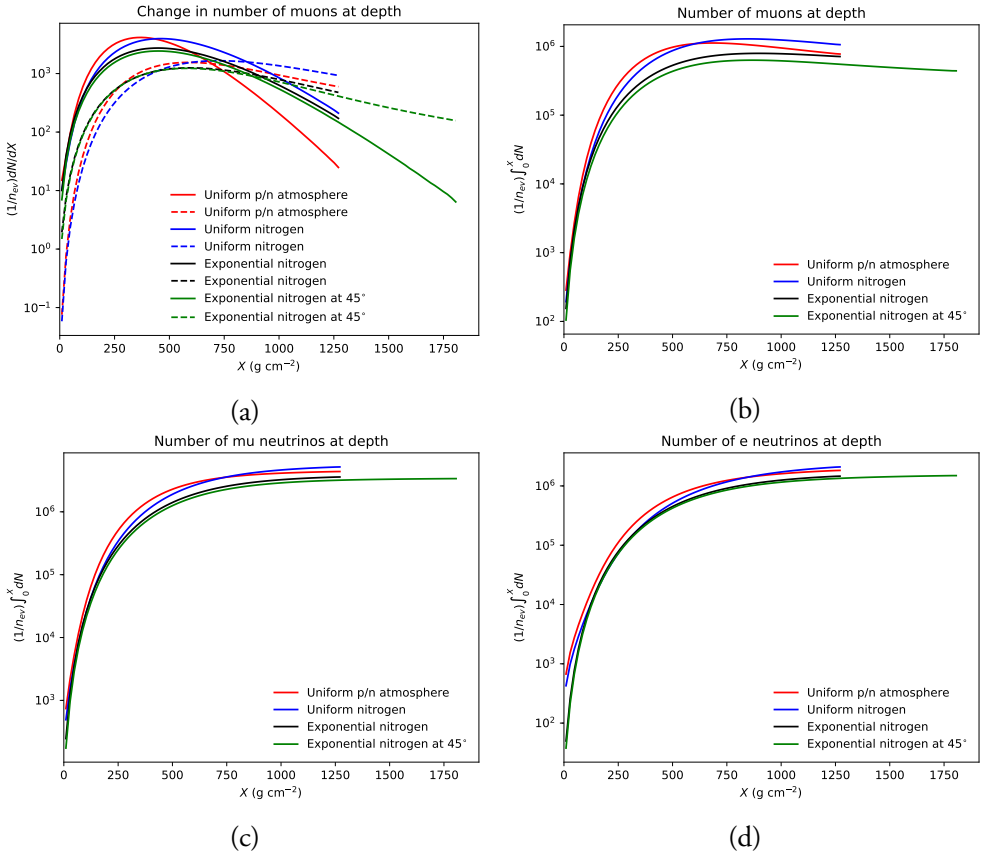


Figure IV.13: Evolution of muons and neutrinos for the same cascades as in Figure IV.12. Shown is the number of (a) muons produced (full) and decayed (dashed), (b) muons remaining, (c) muon neutrinos remaining and (d) electron neutrinos remaining.

lower energy. The early hadron production is higher for the free p/n atmosphere, consistent with expectations but not quite as dramatic as Figure IV.11 might have led one to expect. Most of the hadrons decay reasonably rapidly, leaving mainly protons and neutrons to carry on. Figure IV.12c shows how the number of such undecayed hadrons increases, following the pattern of the previous plots. Specifically, the exponential atmospheres give a reduced number of final hadrons.

The long lifetime of muons, $c\tau = 659$ m, means that muon decays lag behind production, Figure IV.13a. The number of muons reaches a plateau, where production and decay roughly balance, Figure IV.13b. The total number of muons follows the same pattern between the four atmospheric scenarios as noted for hadrons. The production of muon and electron neutrinos, Figure IV.13c,d, is dominated by pion and muon decays, but also

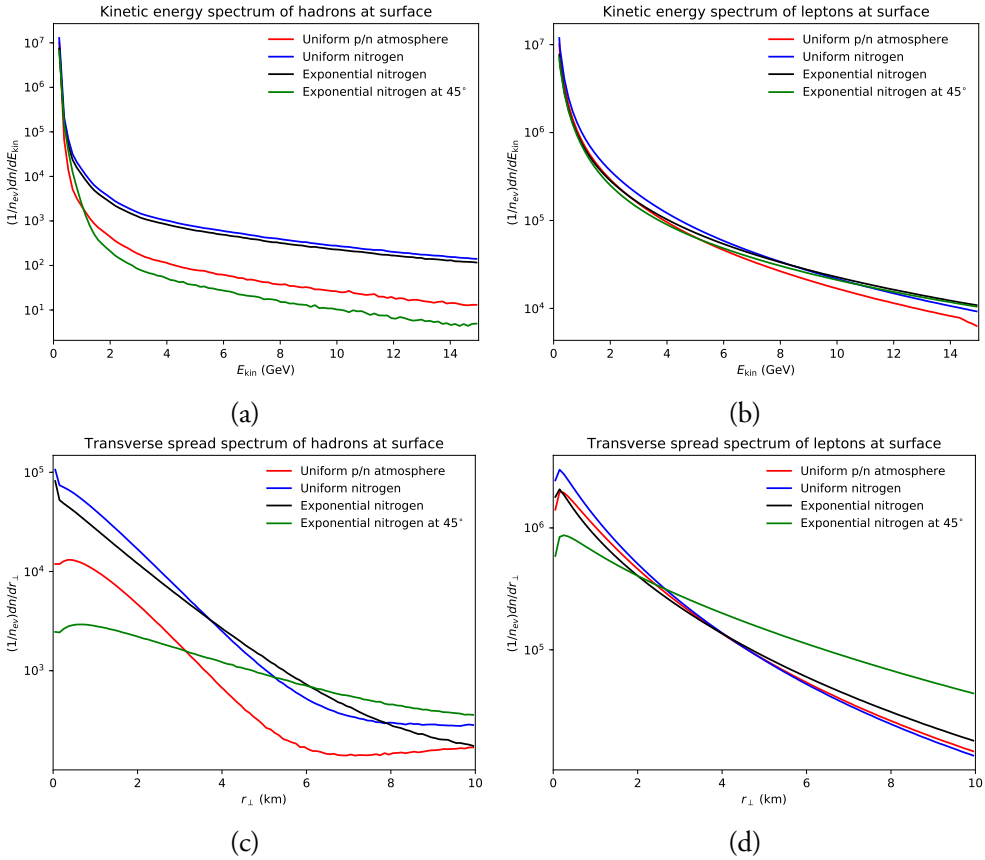


Figure IV.14: Kinetic energy spectra (in a,b) and transverse spread (in c,d) of the particles that reach the surface for the same cascades as in Figure IV.12. Shown are the spectra of (a,c) hadrons and (b,d) muons and neutrinos.

receives contributions from other weak decays. Neutrino oscillations are not considered here.

The bulk of hadrons that apparently reach the ground have very low kinetic energies, even given the cut $E_{\text{kin}} = E - m > E_{\text{kin,min}}$, Figure IV.14a. In reality most of these would be stopped or bent away by the earth magnetic field, so the figure should be viewed as a study of the consequences of hadronic cascades on their own. The uniform and exponential nitrogen atmospheres have comparable rates of higher-energy hadrons. These hadrons are dominated by p and n, which are not affected by decays. That the higher-energy hadron rate is reduced for a non-vanishing zenith angle is to be expected. Also the p/n atmosphere gives a lower rate, presumably as a consequence of the faster split of the original energy into several lower-energy collision chains. The kinetic energy spectra of muons and neutrinos,

Figure IV.14b, again are peaked at lower energies, though not quite as dramatically. The four atmospheric models also come closer to each other for leptons, though the p/n one remains an outlier.

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

4.3 A lead study

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

The main conclusion of Figure IV.15 is that the different incoming hadrons give rise to rather similar cascades. This is largely owing to the rapid multiplication into a fairly similar set of secondary hadrons. Baryons tend to have larger cross sections than mesons, and the proton the largest of them all, so it is understandable why the proton cascade starts somewhat earlier and also dies down earlier. Strange particles have somewhat lower cross sections

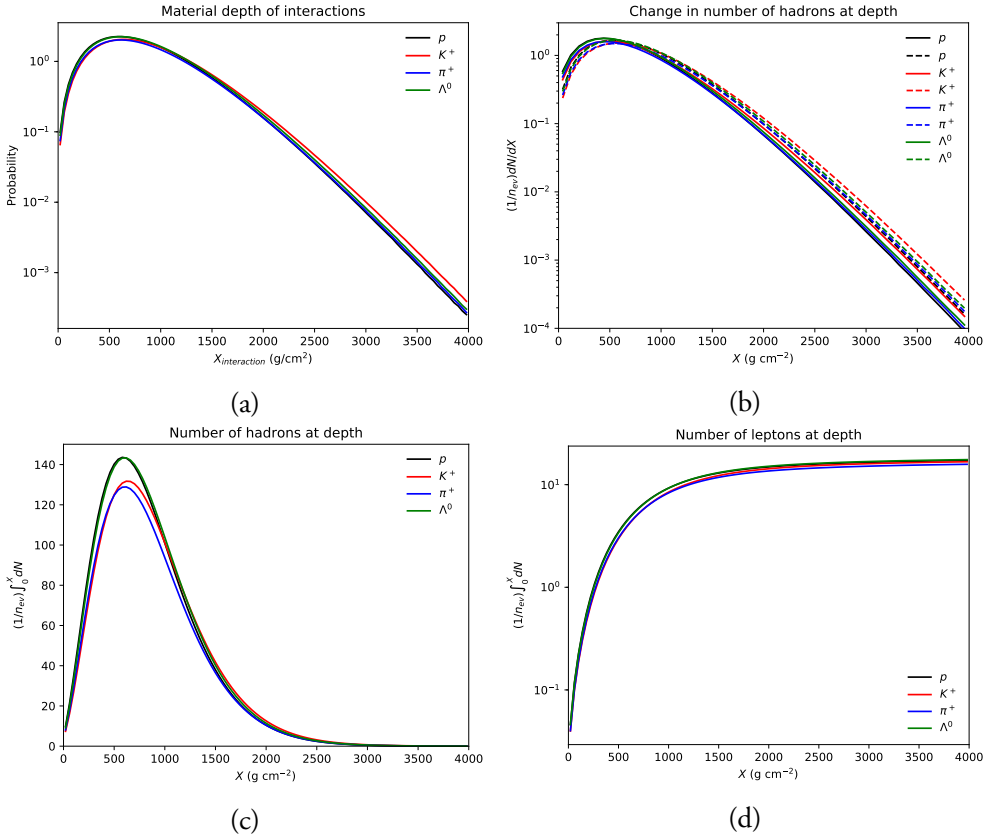


Figure IV.15: Evolution of a cascade initiated by a 1000 GeV proton, π^+, K^+ or Λ^0 passing through a 3.5 m thick slab of lead. Shown is the number of (a) interactions, (b) hadrons produced (full) and decayed (dashed), (c) hadrons remaining and (d) muons and neutrinos remaining. Hadrons that fall below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

than their non-strange counterparts, which explains why the K^+ curve starts slower than the π^+ one. But also other factors may be relevant, like how the leading-particle spectrum of a collision affects the nature of subsequent collisions. Here we expect a baryon beam to give a harder leading hadron than a meson, and a strange hadron a harder spectrum than a non-strange one, within the context of normal string fragmentation. This could partly compensate for the cross section differences. Further studies will be needed to disentangle these and other factors that may contribute to the small differences observed.

5 Summary and outlook

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

Now we have made a deeper study of almost all possible hadron–nucleon collision types. This includes deriving new total and partial cross sections at medium-to-high collision energies, based on the DL and SaS ansätze, extended with the help of the Additive Quark Model and Reggeon systematics where no data is available. It also includes producing some twenty new PDF sets, here denoted SU21. One key assumption has been that heavier valence quarks start out with a larger fraction of the total hadron momentum, at the expense of lighter quarks and gluons, so that all hadron constituents have comparable average velocities. The same constituent-quark-mass ratios as used in the AQM therefore come to characterize our new PDFs. A consistency check then is that the average number of multiparton interactions is comparable in all collision types. This average is the ratio of the integrated (mini)jet cross section, which directly relates to the PDFs used, and the total (nondiffractive) cross section. Both these numbers should reduce at comparable rates when light quarks are replaced by heavier ones.

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

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

We do not claim any fundamentally new results in this article, but still present some nice studies that point to the usefulness of the framework. We show how hadronic cascades

evolve in the atmosphere, spanning energy scales from 10^8 GeV (or higher if wanted) to 0.2 GeV, how the energy rapidly is spread among many hadrons with low energy each, how hadron decays give muon and neutrino fluxes, how the kinematics and dynamics leads to a wide spread of particles that hit the ground, and more. Note that a complete record of all particles is kept, so it is possible to ask rather specific questions, such as e.g. whether hard-jet production in the primary interaction correlate with isolated energy/particle clusters on the ground. We also show, for the solid-target case, how hadrons with larger cross sections also begin their cascades earlier, evolve faster and peter out sooner.

In the current article we have put emphasis on the applications to full cascade evolution, in the atmosphere or in solid matter, rather than on the single collision. One reason is that the full cascade offers further technical challenges on top of modelling the individual collision, which forces us to extend the capabilities of the Pythia code. Previously it has not been feasible to switch collision energy or beam type event by event, at least not without each time doing a complete reinitialization, which then slows down event generation times by orders of magnitude. The other reason is that we would like to be able to benefit from and contribute to the understanding of hadronic collisions in different environments. Currently there is one set of event generators that is mainly used for LHC pp physics, such as Herwig [122], Sherpa [123] or Pythia, and another one for cosmic rays, see the Introduction, with only EPOS as an example of a code used in both environments.

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

Furthermore, hadronic cascades is not the end of the story, but must be part of a larger framework that encompasses all relevant processes, and provides a more detailed modelling of the atmosphere. The hope is that the code will find use in larger frameworks, such as CORSIKA 8 for cosmic rays and GEANT4 for detector simulation. At the very least, we offer a far more powerful replacement to the older Pythia 6 code currently used in some such frameworks. In the future we could also take on some other related tasks, such as photoproduction in the cascades.

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

Acknowledgements

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

References

[1] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP* **05** (2006) 026, hep-ph/0603175.

[2] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.* **191** (2015) 159–177, 1410.3012.

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

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

[5] T. Sjöstrand and M. Utheim, "A Framework for Hadronic Rescattering in pp Collisions," *Eur. Phys. J. C* **80** (2020), no. 10 907, 2005.05658.

[6] C. Bierlich, T. Sjöstrand, and M. Utheim, "Hadronic Rescattering in pA and AA Collisions," 2103.09665.

[7] E. M. Levin and L. L. Frankfurt, "The Quark hypothesis and relations between cross-sections at high-energies," *JETP Lett.* **2** (1965) 65–70.

[8] H. J. Lipkin, "Quarks for pedestrians," *Phys. Rept.* **8** (1973) 173–268.

[9] J. Engel, T. K. Gaisser, T. Stanev, and P. Lipari, "Nucleus-nucleus collisions and interpretation of cosmic ray cascades," *Phys. Rev. D* **46** (1992) 5013–5025.

[10] R. S. Fletcher, T. K. Gaisser, P. Lipari, and T. Stanev, "SIBYLL: An Event generator for simulation of high-energy cosmic ray cascades," *Phys. Rev. D* **50** (1994) 5710–5731.

[11] R. Engel, F. Riehn, A. Fedynitch, T. K. Gaisser, and T. Stanev, "The hadronic interaction model Sibyll – past, present and future," *EPJ Web Conf.* **145** (2017) 08001.

[12] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, "Hadronic interaction model Sibyll 2.3d and extensive air showers," *Phys. Rev. D* **102** (2020), no. 6 063002, 1912.03300.

[13] N. N. Kalmykov and S. S. Ostapchenko, "The Nucleus-nucleus interaction, nuclear fragmentation, and fluctuations of extensive air showers," *Phys. Atom. Nucl.* **56** (1993) 346–353.

[14] N. N. Kalmykov, S. S. Ostapchenko, and A. I. Pavlov, "EAS and a quark - gluon string model with jets," *Bull. Russ. Acad. Sci. Phys.* **58** (1994) 1966–1969.

[15] S. Ostapchenko, "QGSJET-III model: physics and preliminary results," *EPJ Web Conf.* **208** (2019) 11001.

[16] J. Ranft, "The Dual parton model at cosmic ray energies," *Phys. Rev. D* **51** (1995) 64–84.

[17] S. Roesler, R. Engel, and J. Ranft, "The Monte Carlo event generator DPMJET-III," in *International Conference on Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications (MC 2000)*, 12, 2000. hep-ph/0012252.

[18] K. Werner, "Strings, pomerons, and the venus model of hadronic interactions at ultrarelativistic energies," *Phys. Rept.* **232** (1993) 87–299.

[19] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider," *Phys. Rev. C* **92** (2015), no. 3 034906, 1306.0121.

[20] T. Pierog, B. Guiot, I. Karpenko, G. Sophys, M. Stefaniak, and K. Werner, "EPOS 3 and Air Showers," *EPJ Web Conf.* **210** (2019) 02008.

[21] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw, "CORSIKA: A Monte Carlo code to simulate extensive air showers,".

[22] W. R. Nelson, H. Hirayama, and D. W. O. Rogers, "The Egs4 Code System,".

[23] H. Fesefeldt, "The Simulation of Hadronic Showers: Physics and Applications,".

[24] S. A. Bass *et. al.*, "Microscopic models for ultrarelativistic heavy ion collisions," *Prog. Part. Nucl. Phys.* **41** (1998) 255–369, nucl-th/9803035.

[25] S. J. Sciutto, "AIRES: A system for air shower simulations," astro-ph/9911331.

[26] S. J. Sciutto, "The AIRES system for air shower simulations: An Update," in *27th International Cosmic Ray Conference*, 6, 2001. astro-ph/0106044.

[27] R. Engel, D. Heck, T. Huege, T. Pierog, M. Reininghaus, F. Riehn, R. Ulrich, M. Unger, and D. Veberič, "Towards a Next Generation of CORSIKA: A Framework for the Simulation of Particle Cascades in Astroparticle Physics," *Comput. Softw. Big Sci.* **3** (2019), no. 1 2, 1808.08226.

[28] **CORSIKA 8** Collaboration, H. P. Dembinski, L. Nellen, M. Reininghaus, and R. Ulrich, "Technical Foundations of CORSIKA 8: New Concepts for Scientific Computing," *PoS* **ICRC2019** (2020) 236.

[29] I. T. Storehaug, "Improving the Atmospheric Neutrino Flux Estimation in IceCube," *master thesis, University of Copenhagen* (2019).

[30] A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft, "FLUKA: A multi-particle transport code (Program version 2005),".

[31] R. Brun, F. Bruyant, F. Carminati, S. Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban, "GEANT Detector Description and Simulation Tool,".

[32] **GEANT4** Collaboration, S. Agostinelli *et. al.*, "GEANT4–a simulation toolkit," *Nucl. Instrum. Meth. A* **506** (2003) 250–303.

[33] J. Allison *et. al.*, "Geant4 developments and applications," *IEEE Trans. Nucl. Sci.* **53** (2006) 270.

[34] J. Allison *et. al.*, "Recent developments in Geant4," *Nucl. Instrum. Meth. A* **835** (2016) 186–225.

[35] G. Battistoni, A. Margiotta, S. Muraro, and M. Sioli, "FLUKA as a new high energy cosmic ray generator," *Nucl. Instrum. Meth. A* **626-627** (2011) S191–S192, 1002.4655.

[36] P. Paschalis, H. Mavromichalaki, L. I. Dorman, C. Plainaki, and D. Tsirigkas, "Geant4 software application for the simulation of cosmic ray showers in the Earth's atmosphere," *New Astron.* **33** (2014) 26–37.

[37] R. Sarkar, A. Roy, and S. K. Chakrabarti, "Simulation of cosmic rays in the Earth's atmosphere and interpretation of observed counts in an X-ray detector at balloon altitude near tropical region," *Adv. Space Res.* **65** (2020) 189–197.

[38] P. D. B. Collins, *An Introduction to Regge Theory and High-Energy Physics*. Cambridge Monographs on Mathematical Physics. Cambridge Univ. Press, Cambridge, UK, 5, 2009.

[39] J. R. Forshaw and D. A. Ross, *Quantum chromodynamics and the pomeron*, vol. 9. Cambridge University Press, 1, 2011.

[40] S. Donnachie, H. G. Dosch, O. Nachtmann, and P. Landshoff, *Pomeron physics and QCD*, vol. 19. Cambridge University Press, 12, 2004.

[41] V. Barone and E. Predazzi, *High-Energy Particle Diffraction*, vol. v.565 of *Texts and Monographs in Physics*. Springer-Verlag, Berlin Heidelberg, 2002.

[42] C. O. Rasmussen and T. Sjöstrand, "Models for total, elastic and diffractive cross sections," *Eur. Phys. J. C* **78** (2018), no. 6 461, 1804.10373.

[43] A. Donnachie and P. V. Landshoff, "Total cross-sections," *Phys. Lett. B* **296** (1992) 227–232, hep-ph/9209205.

[44] **TOTEM** Collaboration, G. Antchev *et. al.*, "First determination of the ρ parameter at $\sqrt{s} = 13$ TeV: probing the existence of a colourless C-odd three-gluon compound state," *Eur. Phys. J. C* **79** (2019), no. 9 785, 1812.04732.

[45] **Do, TOTEM** Collaboration, V. M. Abazov *et. al.*, "Comparison of pp and $p\bar{p}$ differential elastic cross sections and observation of the exchange of a colorless C-odd gluonic compound," 2012.03981.

[46] L. Lukaszuk and B. Nicolescu, "A Possible interpretation of p p rising total cross-sections," *Lett. Nuovo Cim.* **8** (1973) 405–413.

[47] G. A. Schuler and T. Sjöstrand, "Towards a complete description of high-energy photoproduction," *Nucl. Phys. B* **407** (1993) 539–605.

[48] G. A. Schuler and T. Sjöstrand, "A Scenario for high-energy gamma gamma interactions," *Z. Phys. C* **73** (1997) 677–688, hep-ph/9605240.

[49] S. Okubo, "Phi meson and unitary symmetry model," *Phys. Lett.* **5** (1963) 165–168.

[50] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking. Version 1,".

[51] J. Iizuka, "Systematics and phenomenology of meson family," *Prog. Theor. Phys. Suppl.* **37** (1966) 21–34.

[52] G. A. Schuler and T. Sjöstrand, "Hadronic diffractive cross-sections and the rise of the total cross-section," *Phys. Rev. D* **49** (1994) 2257–2267.

[53] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton Fragmentation and String Dynamics," *Phys. Rept.* **97** (1983) 31–145.

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

[55] T. Sjöstrand, "The Development of MPI Modeling in Pythia," *Adv. Ser. Direct. High Energy Phys.* **29** (2018) 191–225, 1706.02166.

[56] C. O. Rasmussen and T. Sjöstrand, "Hard Diffraction with Dynamic Gap Survival," *JHEP* **02** (2016) 142, 1512.05525.

[57] S. J. Brodsky, P. Hoyer, C. Peterson, and N. Sakai, "The Intrinsic Charm of the Proton," *Phys. Lett. B* **93** (1980) 451–455.

[58] M. Suzuki, "Fragmentation of Hadrons from Heavy Quark Partons," *Phys. Lett. B* **71** (1977) 139–141.

[59] J. D. Bjorken, "Properties of Hadron Distributions in Reactions Containing Very Heavy Quarks," *Phys. Rev. D* **17** (1978) 171–173.

[60] **NNPDF** Collaboration, R. D. Ball *et. al.*, "Parton distributions from high-precision collider data," *Eur. Phys. J. C* 77 (2017), no. 10 663, 1706.00428.

[61] T.-J. Hou *et. al.*, "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC," *Phys. Rev. D* **103** (2021), no. 1 014013, 1912.10053.

[62] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, "Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs," *Eur. Phys. J. C* **81** (2021), no. 4 341, 2012.04684.

[63] **NA3** Collaboration, J. Badier *et. al.*, "Experimental Determination of the pi Meson Structure Functions by the Drell-Yan Mechanism," *Z. Phys. C* **18** (1983) 281.

[64] **NA10** Collaboration, B. Betev *et. al.*, "Differential Cross-section of High Mass Muon Pairs Produced by a 194-GeV/$c\pi^-$ Beam on a Tungsten Target," *Z. Phys. C* **28** (1985) 9.

[65] **NA10** Collaboration, P. Bordalo *et. al.*, "Nuclear Effects on the Nucleon Structure Functions in Hadronic High Mass Dimuon Production," *Phys. Lett. B* **193** (1987) 368.

[66] J. S. Conway *et. al.*, "Experimental Study of Muon Pairs Produced by 252-GeV Pions on Tungsten," *Phys. Rev. D* **39** (1989) 92–122.

[67] **Saclay-CERN-College de France-Ecole Poly-Orsay** Collaboration, J. Badier *et. al.*, "Measurement of the K^-/π^- Structure Function Ratio Using the Drell-Yan Process," *Phys. Lett. B* **93** (1980) 354–356.

[68] M. Glück, E. Reya, and A. Vogt, "Pionic parton distributions," *Z. Phys. C* **53** (1992) 651–656.

[69] P. J. Sutton, A. D. Martin, R. G. Roberts, and W. J. Stirling, "Parton distributions for the pion extracted from Drell-Yan and prompt photon experiments," *Phys. Rev. D* **45** (1992) 2349–2359.

[70] M. Glück, E. Reya, and M. Stratmann, "Mesonic parton densities derived from constituent quark model constraints," *Eur. Phys. J. C* **2** (1998) 159–163, hep-ph/9711369.

[71] M. Glück, E. Reya, and I. Schienbein, "Pionic parton distributions revisited," *Eur. Phys. J. C* **10** (1999) 313–317, hep-ph/9903288.

[72] B. Clerbaux and M. V. Polyakov, "Partonic structure of pi and rho mesons from data on hard exclusive production of two pions off nucleon," *Nucl. Phys. A* **679** (2000) 185–195, hep-ph/0001332.

[73] R. M. Davidson and E. Ruiz Arriola, "Parton distributions functions of pion, kaon and eta pseudoscalar mesons in the NJL model," *Acta Phys. Polon. B* **33** (2002) 1791–1808, hep-ph/0110291.

[74] F. Bissey, J. R. Cudell, J. Cugnon, M. Jaminon, J. P. Lansberg, and P. Stassart, "A Model for the pion structure function," *Phys. Lett. B* **547** (2002) 210–218, hep-ph/0207107.

[75] W. Detmold, W. Melnitchouk, and A. W. Thomas, "Parton distribution functions in the pion from lattice QCD," *Phys. Rev. D* **68** (2003) 034025, hep-lat/0303015.

[76] M. Aicher, A. Schäfer, and W. Vogelsang, "Soft-gluon resummation and the valence parton distribution function of the pion," *Phys. Rev. Lett.* **105** (2010) 252003, 1009.2481.

[77] R. J. Holt and C. D. Roberts, "Distribution Functions of the Nucleon and Pion in the Valence Region," *Rev. Mod. Phys.* **82** (2010) 2991–3044, 1002.4666.

[78] C. Han, H. Xing, X. Wang, Q. Fu, R. Wang, and X. Chen, "Pion Valence Quark Distributions from Maximum Entropy Method," *Phys. Lett. B* **800** (2020) 135066, 1809.01549.

[79] A. Watanabe, C. W. Kao, and K. Suzuki, "Meson cloud effects on the pion quark distribution function in the chiral constituent quark model," *Phys. Rev. D* **94** (2016), no. 11 114008, 1610.08817.

[80] P. C. Barry, N. Sato, W. Melnitchouk, and C.-R. Ji, "First Monte Carlo Global QCD Analysis of Pion Parton Distributions," *Phys. Rev. Lett.* **121** (2018), no. 15 152001, 1804.01965.

[81] A. Watanabe, T. Sawada, and M. Huang, "Extraction of gluon distributions from structure functions at small x in holographic QCD," *Phys. Lett. B* **805** (2020) 135470, 1910.10008.

[82] M. Ding, K. Raya, D. Binosi, L. Chang, C. D. Roberts, and S. M. Schmidt, "Symmetry, symmetry breaking, and pion parton distributions," *Phys. Rev. D* **101** (2020), no. 5 054014, 1905.05208.

[83] I. Novikov *et. al.*, "Parton Distribution Functions of the Charged Pion Within The xFitter Framework," *Phys. Rev. D* **102** (2020), no. 1 014040, 2002.02902.

[84] A. Watanabe, T. Sawada, and C. W. Kao, "Kaon quark distribution functions in the chiral constituent quark model," *Phys. Rev. D* **97** (2018), no. 7 074015, 1710.09529.

[85] A. Watanabe, T. Sawada, and C. W. Kao, "Meson cloud effects on kaon quark distribution functions and the SU(3) flavor symmetry," *Nucl. Part. Phys. Proc.* **300-302** (2018) 121–125, 1810.04032.

[86] C. Shi, C. Mezrag, and H.-s. Zong, "Pion and kaon valence quark distribution functions from Dyson-Schwinger equations," *Phys. Rev. D* **98** (2018), no. 5 054029, 1806.10232.

[87] J. Lan, C. Mondal, S. Jia, X. Zhao, and J. P. Vary, "Pion and kaon parton distribution functions from basis light front quantization and QCD evolution," *Phys. Rev. D* **101** (2020), no. 3 034024, 1907.01509.

[88] Z.-F. Cui, M. Ding, F. Gao, K. Raya, D. Binosi, L. Chang, C. D. Roberts, J. Rodríguez-Quintero, and S. M. Schmidt, "Kaon and pion parton distributions," *Eur. Phys. J. C* **80** (2020), no. 11 1064.

[89] M. Glück, R. M. Godbole, and E. Reya, "Dynamically Generated Parton Distributions for High-energy Collisions," *Z. Phys. C* **41** (1989) 667.

[90] M. Glück, E. Reya, and A. Vogt, "Parton distributions for high-energy collisions," *Z. Phys. C* **53** (1992) 127–134.

[91] M. Glück, E. Reya, and A. Vogt, "Dynamical parton distributions of the proton and small x physics," *Z. Phys. C* **67** (1995) 433–448.

[92] M. Glück, E. Reya, and A. Vogt, "Dynamical parton distributions revisited," *Eur. Phys. J. C* **5** (1998) 461–470, hep-ph/9806404.

[93] M. Glück, P. Jimenez-Delgado, and E. Reya, "Dynamical parton distributions of the nucleon and very small-x physics," *Eur. Phys. J. C* **53** (2008) 355–366, 0709.0614.

[94] **NNPDF** Collaboration, R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo, "Parton distributions with QED corrections," *Nucl. Phys. B* **877** (2013) 290–320, 1308.0598.

[95] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory," *Sov. J. Nucl. Phys.* **15** (1972) 438–450.

[96] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.," *Sov. Phys. JETP* **46** (1977) 641–653.

[97] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys. B* **126** (1977) 298–318.

[98] M. Botje, "QCDNUM: Fast QCD Evolution and Convolution," *Comput. Phys. Commun.* **182** (2011) 490–532, 1005.1481.

[99] S. Alekhin *et. al.*, "HERAFitter," *Eur. Phys. J. C* **75** (2015), no. 7 304, 1410.4412.

[100] V. Bertone, S. Carrazza, and J. Rojo, "APFEL: A PDF Evolution Library with QED corrections," *Comput. Phys. Commun.* **185** (2014) 1647–1668, 1310.1394.

[101] **LHCf** Collaboration, O. Adriani *et. al.*, "Measurement of forward neutral pion transverse momentum spectra for $\sqrt{s} = 7$TeV proton-proton collisions at LHC," *Phys. Rev. D* **86** (2012) 092001, 1205.4578.

[102] **LHCf** Collaboration, O. Adriani *et. al.*, "Measurement of very forward neutron energy spectra for 7 TeV proton–proton collisions at the Large Hadron Collider," *Phys. Lett. B* **750** (2015) 360–366, 1503.03505.

[103] V. Kireyeu, I. Grishmanovskii, V. Kolesnikov, V. Voronyuk, and E. Bratkovskaya, "Hadron production in elementary nucleon–nucleon reactions from low to ultra-relativistic energies," *Eur. Phys. J. A* **56** (2020), no. 9 223, 2006.14739.

[104] B. Andersson, G. Gustafson, and T. Sjöstrand, "Baryon Production in Jet Fragmentation and Υ Decay," *Phys. Scripta* **32** (1985) 574.

[105] B. Andersson, G. Gustafson, I. Holgersson, and O. Mansson, "A Model for the Reaction Mechanism and the Baryon Fragmentation Distributions in Low p_T Hadronic Interactions," *Nucl. Phys. B* **178** (1981) 242–262.

[106] P. Edén and G. Gustafson, "Baryon production in the string fragmentation picture," *Z. Phys. C* **75** (1997) 41–49, hep-ph/9606454.

[107] P. Edén, "A Program for baryon generation and its applications to baryon fragmentation in DIS," hep-ph/9610246.

[108] F. Kling, "Forward Neutrino Fluxes at the LHC," 2105.08270.

[109] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, "LHAPDF6: parton density access in the LHC precision era," *Eur. Phys. J. C* **75** (2015) 132, 1412.7420.

[110] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP* **04** (2008) 063, 0802.1189.

[111] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *Eur. Phys. J. C* **72** (2012) 1896, 1111.6097.

[112] R. J. Glauber, "Cross-sections in deuterium at high-energies," *Phys. Rev.* **100** (1955) 242–248.

[113] A. Bialas, M. Bleszynski, and W. Czyz, "Multiplicity Distributions in Nucleus-Nucleus Collisions at High-Energies," *Nucl. Phys. B* **111** (1976) 461–476.

[114] B. Andersson, G. Gustafson, and B. Nilsson-Almqvist, "A Model for Low p(t) Hadronic Reactions, with Generalizations to Hadron - Nucleus and Nucleus-Nucleus Collisions," *Nucl. Phys. B* **281** (1987) 289–309.

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

[116] C. Bierlich, "Rope Hadronization and Strange Particle Production," *EPJ Web Conf.* **171** (2018) 14003, 1710.04464.

[117] C. Bierlich, G. Gustafson, and L. Lönnblad, "Collectivity without plasma in hadronic collisions," *Phys. Lett. B* **779** (2018) 58–63, 1710.09725.

[118] C. Bierlich, S. Chakraborty, G. Gustafson, and L. Lönnblad, "Setting the string shoving picture in a new frame," *JHEP* **03** (2021) 270, 2010.07595.

[119] **H1** Collaboration, A. Aktas *et. al.*, "Measurement and QCD analysis of the diffractive deep-inelastic scattering cross-section at HERA," *Eur. Phys. J. C* **48** (2006) 715–748, hep-ex/0606004.

[120] National Weather Service, "Layers of the atmosphere." https://www.weather.gov/jetstream/layers, 2019.

[121] ISO, "Standard atmosphere," ISO 2533:1975, International Organization for Standardization, Geneva, Switzerland, 2007.

[122] J. Bellm *et. al.*, "Herwig 7.2 release note," *Eur. Phys. J. C* **80** (2020), no. 5 452, 1912.06509.

[123] **Sherpa** Collaboration, E. Bothmann *et. al.*, "Event Generation with Sherpa 2.2," *SciPost Phys.* 7 (2019), no. 3 034, 1905.09127.