Developing the Angantyr model and extending colour reconnection for heavy-ion collisions in Pythia

Shah, Harsh

2023

Developing the Angantyr Model and Extending Colour Reconnection for Heavy-ion Collisions in Pythia

HARSH SHAH
DEPARTMENT OF PHYSICS | FACULTY OF SCIENCE | LUND UNIVERSITY

Developing the Angantyr Model and Extending Colour Reconnection for Heavy-ion Collisions in Pythia

Developing the Angantyr Model and Extending Colour Reconnection for Heavy-ion Collisions in Pythia

by Harsh Shah

LUND
UNIVERSITY

Thesis for the degree of Doctor of Philosophy
Thesis advisor: Prof. Leif Lönnblad
Faculty opponent: Prof. Andrzej Siódmok

To be presented, with the permission of the Faculty of Science of Lund University, for public criticism in the Rydbergssalen at the Department of Physics on 2^{nd} of November 2023 at 09:15.

Organization	Document name
LUND UNIVERSITY Department of Physics Professorsgatan 1 SE–223 63 LUND Sweden	**DOCTORAL DISSERTATION** Date of disputation 2023-11-02 Sponsoring organization
Author(s) Harsh Shah	

Title and subtitle

Developing the Angantyr Model and Extending Colour Reconnection for Heavy-ion Collisions in Pythia:

Abstract

This thesis is composed of four papers, the first two papers show the attempts made to extend the Pythia event generator to simulate heavy-ion collision events. The other two papers show methods to improve the hadronization mechanism in the Pythia model for a better agreement with the experimental data of pp and heavy-ion collisions.

Paper I introduces a new model named Angantyr for heavy-ion collision events simulation using the Pythia machinery. We developed a technique for sophisticated stacking of the pp-like sub-collisions, so that the Angantyr model can simulate pA and AA type heavy-ion collision events. We show that the Angantyr model can reproduce the general distributions of the produced particles in pp and heavy-ion collision events.

Paper II extends the QCD inspired colour reconnection mechanism in Pythia by introducing spatial constrains and extends the hadronization mechanism involving junction systems. The extended QCD colour reconnection model is used to improve the integration of the stacked pp-like sub-collisions as a single heavy-ion event in the Angantyr model by introducing a global colour reconnection. We show that the upgraded model can reproduce the general distribution of the produced particles in pp collision events and qualitatively in heavy-ion collision events. We also show that the model can reproduce the strangeness enhancement like feature in the baryon sector even though the model has no special treatment for the strangeness enhancement.

Paper III investigates the baryon angular correlations and the hadronization mechanism using Pythia. It highlights the need for improvement in the string fragmentation mechanism to reproduce the baryon angular correlations. We propse a hypothesis for the jets sub-structure studies as an outcome of our investigation.

Paper IV improves the formation of junction like setups in the QCD inspired colour reconnection and the fragmentation of the junction configurations in the string fragmentation. We show that the changes made in this paper improves the charm baryons production rate in pp and heavy-ion collision events.

Key words

Heavy-ion Collisions, Colour Reconnections, Hadronization, QCD Phenomenology

Classification system and/or index terms (if any)

Supplementary bibliographical information	Language English
ISSN and key title	ISBN 978-91-8039-838-1 (print) 978-91-8039-837-4 (pdf)
Recipient's notes	Number of pages 214 Price Security classification

Developing the Angantyr Model and Extending Colour Reconnection for Heavy-ion Collisions in Pythia

by Harsh Shah

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: Picture of Nataraj, taken during one of the visits at CERN.
This cosmic dance form represents five activities: **creation** (symbolized by the drum), **protection** (by the "fear-not" pose of the hand), **destruction** (by the fire), **embodiment** (by the foot planted on the ground), and **release** (by the foot held aloft).

Faculty of Science , Department of Physics

ISBN: 978-91-8039-838-1 (print)
ISBN: 978-91-8039-837-4 (pdf)

Printed in Sweden by E-huset, Lund University, Lund 2023

To my family

Contents

Acknowledgements v
List of publications vii
Popular summary viii

Introduction **1**
1 Heavy-ion Collisions 3
2 Heavy-ion collision observables 6
2.1 Collective flow 7
2.2 Jet quenching 10
2.3 Other observables 11
3 Event generators 12
3.1 Hydro-dynamical approach 13
3.2 Non-thermal models 14
3.3 pp event generator 15
3.4 Heavy-ion event generators 23
3.5 Angantyr 26
4 Conclusion and Outlook 33
References 37
5 Overview of publications 42

Publications

I The Angantyr Model for Heavy-ion Collisions in Pythia8 47
1 Introduction 48
2 Nucleon-nucleon sub-collisions in $\mathbf{pA}$ and $\boldsymbol{AA}$ 52
2.1 Glauber formalism 53
2.2 Fluctuations 53
2.3 From cross sections to probabilities 56
2.4 $\boldsymbol{NN}$ scattering models used in Glauber calculation Monte Carlos 57
2.5 Nucleon fluctuations in $\boldsymbol{AA}$ collisions 59
3 From wounded nucleons to exclusive final states 61
3.1 Multiparton interactions in pp collisions 62
3.2 Multi-parton interactions in a pA collision 63

3.3 Multi-parton interactions in an $\boldsymbol{AA}$ collision . . . 65
4 Generating and combining parton-level $\boldsymbol{NN}$ events . . . 66
4.1 Selecting primary absorptive collisions . . . 66
4.2 Adding secondary absorptive interactions. . . . 67
4.3 Adding diffractive interactions . . . 68
5 Modifications of single diffractive to secondary absorptive . . . 69
5.1 High-mass diffractive excitation and secondary absorptive . . . 71
5.2 Comparing primary and secondary absorptive sub-events . . . 73
6 Sample results . . . 77
6.1 pp results . . . 78
6.2 $\mathbf{p}\boldsymbol{A}$ results . . . 80
6.3 $\boldsymbol{AA}$ results . . . 83
6.4 Collectivity and non-flow estimation . . . 87
7 Model uncertainties . . . 89
7.1 Uncertainties in treating secondary wounded nucleons . . . 90
7.2 Diffractively excited nucleons . . . 94
7.3 Uncertainties in AA collisions . . . 95
8 Relation to other models . . . 96
9 Conclusion and Outlook . . . 99
10 Appendix: Generating absorptively and diffractively wounded nucleons . . 101
10.1 Absorptively wounded nucleons . . . 101
10.2 Diffractively wounded nucleons . . . 102
References . . . 103

II A Spatially Constrained QCD Colour Reconnection in pp, pA, and AA Collisions in the PYTHIA8/Angantyr Model 111
1 Introduction . . . 112
2 The Colour Reconnection . . . 114
2.1 Spatially constrained model . . . 118
2.2 Improved junction handling . . . 119
3 Selection of parameters and retuning strategy . . . 121
3.1 Low-$p_\perp$ suppression for MPIs . . . 122
3.2 m_0 and C_j parameters . . . 124
3.3 Allowed dipole separation . . . 125
3.4 ϵ_{pom} for SND events . . . 127
3.5 CR effects in pA and AA collisions . . . 128
4 Results . . . 129
4.1 pp results . . . 131
4.2 pPb results . . . 134
4.3 PbPb results . . . 135
5 Discussion and outlook . . . 137

6 Appendix: Junction Fragmentation 140
References 144

III Baryon Correlations in Pythia **149**
1 Introduction 150
2 Baryons, popcorn and gluons in the Lund Model 152
2.1 Gluons vs. popcorn 154
3 Junctions and colour reconnections 156
4 Final-state effects on correlations 158
5 Comparison with data 158
5.1 The QCD colour reconnection model 159
5.2 Hadronic rescattering 159
5.3 Suppressing baryon production close to gluon kinks 161
6 Discussion and summary 163
References 165

IV The Dynamic Hadronization of Charm Quarks in Heavy-ion Collisions **171**
1 Introduction 172
2 Heavy ion collisions with the Angantyr model 173
3 Charm hadron production in PYTHIA 174
3.1 Charm quark in hard process and parton shower 174
3.2 Colour reconnection and hadronization 175
3.3 The role of junctions 177
4 Results 183
4.1 Hyperon production in pp collisions 183
4.2 Charmed baryon production in pp collisions 183
4.3 **pPb** collisions 187
5 Conclusion 188
References 191

Acknowledgements

The time spent during my doctorate training was an Odyssey for me. I learned many things within physics and outside of the scientific endeavour during my long stay at Lund University. Throughout this journey, I crossed with many individuals. I am carrying some memories of those encounters, which will remain with me forever to cherish. At this stage, I would like to thank all of them.

While completing the thesis for the highest academic degree, I have to mention three people who inspired and influenced my choice to continue pursuing academic research. First, my high school teacher, Vinod Khandivar, who continuously boosted my curiosity and encouraged me to ask questions. The second person is Bedanga Mohanty. Working with him as a summer project student I encountered the term of *Quark Gluon Plasma* for the first time. Today, I am submitting a doctorate thesis that challenges the Quark Gluon Plasma creation hypothesis. The third person is Satyanarayana Bheesette. I worked with him as a summer project student. He inspired me to develop persistence and patience, which are essential tools for the seeker of the answers to the mysteries of nature.

I am thankful to my supervisor, Leif Lönnblad. Leif has been a unique mentor I have ever experienced. He is calm, thoughtful, and supportive. He let me explore and experience the uncertainties in the life of a researcher. This is the kind of research training I seek in the first place: to learn how to ask the right questions and filter out the most scientific answer to move forward in search of new questions. The most valuable lesson I learned is that the job of a researcher doesn't end with finding the solutions. Instead, it's the beginning of the pain of writing down the details of the exploration with scientific precision in the choice of words. But in the end, I managed to survive, thanks to Gösta and Leif.

I am grateful to Christian Bierlich, Gösta Gustafson, and Torbjorn Sjöstrad for their patience in answering the same questions repeatedly. Christian always grounded my vaguely thought questions with his counter questions, and helped me to formulate my thoughts and to ask better questions. I express my gratitude to Gösta for the mind-boggling depth of his clarity about the fundamentals of the Lund Strings and scientific writing. Torbjorn's insightful responses and extensive comments to my questions related to the Pythia model helped me to grasp the details and nuances of phenomenological concepts.

I am also thankful to all of my office mates, and everyone from the corridor of theoretical physics. I bothered many of you (especially, Christopher, Eliel, Hugo, Johan R., Jonas W., Leif G., and Stefan) with my questions, which are often unconventional and creations of my imagination. You all tried your best to satisfy my curiosity.

I am grateful to Christine and Johannes for their valuable advice and encouragement when I needed it the most, which boosted my self-esteem and helped me to move forward.

I am grateful to have met fellows from different countries, exchanged many pleasant memories, and pondered upon many scientific ideas on multiple occasions during the collaboration meetings and summer schools, as a part of the MCnet and CLASH collaborations.

I would like to thank Leif and Rikkert for their expert comments and suggestions while proofreading my thesis. I am also grateful to Chiara for her suggestions which improved the lucidity of the thesis text.

The magic of Marius, long multi-topic discussions with Torbjörn (young!), special lunches, Fika, and many pub nights with fellow PhDs and Postdocs colleagues: Anca, Andrecia, Andrew, Baptiste, Jasmina, Jarkko, Joan, João, Mårten, Nils, Robin, and Smita. Thank you for the memories of a lifetime. A special thanks to Mattias and Timea for helping out with the thesis template.

I am also thankful to the secretory office for always being fast and efficient in helping me with administrative tasks.

The journey of a PhD life is incomplete without conferences and schools. Both are important tools to spread the word about one's scientific achievements and explore the scientific world's vastness. Such events sometimes gave me opportunities to meet old friends and also to make some new friends. Getting to know Iwona was one of such extra perks of attending conferences and schools. Iwona, thank you for the long weekly talks and debates over a range of topics with your unique insights, a YouTube channel could have earned us some royalty :)!

I thank Bhavya, Reena, Sargam, Vivek (Bela), and Vivek (VP) for many long conversations over a variety of topics. I am also thankful to Arun, Deepak, Dipen, Himanshu, Ishani, Jay, Mrudula, and Pinki, who are also the culprits of many of my non-scientific adventures, and I am always grateful to you guys for the time we spent together. A special thanks to Mrudula for sharing a similar passion for food and getting along in experimenting with different cafes and restaurants.

We need friends and family to overcome the sorrow and share the joys of life. Usually, those who are closest to my heart, often hear thank-you notes last, my parents, sister, and my other family members and friends back home, thank you for your moral support and encouragement whenever I needed. Iwona, thank you once again for being part of my life, and my gratitude to your family for welcoming me with love as a new family member. Archi, Ram, Michel, and Vayu, thank you for being my family and sharing my sorrowful moments as well as joyful events.

This thesis is in part dedicated to you all too.

List of publications

This thesis is based on the following publications:

I **The Angantyr Model for Heavy-ion Collisions in Pythia8**

Christian Bierlich, Gösta Gustafson, Leif Lönnblad, and Harsh Shah
J. High Energ. Phys., 134 (2018)
e-print: `arXiv:1806.10820 [hep-ph]`

II **A Spatially Constrained QCD Colour Reconnection in pp, pA, and AA Collisions in the PYTHIA8/Angantyr Model**
Leif Lönnblad and Harsh Shah
Eur. Phys. J. C 83 (2023) 7, 575
e-print: `arXiv:2303.11747 [hep-ph]`

III **Baryon Correlations in Pythia**

Leif Lönnblad and Harsh Shah
To be submitted to Eur. Phys. J. C
e-print: `arXiv:2309.01557 [hep-ph]`

IV **The Dynamic Hadronization of Charm Quarks in Heavy-ion Collisions**

Christian Bierlich, Gösta Gustafson, Leif Lönnblad, and Harsh Shah
To be submitted to Eur. Phys. J. C
e-print: `arXiv:2309.12452 [hep-ph]`

All papers are reproduced with permission of their respective publisher, with minor stylistic changes in the layout and wording.

Popular summary

The present form of Homo sapiens is the result of thousands of years of evolution. We find enough archaeological evidence of human evolution, which consists of not only as a species but also as an observer of the surroundings and the universe. I believe that at various stages, especially while staring at the night sky, our ancestors had thought about fundamental questions like "Where are we coming from?" and "What is the origin of everything?"

The idea that everything is made from a few fundamental particles was thought centuries ago in ancient India by philosophers like "Kanad" (unclear, between $6^{th} - 2^{nd}$ century BCE), and in ancient Greece by philosophers like "Leucippus"(around 5^{th} century BCE). The idea is associated with the philosophy of "atomism", where the *atom* represents the fundamental building block from which all the elements of nature are formed. This idea remained part of philosophical debates for centuries due to the technological limitations to test its validity experimentally.

Until the late-18^{th} century, neither did we know much about the constituents of the various objects we find in nature nor did we know about the fundamental forces other than gravitation (Thanks to Isaac Newton, who provided the laws of motion). But at the beginning of the 19^{th} century, scientists discovered that another force behaves similarly to that of gravity. This force is called an electric force. Unlike gravity, which only acts as an attractive force, the electric force can be attractive or repulsive. Later, through the work of scientists like Michael Faraday and James Clerk Maxwell, we learn that the electric and widely known magnetic effects are related and they can be combined as electromagnetic effects. It was also the first attempt at unification of two forces known as independent of each other.

The discovery of electromagnetism opened gates for a different way to search for the building blocks of nature. By the end of the 19^{th} century, the first fundamental particle was discovered using electromagnetism and it is named $electron$. The discovery of electrons also expanded our knowledge about the fundamental properties of different objects. The electrons not only carry a mass but also a negative electric charge. Later, the charge will be identified as an intrinsic property of the fundamental particle.

At the beginning of the 20^{th} century, a heavier particle carrying a positive eclectic charge was discovered, named *proton*. This discovery also provided a better picture of the atom, which is no longer a fundamental particle but a structure consisting of a heavy nucleus at the centre, which contains protons and neutrons as more fundamental objects and electrons revolving around the nucleus. At this stage, from the study of electric force, it was also clear that at the scale of these fundamental particles ($\sim 10^{-9} - 10^{-15} m$), the gravitational force has negligible strength compared to the electric force. Henceforth, the gravitational effects

are ignored in the study of fundamental particles.

The 20^{th} century was the golden period of many discoveries in the study of fundamental building blocks of nature. We have entered the era of *quantum* world, where at the atomic scale, the continuum of the classical world is broken by the discreteness in the energy scales. Scientists have developed machines that use electromagnetic force to accelerate particles with electrical charge at a speed close to the speed of light in a vacuum. These machines are called *accelerators*. The accelerators are used to collide a bunch of such particles by accelerating them in opposite directions, such machines are called *colliders*. The results of high-energy collisions of particles are detected in giant machines, and the analysis of those measurements provides information about what they are made of. From some earlier experiments of electron-proton collisions, we learned that protons (and neutrons) are composed of fundamental particles, namely *quarks*. Proton-like particles consist of three quarks and are grouped as *baryons*. A quark and an anti-quark (the same particle with opposite charge) can also form a particle, and they are called *mesons*.

These experiments also discovered two new forces, $weak$, and $strong$ forces. The weak force is partially responsible for radioactive decay, while the strong force is responsible for binding quarks inside the protons. A theoretical framework of Quantum Electro-Dynamics (QED) was developed which describes the electromagnetic interactions between the charged particles. The QED model became an inspiration for further theoretical development of understanding of the strong force, and property like *colour* charge was discovered, which is associated with the strong force interactions. The theoretical framework to study interactions due to the strong force is called the Quantum Chromo-Dynamics (QCD).

From the hard work of many scientists during the last century, a comprehensive theoretical framework was developed whose validity has been tested by many experiments, called "the standard model" (SM) of particle physics. The most recent discovery of one of the fundamental particles, *Higgs* boson, was made in 2012 at the Large Hadron Collider (LHC) facility at CERN. The SM shows that the world around us is made of a few fundamental particles, and these particles interact with each other via **three** fundamental forces, namely electromagnetic, weak, and strong force by exchange of the force carrier particles. Although the SM is the most successful theoretical framework providing information about the interactions between the fundamental particles, it has many limitations, and discussion about those limitations is out of the scope here.

Now we take a different approach and move away from fundamental particles to think about how the universe formed in the first place. According to the widely accepted *Big Bang* model, our universe came into existence more than 13.6 billion years ago. The model predicts an explosion of the cosmic egg and the energy and matter are pushed outward, and cool down with time. It is assumed that during the initial stage of the expansion of

the universe, there might be a situation when the fundamental particles would have been moving freely before forming the bound states like baryons, mesons, and atoms. This state of freely moving fundamental particles is called the *Quark Gluon Plasma*.

While studying the fundamental building blocks of nature at the particle colliders, we may be able to recreate an environment similar to the beginning of the universe. To achieve it, we need to collide large nuclei, where simultaneously many particles from each colliding nuclei interact, and a large amount of energy is deposited in the collision.

The first of such collisions was done at the Relativistic Heavy Ion Collider (RHIC) experiment. The results from Gold nuclei collisions in the RHIC experiment were motivating and suggested that the hypothesis of a QGP formation may be true. Later, similar results were observed in the proton-proton collisions at the Large Hadron Collider (LHC) facility, CERN. The similarity between proton-proton and heavy ion collision results is unexpected, and this thesis is one approach where I studied a possible uniformity of the physics in both types of collisions by developing a computational model.

So far, I highlighted theoretical and experimental developments in search of fundamental building blocks of nature. There is also a phenomenological approach, where one uses analytical solutions from theoretical studies to build a computational model, and the values of its free parameters are obtained by fitting them to the experimental results. The computational model is used to reproduce the different experimental results, predict some new observations at experiments, and test the efficiency of the experimental measurements.

A simple computational model can be developed to reproduce the probability distribution for flipping a coin. We know that a coin flip can result in either "heads" or "tails", only two outcomes. A simple approach is to use a random number generator and assign heads if the value is bigger than some number 'X' otherwise assign tails. But if one wants to make a better, more sophisticated, and general-purpose computational model for the coin-flipping experiment then several details need to be thought through and included in the model. For example, information about some properties of the coin e.g. material, size etc. has to be included. Moreover, the details about the environment, the position and direction of the force vectors, external parameters etc. have to be included in the model. Here, parameters related to the coin are taken from the experimental measurements. The analytical solution for the outcomes can be derived using the probability theory.

One can use this advanced model to generate simulation results and validate the model against the coin-flip experimental results. Now with this validated model, a user can perform many new tests by generating computational results without experimenting. For example, one can change the material properties, the density profile of the coin, the medium, and the position of the force vector, and test if there's any bias in the results or predict new possible outcomes. One can extend the model for the possibility that the coin may have a thick edge, and there's a faint chance that it will stand on the edge and not result in heads

or tails. In short, such modelling can allow us to test many experimental scenarios without experimenting, propose new observables for experiments to look out for, and based on new experimental evidence update the theoretical components of the model to provide a better agreement with the experimental results.

Similarly, computational modelling can be done for particle collisions. Pythia is one such model, which has been developed and improved over four decades to simulate proton-proton collisions. My thesis is a part of developing and extending the Pythia model to perform heavy-ion collision simulations, which we named the Angantyr model. One of the major interests in developing the Angantyr model is to understand the proton-proton and heavy-ion collisions with the same underlying physics principles. Moreover, the results from the Angantyr model will confront the widely assumed hypothesis of the creation of an environment similar to the early universe in the particle colliders. Overall, the work done in my thesis will help us to scrutinise the results from the particle colliders and contribute towards a better understanding of the underlying physics explanations of the observations made at the collider experiments.

Introduction

This thesis is about an extension of the Pythia (PYTHIA8) [1, 2] event generator, primarily used to simulate proton-proton (pp) collision events, to also simulate heavy-ion (HI) collision events. The model we have developed to generate HI collision events is called "the Angantyr model" in Pythia. An important question, that has motivated this study, is whether a HI collision is an accumulation of many nucleon-nucleon collisions or if there is indeed a transition from a hadronic[1] state to a new state of matter; a hot plasma of free *partons*[2], also known as a Quark Gluon Plasma (QGP).

The QGP is thought to exist in the early universe after a few microseconds of the Big Bang [3]. Therefore, if such a new state of matter can be created in the laboratory by HI collisions, it is crucial to verify and validate its existence. For this purpose, we have to verify the observations which are used to support the existence of the hot plasma, because if we can explain the observations without any assumption of the plasma then it would raise serious questions regarding our assumptions about the creation of the hot plasma. In this thesis, we have started to test the limits to which a phenomenological model (Pythia/Angantyr model) can reproduce HI collision observables without assuming the creation of a hot thermalised medium.

A heavy ion (heavy nucleus) can be imagined as a bag containing protons and neutrons, collectively called nucleons. A collision of two such bags leads to simultaneous collisions of many nucleons. In collider experiments, the nuclei collide at almost the speed of light in a vacuum, and at such high energies, the incoming beams are exposed to sub-structures of the nucleons. Protons and neutrons consist of three valence quarks, and depending on the energy-momentum exchange of the collision, an incoming particle may also interact with a sea (anti-) quark or gluon from the target nucleon. These sub-collisions involve the exchange of a large amount of kinetic energy and, as a result, the total number of produced particles in a HI collision is an order of magnitude higher than in a pp collision.

[1] Baryons (e.g. protons and neutrons) and mesons (e.g. pions) are collectively called hadrons. Baryons have three quarks (or three anti-quarks) and mesons have a quark and an anti-quark in their sub-structure.

[2] Quarks and gluons are collectively called partons.

Heavy-ion collision experiments at the Relativistic Heavy Ion Collider (RHIC) and the Large Hadron Collider (LHC) have observed phenomena like *strangeness enhancement*, *near side ridge*, *jet quenching*, and *quarkonia suppression*. These observations (which are described in section 2) have been accepted as the signatures of the new state of matter, which was thought to be created only in HI collisions.

The LHC experiments with pp collisions at $\sqrt{s} = 7$ TeV and 13 TeV show that the high multiplicity pp events also produce *near side ridge* [4] and *strangeness enhancement* [5]. These new results from the LHC not only blur the boundary between HI collisions and pp collisions physics but also raise doubts about a phenomenological understanding of the observations made in the HI collision experiments.

Observations of enhancement or suppression in the multiplicity of specific types of particles and long-range collectivity among produced hadrons are difficult to explain by perturbative Quantum Chromo Dynamics[3] (pQCD) first principles, and we require phenomenological models to investigate such observations. Since phenomenological models are developed based on theoretical understanding and the free parameters[4] of the theories are tuned to the experimental data, they act as the bridge connecting theory predictions and experimental observations.

However, the LHC results of *strangeness enhancement* and *near side ridge* in pp collisions are not reproduced by Pythia-like event generators. Pythia simulates collision events of smaller systems like pp, with an assumption of colour strings interacting at zero temperature in a vacuum. It is important to note that the final state hadrons emerge from the colour string fragmentation mechanism (which is based on the Lund String Model [6]) in Pythia. Moreover, the entire collision simulation in Pythia is performed without any prior assumption of the creation of a thermalised medium (the QGP). On the other hand, the idea of a medium creation and its thermalisation is the basis of conventional HI event generators.

Since these LHC results of pp collisions are similar to some of the HI observations, if they are produced due to the same underlying physics, then these results should be consistent with either of the phenomenological approaches. Hence questions like "*How do we explain such unexpected observations in pp collisions?*", "*Is there a boundary distinguishing HI collisions from pp collisions?*", and "*Can all relativistic hadronic collisions be explained using the same phenomenological approach?*" are necessary to be addressed for better understanding of the results from the collider experiments.

The hydrodynamics-based HI event generators require a certain amount of the number density of produced partons to be achieved by the colliding systems. Hence hydrodynamics-

[3]QCD is a quantum field theory, and it is used to describe interactions through colour charge exchange between partons.

[4]These "free parameters" normally are related to the non-perturbative models. There are parameters in perturbative calculations as well, but they are usually under better control.

based models can not reproduce observables for collisions with a low number density of produced partons (which are the majority of pp and e^+e^- collisions). On the other hand, the colour strings interaction mechanism in a microscopic event generator like Pythia can be modified. As an example, in the Pythia event generator, the string-string interactions can be modified in the high string density environment, which can qualitatively reproduce the new observations in high multiplicity pp collisions [7, 8].

If a HI collision is indeed a sophisticated superposition of multiple pp-like collisions, then the reproduction of HI observables is possible using the string interaction modifications with the Angantyr model, and we can have *one model to rule them all*.

In section 1 I provide an overview of the conventional approach to visualize a heavy-ion collision event. I describe the widely accepted HI observables, which are considered exclusive to HI collisions in section 2. I show how a heavy-ion collision is considered in hydrodynamics-inspired models and how non-thermal models are different from the hydro-models in sections 3.1 and 3.2 respectively. Later, I describe the event generators, I talk about Pythia in 3.3 and then about the Angantyr model after providing short overviews of some of the widely used HI event generators in section 3.4. At last, I conclude my thesis introduction with key outcomes from my publications before sharing a wish list of some of the open questions that can be addressed in future based on the work done in this thesis.

1 Heavy-ion Collisions

A relativistic accelerated nucleus looks like a longitudinally Lorentz contracted flat disc[5], and two such Lorentz contracted nuclei collide in a HI collision (Figure 1). The overlap of the colliding discs is defined by the impact parameter (**b**), the distance between the centres of the two nuclei at the time of the collision. The complete overlap ($\mathbf{b} = 0$) represents a head-on collision, and the smaller the overlap between the nuclei, the larger the impact parameter. Even at the fixed **b**, the radii of the nucleons and the positions of the nucleons inside the nuclei vary event-by-event, which contributes to fluctuations in the interaction probability.

The impact parameter can not be controlled and measured directly in the HI collision experiments. It has to be estimated using some observables, which are assumed to correlate with the collision geometry. The event multiplicity (number density of the produced hadrons) is an observable that strongly correlates with **b**. Therefore by measuring the multiplicity of produced hadrons in the central (or forward) region, we can indirectly estimate **b**. The

[5]The thickness of the disc depends on the relativistic γ factor as $2\mathbf{r}/\gamma$; $\gamma = 1/\sqrt{1 - v_z^2/c^2}$, where **r**, v_z, and c are the nucleus' radius, nucleus' velocity along the beam axis, and the velocity of light in vacuum respectively.

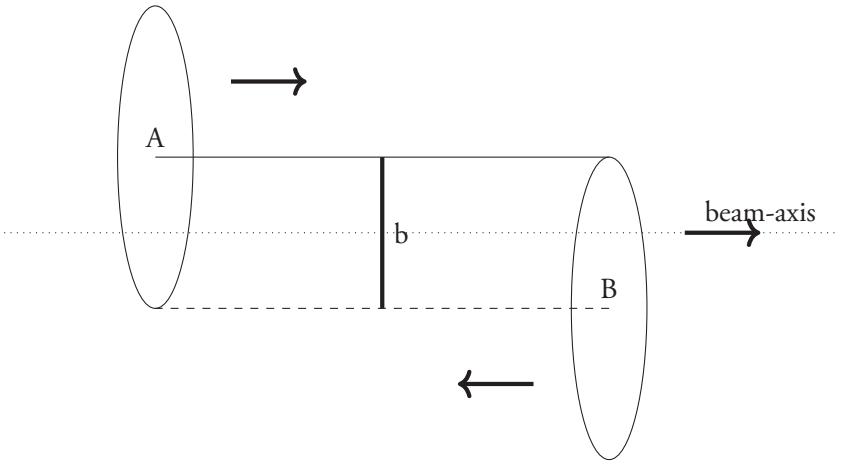


Figure 1: A schematic representation of a HI collision at an impact parameter (**b**) between two Lorentz contracted nuclei **A** and **B**. The relative directions of motion of the two nuclei are shown with arrows.

HI collision events are studied based on this estimate. The events are separated into percentile bins of the measured total charged multiplicity, which is called the event *centrality*. The least percentile values ($\sim 0-5\%$) represent central (or higher multiplicity) collisions, while the gradual increase in the percentile refers to more peripheral (or lower multiplicity) collisions. The centrality-based analyses are used to compare central HI collisions with peripheral HI collisions and to compare results from HI collisions with pp collisions.

Nucleons within the overlapping area often participate in the collision, and those interacting nucleons are called *participants*. Nucleons outside of the overlapped area usually do not collide, and they continue to travel in the respective beam's direction after the collision, and they are called *spectators*. One can not directly measure the number of participants (N_{part}) and number of nucleon-nucleon collisions (N_{coll}) in a HI collision. Therefore one has to estimate N_{part} and N_{coll} using phenomenological models, and the Glauber model [9, 10] is widely used for this purpose. A correlation between the Glauber calculated N_{part}, and the observed event multiplicity is assumed. For example, the highest ($\sim 0-5\%$) centrality bin of the event multiplicity distribution is assumed to correspond to a certain interval of **b**, and based on that impact parameter interval the corresponding average N_{part} or N_{coll} are estimated using the Glauber model (I will discuss more about the Glauber model later).

Lattice[6] QCD calculations show that at extremely high temperatures, the QCD system in equilibrium has the number of degrees of freedom close to a system of free partons. It

[6]A non-perturbative QCD calculation is performed on a lattice of discrete points in a finite space-time volume.

means, there is a phase transition[7] from a state of confined hadronic matter to a deconfined system of free partons. Moreover, Lattice QCD calculations also show that at a temperature around 300 MeV[8], the QCD matter can have an energy density close to 12 GeV/fm^3.

Let me show an estimate of the average energy density in a pseudo-rapidity[9] interval $\eta \in$ [-0.5, 0.5] between two colliding nuclei at time $1\mathrm{fm/c}$ after a PbPb head-on ($\mathbf{b} = 0$) collisions at $\sqrt{s_{NN}}$ = 2.76 TeV. If particle masses are neglected compared to their energy then one can safely consider rapidity[10], $y \approx \eta$. From the relation $y = tanh^{-1}\beta_L$, where β_L is the longitudinal velocity, and $\beta_L \in$ [-0.46, 0.46] for the given η interval. For an approximate total transverse energy of 1.65 TeV [11], the radius of the Pb nucleus ≈ 7 fm, and the separation between two Pb nuclei at $1\mathrm{fm/c}$ time after a collision is 0.92 fm. If we assume approximately a cylindrical system then the energy density with complete overlap between two Pb nuclei is, $1.65\ \mathrm{TeV}/(\pi \times (7\mathrm{fm})^2 \times 0.92\mathrm{fm}) \approx 11\ \mathrm{GeV/fm}^3$, which is close to the Lattice QCD prediction for the matter density at the extreme temperature during the QCD phase transition. Hence it is not unreasonable to expect that a QGP may be formed in HI collisions.

This QGP formation can also be possible either during the early stage of the universe (when the temperature is extreme) or in the core of the neutron stars (where the temperature is low, but the matter density is approximately 3-5 times larger than the nuclear density). Unfortunately, neither of these physical systems can be accessed experimentally to verify the theoretical predictions about the QCD phase transition and creation of the QGP.

However, as discussed above, HI collisions may create an environment similar to a few microseconds after the Big Bang in the laboratory (in the particle colliders). Therefore, the HI physics research is motivated to collide heavy ions at relativistic energies to explore the possibility of the QCD phase transition and the QGP creation. As an example, the QCD phase diagram with temperature vs. baryon chemical potential is explored in Au-Au collisions by varying collision energies ($\sqrt{s_{NN}}$ = 7 - 200 GeV) in experiments at the Relativistic Heavy-Ion Collider (RHIC).

The fluctuations in the amount of energy deposited in the post-collision environment depend on the event-by-event fluctuations of the geometric overlap between the two colliding

[7]The QCD phase transition is analogous to the phase transition of water to vapour at 373 Kelvin and one atmospheric pressure. But while water-to-vapour is a first-order phase transition, the QCD phase transition is a continuous cross-over at zero baryon chemical potential.

[8]In the natural units, $1\ \mathrm{eV/k_B} \approx 1.16 \times 10^4$ Kelvin, where k_B is the Boltzmann constant.

[9]The pseudo-rapidity (η) is used to define the angular position of the referred particle with respect to the beam axis. It is represented as $\eta = \frac{1}{2}\ln\left(\frac{|\mathbf{p}|+p_z}{|\mathbf{p}|-p_z}\right)$, where $\mathbf{p}$ is three momenta and p_z is the momentum of the particle in the direction of the beam axis.

[10]The rapidity (y) is defined as $y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}$, where E and p_z are the energy and the longitudinal momentum. The rapidity (y) can be exchanged with pseudo-rapidity (η) for mass-less or ultra-relativistic particles, which then gives angle (θ) of those particles w.r.t. beam axis because $\eta = -\ln(\tan\frac{\theta}{2})$.

heavy ions. Moreover, these fluctuations are the direct consequence of the fluctuations in the number of nucleon-nucleon collisions in each HI collision. The majority of nucleon-nucleon collisions do not exchange a large amount of transverse momentum (p_T), only a small fraction of hard nucleon-nucleon collisions produce particles with large p_T in HI collisions. These high p_T particles undergo interactions with particles produced in other nucleon-nucleon collisions and lose their energy to the medium. The energy loss results in suppressed signals for the energy deposition in the detectors (more details are in the section "jet quenching"), which is very unlikely to happen in smaller collision systems e.g. e^+e^-, pp etc.

The hadronisation of produced partons and the properties of the partons, which travel through the QGP medium, are other concepts being explored in HI collision experiments. These experiments observe different rates of flavour (types of quarks) production compared to pp collisions. The observed relative production rate of the identified hadrons in HI collision experiments requires a better understanding of the hadronisation mechanism. On the other hand, the hadronisation in collisions like $e^+e^- \longrightarrow q\bar{q}$ is well understood. The produced quarks form jets[11], which are well described and reproduced by the *Lund String Model* [6] implemented in the Pythia event generator for $\mathrm{e^+e^-}$ and pp collisions.

One of the differences between HI collisions and pp collisions is that the QGP is expected to exist only in the HI collisions. However, the new LHC results (mentioned in the previous section) from pp collisions indicate that there may be new hadronisation mechanisms when a large amount of kinetic energy is involved in the collisions irrespective of the colliding system sizes. Following this observation, we need to scrutinise the interpretations of the hadron production mechanism in the HI collisions.

2 Heavy-ion collision observables

The observations of *jet-quenching* and *collective flow* at the RHIC and the LHC are the two most important observations from the HI collision experiments. The jet quenching is considered a result of a high transverse momentum parton losing its energy due to multiple interactions with the partons from the QGP medium created in HI collisions. Whereas, the collective flow of hadrons was predicted in the ideal hydrodynamic models as a consequence of the medium with hydrodynamic properties, and its observation at the RHIC supported the claim of a QGP formation [12] in HI collisions.

HI collision observables are often compared with pp collision observables to study and investigate differences (or similarities) between these colliding systems. These comparisons are performed by analysing the nuclear modification factor R_{AA}. For example, pseudora-

[11]A jet is a collimated spray of hadrons in the direction of a gluon or a quark.

pidity (η) and transverse momentum (p_T) dependent $R_{AA}(\eta, p_T)$ is:

$$R_{AA}(\eta, p_T) = \frac{\frac{d^2N_{AA}}{d\eta dp_T}}{\langle N_{coll}\rangle \frac{d^2N_{pp}}{d\eta dp_T}}, \tag{1}$$

where the numerator is the differential particle yield in HI collisions and the denominator is the average number of nucleon-nucleon collisions ($\langle N_{coll}\rangle$) (obtained from a Glauber model) in HI collisions multiplied with the differential particle yield in pp collisions. This is a general representation of R_{AA} and it can be modified according to the other properties of a study-specific observable e.g. jets or identified hadrons. The suppression (R_{AA} < 1) or enhancement (R_{AA} > 1) will reflect nucleus-induced effects or effects due to a QGP formation in HI collisions (because so far the QGP is believed to be created only in HI collisions.). As an example, for a certain type of identified hadrons, R_{AA} below unity means production of these hadrons in HI collisions is suppressed compared to pp collisions. R_{AA} is also studied for jets [13, 14] to understand medium effects on jets at different centrality, and to explore the possibility of jet quenching in the high multiplicity pp collisions.

2.1 Collective flow

Strong azimuthal correlations are observed among particles widely separated in rapidity [15, 16] in HI collisions. These correlations are stronger than one can expect from a mere superposition of many pp collisions, but the correlations are expected if some kind of plasma is formed, which can have a collective outward flow of matter due to its fluid-like properties. These correlations indicate that they may have originated in the early stage of the collision, and their study may provide indirect information about the earliest stage of the collision.

The arrangement of nucleons inside the overlapped region of the colliding nuclei defines the anisotropy in their spatial distribution (Figure 2). This spatial anisotropy of participants results in an asymmetry in the collision geometry, and that is assumed as the origin of anisotropy in the momentum distribution of the produced particles.

The Fourier decomposed multiplicity as a function of the azimuthal angle (with respect to the reaction plane) gives an indirect insight into the initial state properties of the collision.

$$E\frac{d^3N}{d^3p} = \frac{d^2N}{p_T dp_T dy}\frac{1}{2\pi}\Big[1 + \sum_{n=1}^{\infty} 2v_n \cos n(\phi - \Psi_{RP}^n)\Big]. \tag{2}$$

Here the quantities E, p, p_T, y, ϕ and Ψ_{RP}^n are the energy, total momentum, transverse momentum, rapidity, azimuth angle of the produced particles, and reaction plane angle of the n^{th}-order flow symmetry plane w.r.t. x-axis respectively. The reaction plane angle (Ψ_{RP}^n) (figure 2) for n^{th} harmonic is not an experimentally known quantity. It is derived

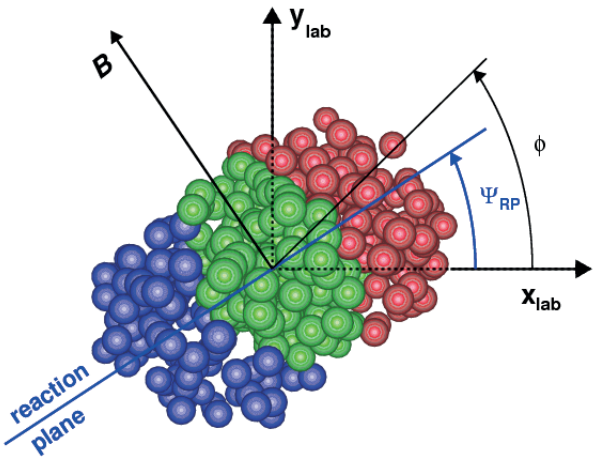


Figure 2: A schematic representation of a HI collision, where the beam axis is perpendicular to the page. Blue and red spheres represent nucleons from two colliding nuclei. Green spheres represent nucleons participating in the collision. The blue solid line and Ψ_{RP} define the reaction plane (formed by the beam axis and the impact parameter axis) and angle of the reaction plane with respect to the x-axis (lab-frame) respectively. The image is from [17].

on the event-by-event basis through the estimated plane which has a maximum number of transversely distributed final state particles. This estimated reaction plane contains the impact parameter (which is obtained from a Glauber model) and the beam axis.

In the absence of fluctuations, the complete overlap ($\mathbf{b} = 0$) of colliding nuclei produce particles moving outward with isotropic momentum vectors, and all flow coefficients $v_n = 0$. The collisions do not always occur without fluctuations and the overlap of the colliding nuclei also varies event-by-event, and the flow coefficients v_n may be non-zero quantities. Non-vanishing v_n coefficients represent contributions from various modes of the generated anisotropy in the non-central HI collisions (usually an almond-shaped collision geometry; Figure 2). The first two coefficients of the cosine function are v_1 (directed flow) and v_2 (elliptic flow), and the second order coefficient v_2 is the most sensitive observable to the "almond-shaped" collision geometry.

The differential v_n coefficients can be determined by the event plane method [18]. The v_n coefficients are obtained as a function of p_T and rapidity (y), and they are averaged over all

events in the given p_T and rapidity bin:

$$v_n(p_T, y) = \langle e^{in(\phi - \Psi^n_{RP})} \rangle = \langle \cos n(\phi - \Psi^n_{RP}) \rangle = \frac{\int \frac{d^3N}{p_T dp_T dy d\phi} \cos n(\phi - \Psi^n_{RP}) d\phi}{\int \frac{d^3N}{p_T dp_T dy d\phi} d\phi}. \tag{3}$$

Another method to obtain v_n coefficients is the multi-particle correlation technique. The two-particle correlation is the simplest case of the multi-particle cumulants method [19, 20], where azimuthal correlations between particle pairs are obtained independent of the reaction plane angle (Ψ^n_{RP}). For example, the $v_n\{m\}$ coefficients of m-particle correlation can be defined from the cumulants $c_n\{m\}$. For $m = 2$, it is:

$$v_n\{2\} = \sqrt{c_n\{2\}} = \sqrt{\langle \{ e^{in(\phi_1 - \phi_2)} \} \rangle} = \sqrt{\langle \{ \cos n(\phi_1 - \phi_2) \} \rangle} = \langle v_n^2 \rangle + \delta_n, \tag{4}$$

where $\{..\}$ represents average over all particles in an event and $\langle .. \rangle$ represents average over all events. Here ϕ_1 and ϕ_2 are the azimuthal angles of the paired particles, $c_n\{2\}$ refers to the two-particle cumulant, which provides $\langle v_n^2 \rangle + \delta_n$, where δ_n refers to the contribution from the *non-flow* components. The non-flow originates from correlations due to jet production and resonance particle decay, and these correlations are assumed to be short-ranged. The non-flow contributions can be minimised by various techniques or by excluding jetty events from the flow analysis. Hence, a rapidity gap ($\Delta y > 1$) between paired particles is often introduced when calculating the cumulant to reduce the non-flow contributions.

Higher order multi-particle cumulants (e.g. $c_n\{4\}$, $c_n\{6\}$ etc.) are calculated to reduce uncertainty and non-flow contribution, because the lower order multi-particle cumulants are subtracted from the higher order multi-particle cumulants. For example :

$$v_n\{4\}^4 = -c_n\{4\} = 2\langle v_n^2 \rangle^2 - \langle v_n^4 \rangle, \tag{5}$$

$$v_n\{6\}^6 = \frac{1}{4} c_n\{6\} = \langle v_n^6 \rangle - 9\langle v_n^4 \rangle \langle v_n^2 \rangle + 12\langle v_n^2 \rangle^3. \tag{6}$$

Here, $\langle v_n^2 \rangle$, $\langle v_n^4 \rangle$, and $\langle v_n^6 \rangle$ represent two-particle, four-particle, and six-particle flow components. In the above two equations, it is evident that the lower-ordered multi-particle components are subtracted while calculating the higher-order multi-particle cumulants.

The initial state fluctuations play a major role in the anisotropic flow study. Therefore spatial and geometrical fluctuations are studied theoretically as the *eccentricity* analysis. A one-to-one correlation between eccentricity and flow coefficients is also observed in models. Some references for more detailed discussions of event-by-event fluctuations and anisotropic flow are obtained in [21–23].

2.2 Jet quenching

To understand jet quenching it is useful to consider a di-jet event, which consists of the observation of two highly energetic collimated sprays of hadrons at approximately 180° angular separation (in azimuth) in the detector. The g + g $\longrightarrow$ g + g scattering is the most dominant partonic interaction in the high energy collisions, and they emerge as jets in the detectors. The time scale of partonic interactions is inversely proportional to the energy scale of the partons. Therefore, hard interactions with the exchange of large transverse energy occur during the earliest stage of the collision, even before a thermalised QGP formation.

In HI events, jets often propagate through the so-called QGP medium and lose some energy due to the jet-medium interactions, which suppress the signal of an expected jet in the opposite direction of a leading jet. This phenomenon of suppressed or missing sub-leading jet is called *jet quenching*. It was first observed in HI collisions at RHIC and now also at LHC, but so far no jet quenching has been observed in pp collisions.

The energy loss depends on the colour charge, the mass of the travelling parton, and the colour-charged density of the medium. The hard gluons travel through the medium and lose energy to the medium either in the form of medium-induced radiation or through the elastic scattering with medium partons. The former is the dominant mode of energy loss in the case of high p_T partons [24, 25]. The energy loss is proportional to the path travelled by the parton inside the medium, and jet quenching analyses provide insight into the medium size and the QGP medium lifetime.

If jet quenching occurs due to the coloured medium interacting with the high energy partons, then there must not be any energy loss for the colourless particles travelling through the medium e.g. photons, $W^{\pm}$and Z^{0}. Figure 3 shows the nuclear modification factor as a function of p_T for colourless particles, which have their values close to one, and suppressed nuclear modification factors for coloured particles for PbPb collisions.

For charged hadrons R_{pA} is consistent with unity for $p_T > 2GeV$ (blue points in the left window of figure 3), hence suppression in R_{AA} can not be an initial state effect, but it must be a final state effect similar to the jet quenching. If the suppression in the R_{AA} would have been an initial state effect then a similar trend should be observed in R_{pA} for high p_T particles as well. The suppression in the charged hadrons yield can be related to the energy loss for jets passing through the coloured medium and losing their energy to the medium particles. Analogously, instead of losing energy, the high p_T charged hadrons may decay within the medium after multiple re-scatterings. Heavy mesons of weakly bound quark-antiquark pairs break their bonds, and the quarks and anti-quarks rebound with different flavoured anti-quarks or quarks from the medium partons. Such effects require a larger volume of the coloured medium, which is more likely to be created in the final state of

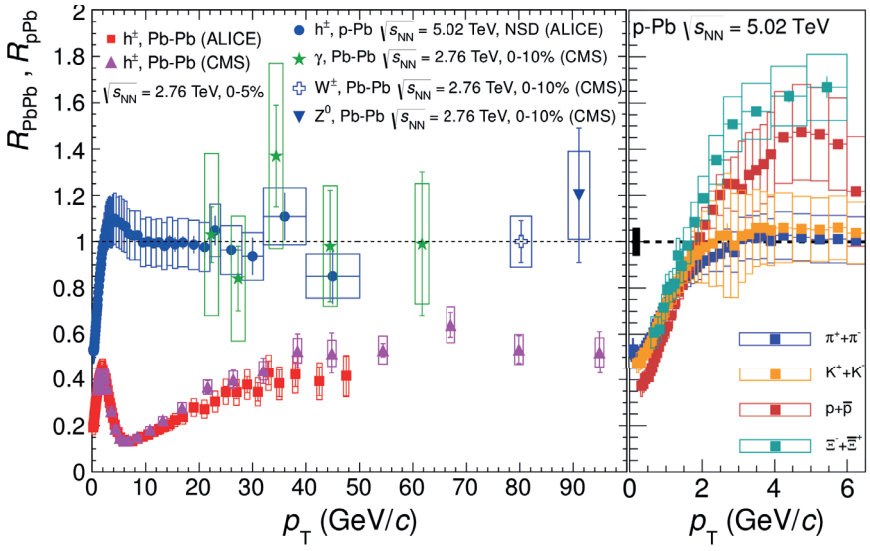


Figure 3: p_T dependent R_{AA} distribution for gauge bosons and charged particles for pPb and PbPb collisions. The right side of the histogram shows the R_{pPb} distribution of identified particles in the low-p_T range. The histogram is taken from ref. [26].

ion-ion (AA) type collisions than the final state of proton-ion (pA) type collisions.

2.3 Other observables

Other interesting observations related to the flavour production are the strangeness enhancement and the quarkonium suppression.

Strangeness enhancement was predicted to be observed in relativistic HI collisions [27, 28] when compared to pp collisions. This was observed first in the NA35, and later at the RHIC, and the LHC HI experiments. The supporting argument is that the QGP temperature is of the order of the strange quark mass (m_s), and because of that s and $\bar{s}$ production should not be suppressed even though the bare mass of the strange quarks is higher than the u and d quark masses; $m_s \gg m_{u,d}$. The abundance of gluons produced in HI collisions is assumed to be the primary source of strangeness enhancement through $s\bar{s}$ pair production.

The Quarkonium sector is explored in HI collisions to study the production and suppression of the heavy quark flavours. The quarkonia are $q\bar{q}$ states where q represents either the charm or the bottom quark. The top quark is excluded here; due to the very short lifetime of top quarks, the $t\bar{t}$ mesons can not be formed. The suppression of the heavy bound state

J/ψ ($c\bar{c}$) in HI collisions has been predicted in theory [29] as a sign of QGP formation. The theoretical argument is based on the electric charge screening effect, where the strength of an electric potential between two test charges weakens if the two charges are surrounded by other electric charges or medium. Similarly, if a quarkonium state has low binding energy then the colour charge of the medium will dilute the binding force between the two quarks and break the bound state. J/ψ suppression in PbPb collisions is observed in the NA50 experiment at the SPS [30], at RHIC [31], and at LHC [32]. The Υ-mesons ($b\bar{b}$) states, $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$ are expected to provide a measure of the QGP temperature based on the relative suppression of the three states. The experimental observations [33, 34] in PbPb collisions show that the suppression is observed the highest for $\Upsilon(3S)$, sequentially lower for $\Upsilon(2S)$, and then relatively less for $\Upsilon(1S)$.

Quarkonium production is a hard interaction process, and it is safe to argue that most of the quarkonium would have been produced in the collision before the QGP formation. Therefore the quarkonia study provides unique observables as perturbative probes to study the production of heavy-quark pairs in the medium, and as probes to understand the non-perturbative mechanism of heavy hadron production. Quarkonia can be studied for all three types of hadronic collision systems i.e. pp, pA, and AA collisions. pp collisions give an insight into quarkonium production in the vacuum, and pA and AA collisions can provide insight into the medium effect (if there is any) on quarkonium production.

3 Event generators

In the previous sections, I have mentioned "event generators", which are used to simulate particle collisions. In this section, I will discuss the event generators in particle physics in general and some of the key components of the Pythia event generator in more detail.

A particle collision produces many new particles, and the process starting from the colliding particles to the production of the resultant particles is called an "event". Different particles have different production rates depending on the interaction cross-sections. The theoretical calculations can provide the probability distributions, and the numerical algorithms use random number generators to select a certain interaction with a certain probability to produce an outcome of a particle collision event. Henceforth the numerical algorithm which performs such calculations is called the event generator or Monte Carlo models. One can not predict all possible interactions by theory calculations, and we have to rely on experimental results to obtain an appropriate value for those unpredictable parameters. The results of the event generator are based on a theoretical understanding of the experimental observations in particle physics.

Particle physicists use these algorithms to analyse experimental data, understand detector

efficiencies, distinguish signal and background events during the data analysis, and verify theory predictions. Moreover, the event generators are also used to design new particle detectors and collision experiments.

3.1 Hydro-dynamical approach

Conventionally, the QGP production and expansion are described by the macroscopic theory of microscopic interactions. Properties of a macroscopic system depend on the initial condition, interactions of constituent particles, conservation laws, and external constraints[12]. A hydrodynamic model [35] of the QGP evolution in space-time can be constructed based on estimates of the equilibrium properties of the QGP, using models inspired by the kinetic theory of gases, distribution functions, and transport theory.

The details of the relativistic hydrodynamic model building are out of scope in this thesis. But I will discuss here a simplified picture of various stages of the HI collisions starting from the instant of collision until the free propagation of final state particles from the hydrodynamic point of view. The HI events have four stages in the hydro-dynamical picture: a) Pre-equilibrium stage, b) Expansion of the QGP, c) Freeze-out stage, and d) Final state hadron scattering.

Pre-equilibrium stage is just after a HI collision when the QGP is just formed and it is in a non-equilibrium state. The partons are correlated strongly, and they are continuously interacting with each other, while the system is expanding and achieving equilibrium. The time taken to achieve local equilibrium is called the thermalisation time of the medium, which is assumed to be within the time scale of $1\ \mathrm{fm/c}$ in the medium rest frame.

The momenta of partons generate a pressure gradient during the **expansion stage**. Hence depending on the magnitude of the pressure gradient, different regions of the QGP medium expand at different rates. The medium expansion reduces the system density and the system cools down, and as the energy density of the system reduces below a critical density, $\rho_{crt} \approx 1GeV/\mathrm{fm}^3$ [36], partons start to hadronise. A mixed state of hardons and the free partons coexist during the initial stage of the hadronisation, but the mixed phase vanishes after the transition from QGP to hadron gas, and the system is left with hadrons only.

The **hadron scattering** and the **freeze-out stage** are connected to each other. The freeze-out stage is sub-divided into *chemical freeze-out* and *kinetic freeze-out*[13]. The produced hadrons initially undergo inelastic scatterings and keep the expanding system in equilibrium. The chemical freeze-out point has been reached when the probability of inelastic collisions re-

[12]Density, temperature, entropy, etc. are characteristics of such macroscopic systems.

[13]Both of these freeze-outs are assumed to occur at certain fixed temperatures ($T_{chemical} > T_{kinetic}$) for all quark flavours. From the freeze-out stage Cooper-Frye prescription [37] is used to obtain an invariant distribution of particles.

duces to almost zero. Prior to the chemical freeze-out, hadrons are producing new hadrons through inelastic scatterings, but after the chemical freeze-out only elastic scatterings occur and the types of hadrons remain constant. As the system expands further, inter-hadronic distances grow larger than the strong interaction range, and the probability of elastic collisions gradually reduces. At this stage, it is not possible to maintain thermal equilibrium, and the system reaches the kinetic freeze-out phase. The hydrodynamic picture can not survive anymore. Hadrons travel freely in the vacuum after the kinetic freeze-out, until they decay into more stable particles or are considered as measured by the detectors.

3.2 Non-thermal models

Apart from hydrodynamic models, there are microscopic models that do not consider temperature or thermalised medium creation in the simulation of particle collisions. All conventional general-purpose event generators for pp-like collision systems, such as Pythia, Herwig [38], and Sherpa [39] etc., belong to this category. The Angantyr model is developed as an extrapolation of Pythia for HI collision simulations without considering a QGP formation. Therefore, the Angantyr model is also a non-thermal model for HI event simulation. In non-thermal models, the post-collision scenario is described based on partonic scattering and radiations in a vacuum. Moreover, unlike hydrodynamic models, where a collision event has four discrete stages, in non-thermal models, the event evolves continuously with various stages, but without any thermalised medium creation.

The final state observables in high multiplicity pp collisions show signatures of *strange- ness enhancement* and a *near side ridge* similar to the HI collisions. String fragmentation is the primary technique for hadron production (see section 3.3) in the Pythia event generator. A model to explain these new observations by introducing the colour string interactions has been included in Pythia. Colour strings in the default Pythia do not interact among themselves except during the *colour reconnection* stage (see section 3.3), but in the light of the new LHC results, an idea emerged that the accumulation of many colour strings due to high parton density can modify the effective string tension in high multiplicity events.

Strings are usually stretched longitudinally, but they also have width, and there may be some overlap between two neighbouring strings in the dense environment of the high multiplicity events. This idea is now implemented in Pythia as *rope hadronisation* [7, 8] and *string shoving* [40, 41] mechanisms for pp events.

The string shoving mechanism gives ridge-like effects in pp collision simulations. Here, the transverse push to the nearby colour strings in the outward direction is generated due to the colour field of neighbouring strings. So far it handles only low p_t strings in Pythia with an assumption that all strings are parallel to each other and to the beam axis as well. In a recent paper [41], non-parallel strings are also included to perform the string shoving

mechanism.

The rope hadronisation is a new model [7, 8] in Pythia. According to this model, the string density will modify the effective string tension (κ) in Pythia. This modification affects the relative yield of strange quarks, because during the string fragmentation the relative strangeness production probability depends on κ as:

$$\rho \propto exp\left(\frac{-\pi(m_s^2 - m_u^2)}{\kappa}\right), \tag{7}$$

where ρ is the relative production probability of strange quarks of mass m_s over up quarks of mass m_u. The increment in the value of κ increases the strangeness production probability (ρ), and κ is modified due to the colour rope formation. This way, Pythia can reasonably reproduce strangeness enhancement as a consequence of higher string density in the high multiplicity pp events. The same ideas may be implemented in the Angantyr model for HI events to explain similar observations of HI collisions without assuming the existence of a QGP.

3.3 pp event generator

Pythia, Herwig, and Sherpa are widely used event generators for pp like small system collisions. These event generators are called general-purpose event generators because they can reproduce a wide range of observables with a detailed picture of the collision. One can find details about the different modules of these generators in [42].

I have used the Pythia model and developed some of its modules during my thesis work, so I will give a short overview of the Pythia model and focus more on the modules I have worked on. Pythia has been proven as an excellent phenomenological model to reproduce most of the pp physics observations, with good accuracy. The core idea in Pythia's phenomenology for event simulation is based on collisions that occur due to multiple parton interactions and parton showers, and the hadrons production based on the colour string evolution and the string fragmentation to produce final state hadrons.

Pythia events are generated by many different modules and they reproduce various stages of a pp event (see Figure 4). Every module is developed to mimic specific physics mechanisms of the collision event. The Pythia modules can broadly be divided into two categories: perturbative and non-perturbative. Three major perturbative modules in Pythia are: a) Parton Distribution Functions (PDFs), b) Multi-parton interactions (MPIs), and c) Parton showers (initial and final state radiation).

Although the proton is identified with its three valance quarks (uud), it contains many partons other than the valance quarks, e.g. sea quarks and gluons. These partons are virtual

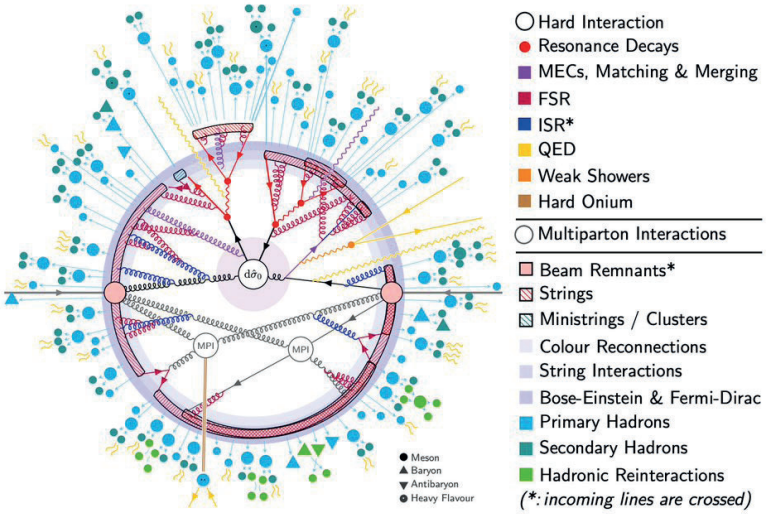


Figure 4: Illustration of a pp collision event modelled in Pythia. Different modules of Pythia are depicted in the event sketch with relevant legends. The two solid lines highlighting "Multiparton Interactions", are also the ones that separate perturbative QCD dependent modules and non-perturbative modules in Pythia. The image is from [2].

fluctuations, which can only be exposed if the proton can be scattered by a particle with high enough energy. This energy can be called the probing scale Q^2. For simplicity, one can consider Q^2 as a resolution scale for an image. One can only catch certain details of the image at a given resolution scale. If one increases the resolution, then more finer details of the image can be revealed to the observer. Similarly, for a given probing Q^2 scale, a probability distribution of different types of partons carrying specific momentum fractions x of the proton's total momentum can be obtained, which is $f_i(x_i, Q^2)$ and called the **Parton Distribution Function (PDF)**.

That means a pp collision is an interaction between two partons, one from each colliding proton. If one can obtain PDFs for the colliding protons then one can calculate the interaction probability, which is the interaction cross-section (σ) for a pp collision.

$$d\sigma = \int dx_a \int dx_b \; f_a(x_a, Q^2) \; f_b(x_b, Q^2) \; d\sigma_{ab\to X}. \tag{8}$$

Here $\sigma_{ab\to X}$ represents a partonic cross section for two partons, a and b, from each colliding proton to interact and produce partons X, and can be obtained from perturbative calculations. However, the PDFs are scale-dependent and only the evolution between two

scales can be calculated. Therefore, the shape of the input distribution from which we can calculate the evolution is fitted to experiments.

Since the proton sub-structure is more than just three valance quarks, it is also possible that multiple partons from each colliding protons can interact with one another. Usually, the interaction with the highest energy exchange is counted as a *hard scattering*. The hard scattering is often the interaction where heavy particles or jets are produced. The other partonic scatterings are called the **multiple parton interactions** (**MPIs**). These MPIs are treated the same way as the hard scattering, but their p_T scales are lower. Every MPI will contribute to the final state event multiplicity. That means the pp collision event producing many particles typically also has to have many MPIs.

The partons have a colour charge, and during the collision, they emit gluons due to the bremsstrahlung effect. These emissions can be calculated as the splitting probability of the parton, where a parton splits into two partons ($q \rightarrow qg$, $g \rightarrow gg$ etc.). The emitted partons can further split into more partons, and a shower of such emitted partons occurs. It is called **Parton shower** (**PS**). The PS can be possible before scattering or after the scattering. They are distinguished as **initial state** and **final state** radiation in Pythia.

For a reader who wants to learn more about the finer details of the above-mentioned and other different modules in Pythia, a good starting point is [2] and references therein.

At low energy, the QCD coupling (α_s) becomes large, and perturbative calculations can not be used to obtain interaction probabilities in particle physics. Hence no direct solutions are available from the QCD first principles for such a non-perturbative scale, and we require phenomenological models to approximately reproduce the experimental results. In Pythia, **colour reconnection** and **hadronization** are the two such non-perturbative modules. During my thesis work, I have made some improvements in these two modules. Hence I will discuss these two modules in more detail.

Colour Reconnection

The MPIs in Pythia draw multiple colour lines from each of the colliding beams, which are connecting the beam remnant and the produced partons. These colour lines stretched between the partons will contribute to the number of produced hadrons. The hadronization of the produced partons is treated as a universal process and independent of how the partons are produced in the Pythia model. Moreover, if no other effects would exist then one can naively think that the contribution to p_T would be independent of the event multiplicity. However, experiments show enhancement in the average transverse momenta with an increase in the event multiplicity ($\langle p_T \rangle$ (N_{ch})).

The discrepancy between Pythia output and experimental results for the event multiplicity

and $\langle p_T \rangle\,(N_{ch})$ of the hadrons indicates some collectivity among the produced hadrons. In Pythia, the partons produced after MPIs and PS are colour-connected by colour strings stretched between them and also between produced partons and the beam remnants; quarks are connected to one colour string, and gluons to two colour strings[14].

The idea of assuming that the produced partons are always colour-connected comes from the concept of *QCD confinement*. The QCD confinement points to a unique property of the strong force. It shows that, unlike the particles with only electric charges, the colour-charged particles can not be found as free particles at lower energies. That means we can have a free electron, which has only an electric charge, but we can not have a free quark, which has a colour charge and an electric charge. Hence the colour charged particles are always connected with an anti-colour charged particle. A simple analogy is of a magnet. One can not have a magnet with just one magnetic pole. Similarly, the colour field can be treated as confined to a fluxtube connecting the two colour charges. This colour fluxtube is treated as a massless relativistic colour string in Pythia.

Unlike $\mathrm{e^+e^-}$ collisions, many colour strings are produced in the MPIs and PS in pp collision events. Hence one can think about, what would be the right choice for the colour connections among the partons. Consider Figure 5 (a) where on the left, two quark–anti-quark dipoles are shown, and on the right, a different configuration between the same four partons is shown. There is no direct way to tell which configuration exists in an actual collision event.

In Pythia, such a decision is made by using an algorithm that rearranges the colour connections between the partons, and it is called **Colour Reconnection** (**CR**) model. The common practice in different CR models is to minimise the net string length (λ-measure) between the two partons. Different models use different definitions of λ-measure and also constrain which partons are allowed to undergo reconnections. Pythia's CR model modifies the colour between the colour dipoles in such a way that it conserves the total colour charge but reduces the net string length in the event. In the end, CR reconnects collinear partons, which generate shorter but boosted strings. The amount of CR is inversely proportional to the event multiplicity in the final state. As a result, CR not only reduces the number of produced hadrons but also increases $\langle p_T \rangle\,(N_{ch})$, and both the results are in agreement with the experiments.

The CR model that plays an important role in this thesis is the QCD-inspired model [43] (QCDCR model) for colour reconnection in Pythia. This model introduces additional constraints based on the colour algebra on the colour dipoles to be reconnected.

In the default CR, the only constraint is to minimise the dipole length by deciding recon-

[14]The colour strings represent colour flux tubes, (anti-)quarks carry an (anti-)colour charge. In contrast, a gluon carries both a colour and an anti-colour charge.

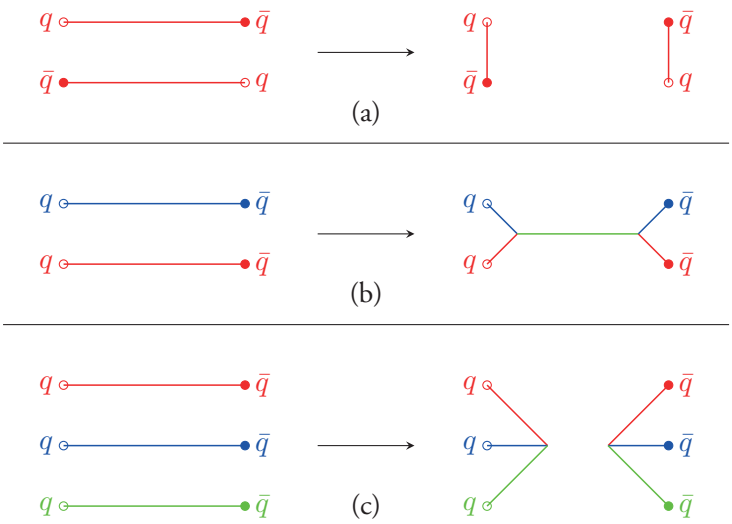


Figure 5: Two dipoles and three dipoles CR possibilities. For two dipoles, they can either have (a) a simple reconnection or (b) a formation of a colour connected junction and anti-junction system. Three dipoles can form (c) disconnected junction and anti-junction systems.

nections like Figure 5 (a). Here if the new configuration (on the right side of figure 5 (a)) reduces the net dipole length then the algorithm favours the new colour connections over the previous configuration ((on the left side of figure 5 (a)).

The QCD CR model allows reconnecting colour dipoles if the QCD colour rules are satisfied. Here, if the two dipoles have exactly the same colours then only they can perform "swing" and reconnect to the closest anti-colour ends of the other dipoles, which is similar to the default CR model. But at the same time, the colour algebra allows a possibility of three colour string pieces being connected to a single "junction" point and forming a colour singlet system. Hence two or three colour dipoles can also reconnect (see Figure 5), and two new configurations (Figure 5 (b) and (c)) of colour singlet systems can be formed. These string systems containing at least one junction point are called *junction systems*. For more technical details, I suggest reading either paper II or Ref. [43].

These junction systems are always produced in a pair of a junction–anti-junction, and each contribute to the baryon production later during the hadronization, which I will discuss in the next subsection.

Hadronization

After the CR, the colour singlet strings fragment and produce primary hadrons according to the Lund string fragmentation model (LSM) [6] in Pythia.

The simplest scenario is a quark–anti-quark connected with each other by a colour string. In the LSM the string can break up from either end of the string with equal probability. Hence, once the string breaks at some point into a $q\bar{q}$ pair, it forms a hadron using a quark (or anti-quark) from the pair and a parton from the string end. The remaining quark (or anti-quark) from the pair will act as the new end of the remaining string piece. The energy sharing between the hadron and the new string piece is defined by the fragmentation function:

$$f(z) \propto \frac{1}{z}(1-z)^a \ \exp\left(\frac{-b\,m_\perp^2}{z}\right), \tag{9}$$

where a and b are free parameters, and z is the energy fraction taken by the hadron.

At first quark and anti-quark in the pair are produced at the same point on the string as massless virtual particles. They come on-shell by consuming the energy stored in the string according to the *tunnelling mechanism*.

Since the tunnelling process requires string energy to bring the $q\bar{q}$ pair on-shell, it suppresses the production of massive quark pairs by $\approx \exp(-\pi m_\perp^2/\kappa)$, where $m_\perp$ and κ are transverse mass of the quark and string tension respectively. Hence heavy quarks (charm, bottom and top) have a very low probability of being produced in string fragmentation and one can safely considered them not being produced in string fragmentation in Pythia. Therefore, one can say that the hadrons with heavy quarks can only be produced in Pythia if the heavy quarks already exist at a string end.

One (anti-)quark of virtual $q\bar{q}$ pair forms a hadron with the string end, and the remaining string piece will undergo further fragmentation until all the string energy is consumed in forming hadrons. The fragmentation sequence of $q\bar{q}$ pairs will produce only mesons if the string was originally stretched between a quark and an anti-quark at the beginning of the fragmentation.

To have baryons the string has to fragment by a diquark–antidiquark breakup, where a diquark (or antidiquark) forms a baryon (or anti-baryon) with one of the string ends and the other becomes the end of the leftover string. A diquark can be considered as a combination of two quarks in some effective state, where according to the QCD colour rule it carries an anti-colour similar to an anti-quark. The baryon number is always conserved because the diquark and anti-diquark are always produced in pairs. Diquark being an effective state of two quarks is more massive than that of a quark, which means its production is suppressed in the fragmentation. Hence the overall baryon production is suppressed in the

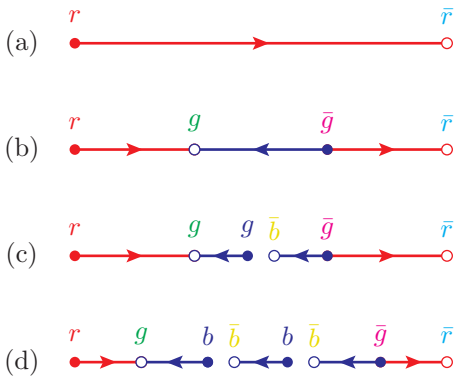


Figure 6: Illustration of the popcorn mechanism. In (a) no fluctuation has occurred, and a full string is spanned between a red–antired $q\bar{q}$ pair. In (b) a green–antigreen pair has appeared on the string as a quantum fluctuation. If the red and green quarks form an antiblue triplet, this reverses the colour flow in this part of the string, and the net force acting on the green quark is zero. In (c) the string breaks by the production of a blue–anti-blue $q\bar{q}$ pair, resulting in two string pieces with diquark ends. In (d) another breakup in the blue triplet field results in an additional meson. The image is from (paper III).

string fragmentation.

The baryons produced by diquark pairs always share two quark flavours and are produced next in rank (next to each other along the string). Such a baryon production is close in rapidity and will give strong correlations that are not found in experiments (this also gives background to paper III). The experiments show that such a baryon distribution does not always happen. Hence a new mechanism is developed, the *popcorn mechanism*. Figure 6 shows a simple scenario of how the popcorn mechanism produces a baryon anti-baryon pair with a meson between the two barons. This mechanism allows a multiple $q\bar{q}$ pair production as a result of a vacuum fluctuation without breaking the string. Later, the string fragments at multiple points as shown in Figure 6 (d), and produce the baryons with a meson in between (for technical details see section 2 in paper III).

There is an alternative way for baryon production by contribution from the junction systems. The junctions are formed either close to the beam remnant or by three colour dipoles connected to a junction point in the QCD CR model. Figure 7 illustrates the fragmentation of the junctions in two stages. The junction legs are distinguished as low, middle, and high-energy legs by calculating the sum of the energies of the partons on each leg. The first two low-energy junction legs are fragmented, and the partons closer to the junction point are combined to form a diquark. Later, the remaining junction leg is fragmented as a string having a diquark at one end and a quark at the other end. Hence every junction sys-

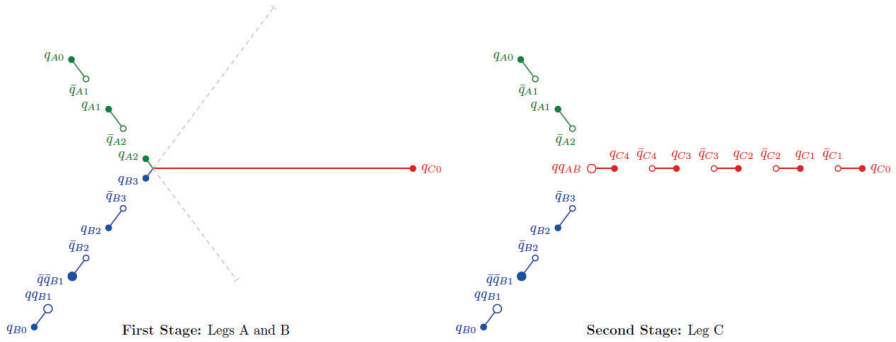


Figure 7: Illustration of the junction fragmentation mechanism.**Left:** At first the low and middle energy legs are fragmented. **Right:** The partons from the low and middle energy are combined to form a diquark connected to the remaining high-energy leg, which fragments at last and the junction system produces at least one baryon. The image is from [2].

tem produces at least one baryon (or anti-baryon) in addition to the baryons produced due to the regular string fragmentation. Since junctions are always produced in the junction–anti-junction pairs in the QCD CR model, they always conserve the baryon number of QCD.

In high multiplicity events, it often happens that a junction system has very low energies in its legs and it can not fragment by the regular string fragmentation. I have introduced a new mechanism called *junction collapse*, where depending on the types of the quarks or diquarks (and respective anti-particles) the junction will try to produce two hadrons (see Appendix in paper II). This mechanism is inspired by Pythia's handling of the short strings, where the string produces one or two hadrons.

So far I have discussed scenarios where a $q\bar{q}$ pair is directly connected by a colour string piece, and quarks or diquarks in junction systems are connected to the junction point. But there are also gluons produced in Pythia events. Unlike (anti-) quarks, gluons are connected by two string pieces as they carry a colour and an anti-colour charge. Hence they are often connected by a chain of gluons or one or more gluons connecting a quark and an anti-quark. They can also separate two junctions by a long chain of gluons in between, which will provide an additional separation between the two junctions.

A gluon is represented as a "kink" in the transverse direction on the string connecting a quark and an anti-quark. The direction of the momentum vector changes on either side of the gluon kink due to the strings being stretched by the gluon momentum. Hence the baryons produced due to the popcorn mechanism near gluons gain non-trivial angular

correlations. We have studied these effects in pp collision events. We show that such a baryon production should be suppressed to improve the results of angular correlation in the baryon sector (see paper III).

The hadrons produced by string fragmentation are primary hadrons. Many of them are unstable hadrons with varying lifetimes. All these unstable hadrons decay according to their respective lifetime and experimentally measured decay channels.

Later, the hadrons can interact with each other due to **hadronic rescattering** [44, 45] in Pythia. The hadronic rescattering adds low-energy interactions between hadrons moving in similar directions in the Pythia/Angantyr model. The model traces the linear path of the produced hadrons, and at some time t (in the hadron pair's rest frame) if any two hadrons cross each other's path then depending on the interaction cross section for the given hadron pair they can have either elastic or inelastic scattering. In the current implementation, it includes $2 \to 2$ and $2 \to 3$ processes only, which modifies the p_T of the identified hadrons and slightly enhances the event multiplicity.

3.4 Heavy-ion event generators

HIJING [46–48], AMPT [49, 50], and EPOS [51, 52] are widely used event generators for HI collisions. Angantyr is developed with the idea that the model stand alone can simulate HI events. I chose to mention the above-mentioned three models[15] because, unlike other hydrodynamics-based simulation models, these three models can also simulate complete HI events without any additional initial or final state modelling outside of the event generator.

I start with an overview of the Glauber model then I will provide a short description of these three event generators before discussing the Angantyr model in more detail.

Glauber modelling

A realistic calculation to obtain N_{coll} and N_{part} requires information about the position of the nucleons inside the colliding nuclei. N_{coll} is the number of inelastic nucleon-nucleon sub-collisions, and N_{part} is the number of nucleons that have had at least one inelastic sub-collision. The Glauber model [9, 10] provides an arrangement of nucleons inside the nucleus according to the Woods-Saxon distribution of nucleon density:

$$\rho(r) = \frac{\rho_0}{1 + \exp\left((r - R)/a\right)}. \tag{10}$$

[15]Recently, a model the dynamic initialisation based on core-corona (DCCI2) has been developed, which can simulate pp and heavy-ion events, but I will not discuss it here. A curious reader can read about it in Ref. [53] and references therein.

Here ρ_0 is the central density, **r** is the distance of a nucleon from the centre of the nucleus in 3-dimension, R is the radius of the nucleus, and a is surface diffusion or "skin width" parameter.

The probability density of finding a given nucleon at some transverse position **s** inside the projectile nucleus is $T_{proj}(\mathbf{s}) = \int \rho_{proj}(\mathbf{s}, z)dz$. The interaction probability between the nucleons from the projectile and the target can be calculated from the nuclear overlap function T(**b**) for the given impact parameter (**b**) of the collision:

$$T(\mathbf{b}) = \int T_{proj}(\mathbf{s})\, T_{tar}(\mathbf{s} - \mathbf{b})\, d^2\mathbf{s}, \tag{11}$$

where "**s**" is the position of a projectile nucleon from the centre of the nucleus and "$\mathbf{s} - \mathbf{b}$" is the relative position of a target nucleon in the incoming nucleus. The overlap function is normalised as $\int d^2\mathbf{b}T(\mathbf{b}) = N_{proj}N_{tar}$ by integrating over all impact parameters, where N_{proj} and N_{tar} are the number of nucleons in the projectile and the target nucleus respectively.

A Glauber model provides N_{coll} and N_{part} according to the geometric overlap between two colliding nuclei for the given impact parameter **b**, and the nucleon-nucleon inelastic interaction cross-section σ_{inel}^{NN}. The overlap function describes the combined density of colliding nucleons in the projectile and the target nuclei in a HI collision with impact parameter **b**. For a given σ_{inel}^{NN}, the average number of nucleon-nucleon interactions is $T(\mathbf{b})\sigma_{inel}^{NN}$.

Nucleons are assumed to travel in a straight line without deflection at relativistic velocity, so the participants can interact with multiple nucleons from the target nucleus while penetrating through the incoming nucleus. Experimental analyses show that the rate of hard interactions is proportional to N_{coll}, while the charged particles multiplicity scales better with N_{part} [54]. This correlation implies the need for sophisticated handling of N_{part}.

All HI event generators use a Glauber calculation to obtain N_{coll} and N_{part}. The significant difference among the HI event generators occurs in the process of their treatment of the underlying scatterings. In the Angantyr model, we distinguish the type of each individual collision while obtaining N_{coll} and N_{part}. Moreover, we have included the fluctuations both in the positions of the nucleons inside the nuclei and the fluctuations in the nucleon-nucleon interaction cross section, as discussed further in section 3.5.

HIJING

The Heavy Ion Jet INteraction Generator (HIJING) model has two primary components: *mini-jets* and *soft interactions*. In HIJING parton interactions are classified into semi-hard or soft. A semi-hard process has at least one pair of jets with $p_T > p_0$. Partonic interactions

with $p_T < p_0$ are treated non-perturbatively and characterised by soft parton cross section (σ_{soft}).

HIJING uses Pythia for the kinematics of jet pairs and associated initial and final state radiations, and for the hadronisation of partons according to the Lund string fragmentation model. HIJING has an impact parameter-dependent parton structure-function to study the effects of nuclear shadowing on the observables. The jet quenching in HIJING is modelled for moderate and high p_T observables, and it is based on assumed energy loss per unit distance ($\frac{dE}{dx}$) of the particles travelling through the medium.

AMPT

The AMPT (A Multi-Phase Transport) model is primarily used to study the flow (collectivity) in HI collisions at RHIC and now also at LHC. It has four components: 1) Initial states, 2) Parton cascade, 3) Hadronisation, and 4) Hadron cascade. It has two modes of event simulation: 1) Default mode and 2) String melting mode. The string melting mode is developed assuming a QGP formation.

The AMPT uses the initial particle distribution from HIJING. The model performs parton-parton scattering according to the Zang Parton Cascade (ZPC) [55] model. For hadronisation, in the default mode, the AMPT uses the Lund string model, and in the String Melting mode, it uses a quark coalescence model [56]. The AMPT string melting model assumes the production of quarks and anti-quarks only, which allows it to have consistent recombination into hadrons by the coalescence model. The coalescence allows every $\bar{q}$ and q to choose partners from surrounding quarks and anti-quarks based on their separation from each other, but a parameter r_{BM} controls the relative formation of baryons over mesons. The ART(A Relativistic Transport model [57] performs the final state inelastic and elastic scattering of the produced hadrons in both modes.

EPOS

EPOS is a semi-hydrodynamic event generator. Hydrodynamic interactions are restricted as a part of the *core* and hydro-independent interactions are part of the *corona*. The separation of the core and the corona is based on string density at some proper time τ_0; the core is the region with string density higher than a certain value ρ_0, and the corona is the region with lower string density [58]. The model treats scatterings according to a Gribov-Regge theory [59] of parton ladders stretched between two colliding nuclei. These parton ladders include the initial evolution of partons from projectile and target nucleons. In the core region, the model generates MPIs and the parton density is high, which leads to a higher number of parton ladders. The corona region has fewer MPIs which results in a low density

of partons.

The hadronisation in the core is performed using a statistical model. The corona is low density, which typically means in the peripheral regions in the transverse plane and far away in rapidity, and here hadronisation is done similarly to Pythia's string fragmentation mechanism. The primary difference between the string fragmentation mechanisms in Pythia and EPOS is in their treatment of the beam remnants. In Pythia, the two beam remnants are connected through the strings extended by the produced partons, but EPOS treats beam remnants individually and independently from the produced partons.

Unlike HIJING and AMPT, EPOS can also generate pp events in addition to the HI events. The core appears only if the local density of strings is higher than ρ_0. Hence, for the high multiplicity pp collision events, at mid-rapidity particles are produced through the core, otherwise, they all are produced according to the normal string fragmentation.

3.5 Angantyr

The motivation to develop Angantyr was to test the limits of an extrapolation based on pp collision dynamics to describe the observed HI data without considering any thermalised QGP medium during the event simulation. The Angantyr model is a generalisation of the previous work [60] done for pA to AA. This new model has improved the mechanism for sub-collisions (nucleon-nucleon collisions) in heavy-ion collision events.

The multiplicity distribution in pPb collisions shows that the projectile proton participates in multiple sub-collisions with nucleons from the target Pb nucleus. Moreover, the asymmetric multiplicity distribution as a function of pseudo-rapidity in pA collisions discards the possibility of a mere superposition of many independent pp-like collisions to generate an HI event. Hence, if one wants to stack multiple pp events to generate a HI event, it will require a sophisticated treatment of superposition of several pp-like collisions.

The nucleons are called "participants" above and calculated using the Glauber model, they are also called "wounded nucleons". This terminology is e.g. used in the old FRITIOF model [61, 62], which assumed that the event multiplicity is approximately proportional to the number of such wounded nucleons.

Glauber's original calculation did not include the effects of fluctuations in the scattering process. They are quite important, especially in pA collisions. Fluctuations in the nucleon position within a nucleus are accounted for by event generator simulations. Gribov pointed out that it is important to account for diffractively excited nucleons as intermediate states in multiple NN sub-collisions [63]. In Angantyr we use the Good-Walker formalism [64] to describe diffraction as a result of fluctuations in the projectile and target wave functions.

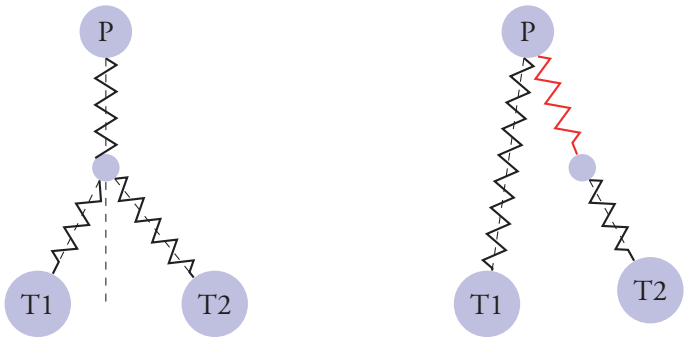


Figure 8: (Left) A schematic diagram of a projectile particle colliding with two target particles represented with a triple cut-Pomeron vertex. The cut-Pomeron vertex represents the non-diffractive interaction between the participants.
(Right) A schematic diagram of a projectile particle colliding with a target particle (T1) with a cut-Pomeron exchange (non-diffractive) interaction, and it is interacting with a target (T2) via a Pomeron exchange (single-diffractive) interaction.
(A cut-Pomeron is represented as a zigzag with a straight line, and an uncut-Pomeron is represented as a zigzag.)

The discrimination between different collision types when calculating the number of binary collisions is essential for a realistic simulation of HI events, because different types of collisions produce different multiplicity distributions in different pseudo-rapidity ranges. The old FRITIOF model did not distinguish wounded nucleons as diffractively or non-diffractively wounded. We modified the old FRITIOF model and introduced the classification of the wounded nucleons in different collision types (inelastic collisions: non-diffractive, single or double diffractive collisions) in the Angantyr model.

After obtaining the number of wounded nucleons and types of sub-collisions, the interactions are ordered based on the impact parameter between colliding nucleons. The sub-collisions are generated using Pythia's machinery as described in sections 3 and 4 of the paper I.

The event-by-event fluctuations are simulated by fluctuating nucleon radii, which influence the nucleon opacity and the interaction cross-sections at a given impact parameter. Moreover, even at fixed impact parameters, the final state event multiplicity fluctuates in HI collisions, and in the event simulation, we want to reproduce these fluctuations.

Secondary non-diffractive events

During a HI event simulation with the Angantyr model, a crucial part is when the Glauber calculation predicts multiple non-diffractive (ND) interactions between a projectile nucleon and several target nucleons or vice-versa. This is also the part in the Angantyr model, where I worked the most, so I am taking an opportunity to go into more detail about the physics description in the text below.

A schematic diagram of a three-particle interaction (a collision between a projectile and a nucleus with two nucleons) is shown in figure 8 (Left), where a projectile (P) is interacting with two target nucleons (T1 and T2). The diagram shows a triple cut-Pomeron[16] vertex to represent ND interactions between the projectile and two target nucleons.

Pythia is fundamentally a $2 \to n$ type event generator, which can have beams of electrons and/or protons only. Since we aim to use Pythia's machinery to extrapolate the pp collision dynamics for HI collisions, in the case of a three-particle collision (figure 8), we opt for simulating two collisions of two particles in such a way that the projectile is the participant in both collision events. The simulation of two ND events, in this case, is a non-trivial task because not only the vertex position of the triple cut-Pomeron can vary in rapidity, but also the momentum fraction shared between the two cut-Pomerons from the targets can vary.

One possibility is to ignore the ND interaction between the projectile and the target nucleon, which has a larger impact parameter, and only simulate the ND interaction with a smaller impact parameter. The effect of such a hard vetoing will reduce N_{coll} and N_{part}, and that will directly affect the final state particle production in HI events. We show the contributions of diffractive and multiple ND interactions for pPb and PbPb events in Figure 21 of paper-I, and it reflects that a large number of multiple ND interactions occur in the simulation of both types of events. Therefore, such a hard veto on ND interactions for larger impact parameters, if the participant nucleon is already identified as a participant in another ND interaction with a smaller impact parameter, will reduce the final state particle production excessively.

Since we can not ignore the Glauber predictions for the same nucleon participating in multiple ND interactions, we handle such situations by modifying the tags of collision types, and we re-tag the interaction with a smaller impact parameter as the primary ND collision. Once the primary ND collision is fixed, the other ND collision is re-tagged as the secondary ND collision. In a situation shown in figure 8, the impact parameter between the

[16]A cut-Pomeron diagram represents a non-diffractive interaction between interacting particles, while a single uncut Pomeron exchange is argued for the observed rapidity gaps in a diffractive scattering in pp collisions. The Pomeron is proposed in the Regge theory (which is a predecessor of the QCD) to explain hadronic cross-section at high energy. The Pomeron is treated as a mediator particle with quantum numbers of the vacuum. It has neither an electric charge nor a colour charge. A good reference for more details about the Pomerons in the Regge theory is [65].

projectile (P) and the target nucleon (T1) is smaller than the impact parameter between the projectile (P) and the target nucleon (T2). Therefore, the interaction between the projectile (P) and the target (T1) is the primary ND (figure 8 (Right)), and it is generated using Pythia's default ND mechanism.

Now, we have to handle the secondary ND interaction between the projectile (P) and the target (T2). The secondary ND sub-event is simulated with a new modified mechanism, which is detailed in section 5 of the paper-I. In a nutshell, a modification is performed in such a way that the rapidity span filled with the produced particles in the modified high-mass single-diffractive (SD[17]) collision should look similar to an ND collision.

At this stage, it is important to note that this is a non-trivial attempt to include secondary ND collisions in the HI event simulations driven by technical limitations, and we are not advertising that this is the only way to handle the multiple ND collisions involving the same participants. We want to come up with a mechanism, which uses Pythia's event generation setup and allows us to simulate secondary ND events. The secondary ND collision can be mimicked if we use the high-mass SD mechanism in Pythia because an ND-like interaction is simulated between the Pomeron and the proton, where the proton, which emits the Pomeron, remains intact but loses momentum.

Pythia's high-mass SD events are simulated with a Pomeron emission from the projectile and the interaction will occur between the Pomeron and the incoming target. The rapidity gap (Δy) in the diffractive event depends on the diffractive mass (M_{diff}) for the given centre of mass energy ($\sqrt{s}$) as $\Delta y = \ln(s/M^2_{diff})$. The particles are produced in the rapidity range between the Pomeron-proton interaction vertex and the remnants of the target proton. In the rapidity region close to the target proton the high-mass SD event on average looks similar to the ND event (Figure 5 in paper I).

In the situation of Figure 8 the projectile interacts with both of the target nucleons. The projectile proton is replaced with a Pomeron for the secondary ND collision. We modified the Pomeron PDF so that it looks like a re-scaled proton PDF, and the secondary ND proton-proton collision is simulated as a modified Pomeron-proton collision, while regular diffractive events are generated according to the default Pomeron-proton collision set-up in Pythia.

After generating the primary ND event, the energy and the momentum of the beam remnants are tracked. We assume that the Pomeron is emitted by the projectile proton's rem-

[17]The diffractive scatterings in Pythia are simulated according to non-perturbative soft Pomeron exchange. The diffraction in Pythia is separated into soft and hard diffraction. The soft diffraction is further subdivided into low-mass and high-mass diffraction based on the diffractive mass (M_{diff}) of the event. The diffractive event is classified as low-mass diffraction if $M_{diff} \leq 10GeV$, otherwise it is classified as high-mass diffraction in Pythia. The high-mass diffractive events are simulated using MPIs and Pomeron PDFs, which are similar to those in the non-diffractive event simulation in Pythia.

nants, and if these do not have enough energy left to produce the Pomeron, then we cannot conserve the energy and momentum of the system. Hence while generating the secondary ND using the new mechanism, and if the event does not conserve the energy-momentum then it is vetoed. The diffractive mass (M_{diff}) of a diffractive event depends on the momentum of the emitted Pomeron, and we can produce the Pomeron at different energies in the event simulation. We tested the influence of the energy-momentum conservation by allowing few retrials of generating previously vetoed secondary events by reproducing them. If the Pomeron has lower energy than the previously vetoed event, then it is likely to be accepted. The result of our test is shown in Figure 25 of paper I, but in the default setup, only one attempt to generate a secondary ND event for the given nucleon is allowed.

After generating all sub-collisions, the final state partons are stacked together by shifting the collision points of the participating nucleons with respect to the impact parameter of the nucleus collision. The production vertices of partons are shifted according to the relative shift in the positions of the collision points of the participating nucleons inside the nucleus at the time of the collision. The colour strings produced in the nucleon-nucleon collisions undergo colour re-connection (CR) within the same sub-collision event, and there is no colour flow between the two sub-collisions. Later, these strings hadronise by string fragmentation, every string breaks independently, and the produced hadrons are assumed to travel freely without any further interactions, and unstable hadrons decay into stable particles. Hence, in the default version, final state hadrons do not show long-range collectivity, and they also do not interact among themselves.

The Angantyr model at its current stage can reproduce final state multiplicity and momentum distribution quantitatively in different centrality bins for HI events. The model can be used to study the effects of the initial state fluctuations and non-flow contributions in HI collisions. We have compared Angantyr generated multiplicity distributions with other HI event generators and we are able to reproduce competitive results (see paper I).

Global colour reconnection

It is difficult to argue that the NN sub-collisions would remain independent and that the partons from different sub-collisions do not interact in a heavy-ion collision experiment. Hence a natural choice is to extend the default Angantyr model to include interactions among the partons from different sub-collisions. Since the partonic interactions are among the colour charges, and the strong force has a short range, we can not allow *all* the partons in a heavy-ion event to interact with each other.

The CR is the stage where colour dipoles from different MPIs interact with each other in pp collision events in Pythia. Hence we decided to use colour reconnection (see section 3.3) as the stage in the event simulation in Angantyr where the partons from different

sub-collisions can interact. We have decided to use the QCD CR model. Prior to my work, all the colour dipoles in a pp collision event were allowed to participate in the CR in the QCD CR model. We introduced an impact parameter dependent constrained for the colour dipoles to be colour reconnected. We stack multiple sub-collisions after the parton showers and perform colour reconnection on all the partons in a heavy-ion event (see paper II).

The Angantyr model is developed with the idea that we do not tune the model with heavy-ion collision results. The heavy-ion events shall produce the results as the natural outcome of the extrapolation of the pp-like events. Hence I tuned the spatial constraint and some of the Pythia parameters in pp collisions. The QCD CR model has an existing shortcoming of relatively bad p_T distribution of the produced hadrons in pp collisions, and the re-tuning of the parameters worsened the distribution. Our preliminary study showed that the technical handling of short dipoles requires improvements.

In the pA collisions, we still have some freedom to modify the secondary non-diffractive events. We have already made changes in the Pomeron PDFs and Pomeron-proton interaction cross-section for the secondary events in the paper I. This time we decided to modify the Pomeron Flux in such a way that it will contribute to enhancement in the multiplicity distribution of the secondary non-diffractive events. We decided to use the supercritical description of the Pomeron Flux, where the Pomeron Regge trajectory is parameterized as:

$$\alpha(t) = 1 + \epsilon_{pom} + \alpha^{'}(t), \tag{12}$$

which gives the mass distribution $dm^2/m^{2(1+\alpha(t))}$. We decided to vary ϵ_{pom} and set its value to -0.04, which is a parameter to modify the mass distribution in the diffracted system. Our choice enhanced the event multiplicity in pA collisions by 10% (see paper II).

The AA events are generated without further tuning, and we show that the event multiplicity in the central events is suppressed due to the global CR. The central events are highly dense with many colour dipoles, which results in a large amount of CR, and as I have already discussed in section 3.3, CR suppresses the event multiplicity. We also notice that the p_T distribution of the hadrons maintains its bad shape (from the pp collision) in AA results. Hence I hope that if we can improve the p_T distribution in pp events we can expect similar improvement for the results in AA events as well.

Production of heavy baryons

After extending the QCD CR model with spatial constraint, we have noticed that the λ-measure for the junctions in the QCD CR model does not account for the types of the parton in the colour dipole while calculating the possibility for CR. The default λ-measure calculates the average potential energy available for the hadronisation for the given dipole

in the junction rest frame. Hence it provides an indirect measure of the "string length". The default setup works well for the light parton, whose mass is negligible compared to its energy. But for a heavy quark, the quark mass has a non-negligible contribution to its energy, especially if it also has low p_T. That means that while calculating the λ-measure, we should use a different way to obtain a measure of the effective "string-length". We decided to use the rapidity span of the heavy quark in the junction rest frame as the λ-measure for heavy quarks (see paper IV). These changes enhance the probability of the heavy quarks being in junction systems. ALICE experiment has shown that the relative fraction of charm baryons in pp collisions increases compared to $\mathrm{e^+e^-}$ and ep results.

Moreover, we noticed that the junction systems (Figure 7) containing heavy quarks sometimes fail to obtain the junction rest frame. In such a system, a heavy quark with low momentum is connected to the junction point by a short colour dipole. Hence it is difficult to achieve the required 120° angle between all the junction legs so that the outward momentum vectors balance each other, which is necessary to fragment the junction legs in the junction rest frame [2]. Therefore we decided to improve the fragmentation frame selection for such cases (paper IV).

Since the heavy quark is very close to the junction point, assuming that the junction system fragments in the heavy quark's rest frame is a good starting point. If we select the heavy quark at rest, we can easily visualise the remaining two legs spread out in the opposite direction of the heavy quark. We also improved the algorithm to calculate low-, middle-, and high-energy legs. Our changes in the junction systems more often make the heavy quark leg a low-energy leg which does not fragment any further and forms a diquark with a parton from the middle-energy leg resulting in a heavy baryon.

The junction systems are a complex arrangement of the Lund strings. We know that the majority of the charm quarks are usually produced in high multiplicity events, and in such a scenario the strangeness enhancement can also occur even in pp collision events. The rope hadronization model is an ideal choice for the strangeness enhancement in Pythia/Angantyr, but it is yet not optimised with the junction systems. Hence we used a reasonably enhanced strange quarks production rate to improve Pythia's description of charm baryon production in pp collision events at various LHC energies. We maintained a good description of the yields of Λ_c^+ baryon and Λ_c^+/D^0 ratio in pp collisions while reproducing them in pPb collisions with our global CR tune (see paper IV), which was not possible using the default Angantyr model.

Baryon correlations

The angular correlation results from the collider experiments of the produced hadrons provide insight into modelling the hadronisation mechanism. The results from the ALICE ex-

periment for pp collisions show that the angular distribution of the baryon pairs depends on the types of baryons forming the pair. Pythia uses the LSM to fragment the colour strings, where the fragmentation mechanism is symmetric and independent of the parton production mechanisms. In paper III we show that the QCD CR model and hadronic rescattering can influence the angular distribution of the baryons in Pythia. We also highlight that gluons in the Lund strings affect the baryon production near gluon kinks.

Paper III is our first step towards solving a long-standing question about understanding the hadronisation mechanism. We believe that if our study progresses towards a better understanding of the hadron production, then it should be reflected as an observation of the suppressed baryon-to-meson ratio in the substructure of the gluon jets, a hypothesis we have proposed in paper III.

4 Conclusion and Outlook

At the beginning of my thesis, I proposed some questions as a part of the motivation for my thesis work. At this stage, let me look back and introspect where we stand in providing a phenomenological model with an alternative explanation of the conflicting observations made in pp and AA.

We have developed the Angantyr model based on pp collision dynamics, and it generates heavy-ion events that reproduce general features of heavy-ion collision observables. In paper I, we show that the Angantyr model reproduces pp, pA, and AA multiplicity distributions as a function of pseudo-rapidity in different centrality bins. The momentum distribution is also reproduced without any additional re-tuning to HI data. We also extended the model in paper II with spatially constrained colour reconnection, which allowed the interactions among the partons after the parton showers irrespective of the subcollisions in heavy-ion collision events. This treatment of global CR enables a more realistic heavy ion events simulation in the Angantyr model. Although there is no special treatment for strangeness enhancement, we show that the spatially constrained QCD CR model in Pythia/Angantyr enhances the strange baryon production as the event multiplicity increases (paper II). We also show that the description of the heavy flavour baryon production (paper IV) in the Pythia/Angantyr model can be improved with appropriate changes in the junction formation and fragmentation mechanisms. The angular correlations in the baryon sector have been studied and improved, and we proposed a hypothesis of a relatively low baryon-to-meson ratio for the gluon jet sub-structure (paper III), experimental validation or denial of our hypothesis will provide us with a direction towards a better understanding of the hadronisation mechanism.

In short, we have made significant progress in developing the Angantyr model. The model

not only reproduces some of the general features of heavy ion events but also shows a promising future in reproducing some of the heavy ion events' specific observables without any thermalised medium. It is now time to equip the Angantyr model with the string interactions and hadronic rescattering models to test the potential of the Angantyr model in reproducing a variety of the experimental observables by simulating events from pp to AA collisions with the same underlying principles and without any assumption of a QGP medium. I believe the work done in this thesis will act as a model with an alternative explanation related to some unique observations scientists have made at the RHIC and the LHC facilities for small and large colliding systems.

I would like to allow myself to look forward and speculate on scientific desires and technical extensions, which involve the work done in this thesis.

> There are two possible outcomes: if the result confirms the hypothesis, then you've made a measurement. If the result is contrary to the hypothesis, then you've made a discovery. - Enrico Fermi

The Lund tune

Along with the developments in this thesis, some other models have also been developed in the recent past. The rope hadronization and shoving models introduce a possibility of strangeness enhancement and correlations among the produced particles respectively in high multiplicity events. The hadronic rescattering model allows produced hadrons to undergo elastic and inelastic rescattering. It may be the time to merge all these models along with the recently developed global CR model for heavy ion collisions and provide a new tune not only for Pythia but also for the Angantyr model.

Fixing p_T distribution in Angantyr

The spatially constrained QCD CR worsens the ability to reproduce the experimentally measured p_T distribution of the produced hadrons in pp collisions in Pythia and also in heavy ion collisions in the Angantyr model. We have done some preliminary studies and figured out that the technical handling of the short colour dipoles in the QCD CR model requires improvements. We also know that the bad p_T distribution has emerged already in pp collisions and continued in heavy-ion collisions. One hope is that if it can be improved in pp collisions then the results in the Angantyr model will also be improved. Moreover, string interactions and hadronic rescattering can modify the momentum of the hadrons. Hence collectively these models may improve the p_T distribution.

AA event multiplicity in Angantyr

We observe that the global CR reduces the event multiplicity in the central AA collision events. To resolve this shortcoming, one possibility is not to fix the value of the spatial constraint in the CR at pp collisions. Instead, its value can vary depending on the average number density of the colour dipoles in the system or simply depending on the colliding systems. The idea is vaguely inspired by the electric field penetration depth, where the penetration depth varies depending on the screening charge.

Reproducing the flow coefficients

The string shoving and the hadronic rescattering models can contribute to long-range collectivity in high multiplicity events. The CR model cannot generate long-range collectivity among produced hadrons. However, it can restructure the orientation of the colour dipoles, which has a significant effect, especially in the radial flow. Hence the colour reconnection can influence the production of flow coefficients together with the shoving and hadronic rescattering models. It will be interesting to check how well all three models combined can perform in reproducing different flow coefficients in small and large collision systems with the Pythia/Angantyr model.

Quarkonium suppression

The observations of $c\bar{c}$ and $b\bar{b}$ mesons suppression in heavy-ion collisions are yet to be tested against the Angantyr model. I am more curious about the reproduction of the experimental observations of suppression in $\Upsilon(b\bar{b})$-mesons. The experimental results are explained as the consequence of the interaction of the bound states with the QGP medium. The widely accepted assumption in the heavy ion physics community is that the loosely bound Υ-mesons have b and $\bar{b}$ quarks farther separated in space compared to strongly bound states. Hence they would break more often due to the thermalised medium.

Since we don't have any thermalised medium in the Angantyr model, we need a different approach to explain the observations. The heavy quarks can only be produced in hard scattering or in parton showers in Pythia/Angantyr model. The QCD CR model can reconnect longer b quark dipoles either to shorter dipoles or to junctions. In either case mesonic states formed with widely separated $b\bar{b}$ quarks should be suppressed more than the ones that are produced closely in space. Moreover, the effect of CR is stronger in high multiplicity events, and one can expect that the global CR in Angantyr will have stronger effects in heavy ion collisions compared to pp collisions in Pythia. Hence Quarkonia suppression results can be a good observable to test our alternative approach of explaining the heavy ion observables.

Jet quenching

Jet quenching is another HI observable that has not been explored with the Angantyr model. Jet quenching is a complex observable to reproduce. A naive possibility from the work done in my thesis is to check the effects of the global colour reconnection because it can modify the jet sub-structure depending on the number of colour dipoles within its vicinity. Hence the effects of the global CR will be stronger on the jets produced in high multiplicity heavy-ion collision events.

At last, apart from providing an alternative phenomenological explanation to HI collision observations, the Angantyr model also opens a door for a new set of questions e.g. what can be the signature observations to support or disfavour either of the phenomenological approach for the HI collision? Is a thermalised medium created at all in the HI collisions? Is there a similarity between the string picture and the hydrodynamic description of particle production and the transport mechanism?

References

[1] Sjöstrand, T. *et al.* "An Introduction to PYTHIA 8.2". *Comput. Phys. Commun.* **191,** 159 (2015).

[2] Bierlich, C. *et al.* "A comprehensive guide to the physics and usage of PYTHIA 8.3". *SciPost Phys. Codebases,* 8. arXiv: 2203.11601 [hep-ph] (2022).

[3] Collins, J. C. & Perry, M. J. "Super dense matter: Neutrons or asymptotically free quarks?" *Physical Review Letters* **34,** 1353–2356 (1975).

[4] Et al. (CMS collaboration), V. K. "Evidence for collectivity in pp collisions at the LHC," *Phys. Lett.B.* **765.** arXiv: 1606.06198 (2017).

[5] Et al. (ALICE collaboration), J. A. "Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions". *Nature Phys.* **13.** arXiv: 1606.07424 (2017).

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

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

[8] Bierlich, C. "Rope Hadronization and Strange Particle Production". *EPJ Web Conf.* **171,** 14003. arXiv: 1710.04464 [nucl-th] (2018).

[9] Glauber, R. "Cross Sections in Deuterium at High Energies". *Physical Review* **100,** 242–248 (1955).

[10] Miller, M. L., Reygers, K., Sanders, S. J. & Steinberg, P. "Glauber Modeling in High Energy Nuclear Collisions". *Ann.Rev.Nucl.Part.Sci.* **57,** 205–243 (2007).

[11] ALICE, Toia, A. "Bulk Properties of Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV measured by ALICE". *J. Phys. G* **38** (eds Schutz, Y. & Wiedemann, U. A.) 124007. arXiv: 1107.1973 [nucl-ex] (2011).

[12] Gyulassy, M. & McLeran, L. *Nucl. Phys. A* **30** (2005).

[13] Al.(ATLAS Collaboration), G. A. "Measurements of the Nuclear Modification Factor for Jets in PbPb $\sqrt{s_{NN}}$ =2.76TeV with the ATLAS Detector". *Phys. Rev. Lett.* **114** (2015).

[14] d'Enterria, D. "Jet quenching". *Landolt-Bornstein* **23,** 471. arXiv: 0902.2011 [nucl-ex] (2010).

[15] CMS, Chatrchyan, S. *et al.* "Centrality Dependence of Dihadron Correlations and Azimuthal anisotropy Harmonics in PbPb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV". *Eur. Phys. J. C* **72,** 2012. arXiv: 1201.3158 [nucl-ex] (2012).

[16] ALICE, Aamodt, K. *et al.* "Harmonic decomposition of two-particle angular correlations in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV". *Phys. Lett. B* **708,** 249–264. arXiv: 1109.2501 [nucl-ex] (2012).

[17] Wen, F., Bryon, J., Wen, L. & Gang, W. "Event-shape-engineering study of charge separation in heavy-ion collisions". *Chin. Phys.* **C42,** 014001. arXiv: 1608.03205 [nucl-th] (2018).

[18] Poskanzer, A. M. & S.A.Voloshin. "Methods for analyzing anisotropic flow in relativistic nuclear collisions". *Phys. Rev. C* **58,** 1671–1678 (1998).

[19] Borghini, N., Dinh, P. M. & Ollitrault, J.-Y. "Flow analysis from multiparticle azimuthal correlations". *Phys. Rev. C* **64** (2001).

[20] Bilandzic, A., Snellings, R. & Voloshin, S. "Flow analysis with cumulants: Direct calculations". *Phys. Rev. C* **83** (2011).

[21] Heinz, U. & Snellings, R. "Collective flow and viscosity in relativistic heavy-ion collisions". *Ann. Rev. Nucl. Part. Sci.* **63,** 123–151. arXiv: 1301.2826 [nucl-th] (2013).

[22] Gale, C., Jeon, S. & Schenke, B. "Hydrodynamic Modeling of Heavy-Ion Collisions". *Int. J. Mod. Phys. A* **28,** 1340011. arXiv: 1301.5893 [nucl-th] (2013).

[23] Huovinen, P. "Hydrodynamics at RHIC and LHC: What have we learned?" *Int. J. Mod. Phys. E* **22,** 1330029. arXiv: 1311.1849 [nucl-th] (2013).

[24] Baier, R., Dokshitzer, Y., Mueller, A. & Schiff, D. "Angular dependence of the radiative gluon spectrum and the energy loss of hard jets in QCD media". *Phys. Rev.* **C60** (1999).

[25] Salgado, C. . & Wiedemann, U. . "Medium modification of jet shapes and jet multiplicities". *Phys. Rev. Lett.* **93** (2004).

[26] Foka, P. & Janik, M. A. "An overview of experimental results from ultra-relativistic heavy-ion collisions at the CERN LHC:Hard probes." *Rev. Phys.* **1,** 172–194 (2017).

[27] Rafelski, J. & Müller, B. "Strangeness Production in the Quark - Gluon Plasma". *Phys. Rev. Lett.* **48.** [Erratum: Phys.Rev.Lett. 56, 2334 (1986)], 1066 (1982).

[28] Koch, P., Müller, B. & Rafelski, J. "Strangeness in Relativistic Heavy Ion Collisions". *Phys. Rept.* **142,** 167–262 (1986).

[29] Matsui, T. & Satz, H. "J/ψ Suppression by Quark-Gluon Plasma Formation". *Phys. Lett. B* **178,** 416–422 (1986).

[30] NA50Collaboration, B. A. e. a. "A new measurement of J/ψ suppression in Pb-Pb collisions at 158-GeV per nucleon". *Eur. Phys. J.* **C39,** 335–345. arXiv: hep-ex/0412036 [[nucl-ex]] (2005).

[31] PHENIXCollaboration, A. A. e. a. "J/ψ suppression at forward rapidity in Au+Au collisions at $\sqrt{s_{NN}}$=200 GeV". *Phys. Rev.C84(2011) 054912,* **84.** arXiv: 1103.6269 [[nucl-ex]] (2011).

[32] ALICE collaboration, B. A. a. "J/ψ Suppression at Forward Rapidity in Pb-Pb Collisions at $\sqrt{s_{NN}}$=2.76 TeV". *Phys. Rev. Lett.* **109.** arXiv: 1202.1383 [[hep-exp]] (2012).

[33] CMS, Sirunyan, A. M. *et al.* "Measurement of nuclear modification factors of Υ(1S), Υ(2S), and Υ(3S) mesons in PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV". *Phys. Lett. B* **790,** 270–293. arXiv: 1805.09215 [hep-ex] (2019).

[34] ATLAS, Aad, G. *et al.* "Production of Υ(nS) mesons in Pb+Pb and pp collisions at 5.02 TeV". *Phys. Rev. C* **107,** 054912. arXiv: 2205.03042 [nucl-ex] (2023).

[35] Kolband, P. F. & Heinz, U. "Hydrodynamic description of ultrarelativistic heavy ion collisions". (eds Hwa, R. C. & Wang, X.-N.) 634–714. arXiv: nucl-th/0305084 (May 2003).

[36] Hatsuda, T., Yagi, K. & Miake, Y. in, 49 (Cambridge University Press, 2008).

[37] Cooper, F. & Frye, G. "Comment on the Single Particle Distribution in the Hydrodynamic and Statistical Thermodynamic Models of Multiparticle Production". *Phys. Rev. D* **10,** 186 (1974).

[38] Et al., J. B. "Herwig 7.0/Herwig++ 3.0 release note". *Eur. Phys. J.* **C76** (2016).

[39] Gleisberg, T. *et al.* "Event generation with SHERPA 1.1". *JHEP* **02** (2009).

[40] Bierlich, C., Gustafson, G. & Lönnblad, L. "Collectivity without plasma in hadronic collisions". *Phys. Lett. B* **779,** 58–63. arXiv: 1710.09725 [hep-ph] (2018).

[41] Bierlich, C., Chakraborty, S., Gustafson, G. & Lönnblad, L. "Setting the string shoving picture in a new frame". *JHEP* **03,** 270. arXiv: 2010.07595 [hep-ph] (2021).

[42] Et al., A. B. "General-purpose event generators for LHC physics". *Phys. Rept.,* 145–233 (2011).

[43] Christiansen, J. R. & Skands, P. Z. "String Formation Beyond Leading Colour". *JHEP* **08,** 003. arXiv: 1505.01681 [hep-ph] (2015).

[44] Sjostrand, T. & Utheim, M. *A Framework for Hadronic Rescattering in pp Collisions* 2020. arXiv: hep-ph/2005.05658v1.

[45] Bierlich, C., Sjöstrand, T. & Utheim, M. "Hadronic rescattering in pA and AA collisions". *Eur. Phys. J. A* **57,** 227. arXiv: 2103.09665 [hep-ph] (2021).

[46] Wang, X.-N. & Gyulassy, M. "HIJING: A Monte Carlo model for multiple jet production in pp, pA and AA collisions". *Phys. Rev. D* **44** (1991).

[47] Wang, X.-N. "pQCD based approach to parton production and equilibration in high-energy nuclear collisions". *Phys. Rept.* **280,** 287–371. arXiv: hep-ph/9605214 (1997).

[48] Deng, W.-T., Wang, X.-N. & Xu, R. "Hadron production in p+p, p+Pb, and Pb+Pb collisions with the hijing 2.0 model at energies available at the CERN Large Hadron Collider". *Phys. Rev. C* **83** (2011).

[49] Lin, Z.-W., Ko, C. M., Li, B.-A., Zhang, B. & Pal, S. "A Multi-phase transport model for relativistic heavy ion collisions". *Phys. Rev.* **C72** (2005).

[50] Lin, Z.-W. & Zheng, L. "Further developments of a multi-phase transport model for relativistic nuclear collisions". *Nucl. Sci. Tech.* **32,** 113. arXiv: 2110.02989 [nucl-th] (2021).

[51] Pierog, T., Karpenko, I., Katzy, J. M., Yatsenko, E. & Werner, K. "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider". *Phys. Rev.* **C92** (2015).

[52] Werner, K. "On a deep connection between factorization and saturation: new insight into modeling high-energy proton-proton and nucleus-nucleus scattering in the EPOS4 framework". arXiv: 2301.12517 [hep-ph] (Jan. 2023).

[53] Kanakubo, Y. *Unified description of high-energy nuclear collisions based on dynamical core–corona picture* PhD thesis (Sophia U., 2022). arXiv: 2208.07029 [nucl-th].

[54] Et al., B. A. "Phobos results on charged particle multiplicity and pseudorapidity distributions in Au+Au, Cu+Cu, d+Au, and p+p collisions at ultra-relativistic energies". *Phys. Rev.* **C83,** 024913. arXiv: 1011.1940 (2011).

[55] Zhang, B. "ZPC 1.0.1: A Parton cascade for ultrarelativistic heavy ion collisions". *Comput. Phys. Commun.* **109,** 193–206. arXiv: nucl-th/9709009 (1998).

[56] Lin, Z.-w. & Ko, C. M. "Partonic effects on the elliptic flow at RHIC". *Phys. Rev. C* **65,** 034904. arXiv: nucl-th/0108039 (2002).

[57] Li, B. .-. & Ko, C. M. "Formation of superdense hadronic matter in high energy heavy-ion collisions". *Physical Review C* **52,** 2037–2063 (1995).

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

[59] Drescher, H., Hladik, M., Ostapchenko, S., Pierog, T. & Werner, K. "Parton based Gribov-Regge theory". *Phys. Rept.* **350,** 93–289. arXiv: hep-ph/0007198 (2001).

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

[61] Andersson, B., Gustafson, G. & Nilsson-Almqvist, B. "A Model for Low p(t) Hadronic Reactions, with Generalizations to Hadron - Nucleus and Nucleus-Nucleus Collisions". *Nucl. Phys.* **B281,** 289–309 (1987).

[62] Pi, H. "An Event generator for interactions between hadrons and nuclei: FRITIOF version 7.0". *Comput. Phys. Commun.* **71,** 173–192 (1992).

[63] Gribov, V. "Glauber corrections and the interaction between high-energy hadrons and nuclei". *Sov. Phys. JETP* **29,** 483–487 (1969).

[64] Good, M. L. & Walker, W. D. "Diffraction dissociation of beam particles". *Phys. Rev.* **120,** 1857–1860 (1960).

[65] Levin, E. "An Introduction to pomerons". 261–336. arXiv: `hep-ph/9808486 [hep-ph]` (1998).

5 Overview of publications

In the field of theoretical particle physics, all authors are listed alphabetically, rather than in order of the level of contribution. Below follows a short description of each of the publications included in this thesis and my contribution to each one.

Paper I

The Angantyr Model for Heavy-ion Collisions in Pythia8

Christian Bierlich, Gösta Gustafson, Leif Lönnblad, and Harsh Shah
J. High Energ. Phys., 134 (2018)
e-print: `arXiv:1806.10820 [hep-ph]`

The seed of developing the Angantyr model for heavy-ion collisions in Pythia was sown in the earlier work done by my co-authors. They showed that the secondary collisions are crucial, and they can be treated as modified single-diffractive, and require further improvements to simulate heavy-ion events using Pythia machinery.

I tested available single-diffractive interaction models of Pythia and modifications in the Pomeron PDFs to obtain an acceptable set of tuned parameters, to be used to simulate secondary non-diffractive interactions. Later, this is used in the Angantyr model to generate secondary non-diffractive collisions in pA and AA event simulations. I am responsible for obtaining the tuned parameters associated with secondary non-diffractive events in the Angantyr model, generating histograms included in section 5, and writing the first draft of section 5, which was later heavily edited by Leif.

Paper II

A Spatially Constrained QCD Colour Reconnection in pp, pA, and AA Collisions in the PYTHIA8/Angantyr Model

Leif Lönnblad and Harsh Shah
Eur. Phys. J. C 83 (2023) 7, 575
e-print: `arXiv:2303.11747 [hep-ph]`

The idea of extending the Angantyr model with parton-level interactions between the partons from different sub-collisions using colour reconnection is given by Leif. During the initial discussions on how to extend a colour reconnection model for our purpose, Leif came up with the idea of using QCD CR and a combination of Pythia parameters which allow us to enable colour reconnection only after the stacking of all sub-collisions in Angantyr.

I then took over the project. Although there were continuous discussions with Christian, Gösta, Leif and Torbjörn Sjöstrand throughout the development of the project, I figured out how and where to implement spatial constraint in the colour reconnection module, the limitations of the junction fragmentation and extended the junctions handling in the Mini-string fragmentation module of Pythia, which exponentially improved the computational limitations of the upgraded model. I decided on the parameters to be re-tuned and obtained the new tuned values for the spatially constrained QCD CR in Pythia and in Angantyr. I and Torbjörn Lundberg (another PhD student) fixed a crucial bug in Pythia8.3, fixing of which allowed the new changes from this work to be applicable both in Pythia8.2 and Pythia8.3. I am responsible for the first draft of the paper and all the results. The final version of the paper was achieved with the help of Gösta and Leif.

Paper III

Baryon Correlations in Pythia

Leif Lönnblad and Harsh Shah
To be submitted to Eur. Phys. J. C
e-print: `arXiv:2309.01557 [hep-ph]`

I came across the discrepancy between Pythia results and ALICE experimental data for the angular correlations in the baryon sector. Later, I learned that it was a known problem and we decided to investigate it.

I had multiple discussions with Christian, Gösta, Leif and Torbjörn Sjöstrand throughout the project. I tested the effects of the various parameters in Pythia and narrowed it down to colour reconnections and hadronic rescattering can influence the angular correlations. The idea of the effects of baryons produced around ”gluon-kinks” in generating angular correlations came from Gösta and Leif. I later used Pythia’s USERHOOK class with help from Leif and Torbjörn S. to impose a veto on baryons production near ”gluon-kinks”. I am responsible for writing the first draft of the paper and the results. Leif tested a toy model and generated histograms, which showed how a gluon influences the angular correlations between baryon pairs and edited the draft.

Paper IV

The Dynamic Hadronization of Charm Quarks in Heavy-ion Collisions

Christian Bierlich, Gösta Gustafson, Leif Lönnblad, and Harsh Shah
To be submitted to Eur. Phys. J. C
e-print: `arXiv:2309.12452 [hep-ph]`

The idea of focusing on charm baryons came from Christian when we were discussing testing the model developed in Paper II for various flavours production and flow generation.

I tested various possibilities and we had multiple discussions with Torbjörn Sjöstrand and together we figured out the need for modifications in the junction formation and fragmentation if a heavy quark is involved. Leif implemented the changes in the junction fragmentation module, and I made the changes in the junction formation and finalised the other parameters to be used in this paper. I obtained the results and wrote the first draft of the paper. Later, the manuscript was revised in collaboration.

Paper I

I

The Angantyr Model for Heavy-ion Collisions in Pythia8

Christian Bierlich, Gösta Gustafson, Leif Lönnblad, and Harsh Shah

J. High Energ. Phys., 134 (2018)
DOI: 10.1007/JHEP10(2018)134
e-print: arXiv:1806.10820 [hep-ph]
LU-TP 18-19, MCnet-18-12

Abstract: We present a new model for building up complete exclusive hadronic final states in high energy nucleus collisions. It is a direct extrapolation of high energy pp collisions (as described by PYTHIA), and thus bridges a large part of the existing gap between heavy ion and high energy physics phenomenology. The model is inspired by the old Fritiof model and the notion of wounded nucleons. Two essential features are the treatment of multi-parton interactions and diffractive excitation in each NN sub-collision. Diffractive excitation is related to fluctuations in the nucleon partonic sub-structure, and fluctuations in both projectile and target are here included for the first time. The model is able to give a good description of general final-state properties such as multiplicity and transverse momentum distributions, both in pA and AA collisions. The model can therefore serve as a baseline for understanding the non-collective background to observables sensitive to collective behaviour. As PYTHIA does not include a mechanism to reproduce the collective effects seen in pp collisions, such effects are also not reproduced by the present version of Angantyr. Effects of high string density, shown to be able to reproduce *e.g.* higher strangeness ratios and the ridge in pp, will be added in future studies.

1 Introduction

At hadron collider experiments at RHIC and LHC, protons as well as large nuclei, are collided, and the results are interpreted to obtain better knowledge about the dynamics of the fundamental interactions at high energies. The strong nuclear force plays a central role, but the studies of proton–proton (pp) collisions and heavy ion collisions respectively, are often carried out in quite different ways.

In the case of pp collisions, so–called "general purpose Monte Carlo event generators", such as SHERPA [1], Herwig 7 [2] and PYTHIA8 [3], have been established as cornerstones in aiding our understanding. These event generators have over the last three decades succeeded in simultaneously simulating the dynamics of strong and electroweak processes from very high momentum transfer scales where perturbation theory is applicable, down to scales around Λ_{QCD}, where one must rely on models inspired by analogies to electrodynamics or results from lattice QCD. This has resulted in a remarkably precise description of the majority of observations in proton–proton collisions, which both further experimental and theoretical developments often rely heavily upon.

In high energy heavy ion collisions, the landscape is quite different. Here efforts are more often directed towards signals for the formation of the Quark–Gluon Plasma (QGP), and studies of its properties. The existence of such a phase is demonstrated in lattice calculations and it is presumed to have existed in the hot, early Universe. In this area event generators also exist, but are usually more "special purpose" than "general purpose", each attempting to describe a specific array of observations ascribed to the formation of a QGP. Event generators generating full exclusive events also exist, and the ones most frequently used in analyses investigating particle production mechanisms are, arguably, EPOS-LHC [4], AMPT [5] and HIJING [6]. At least for the bulk event properties, these three generators have for many years defined the "golden standard" for Monte Carlo comparisons to experimental data. In section 8 we outline some of the main similarities and differences between these models and our own.

Several features, which in heavy ion physics are interpreted as a QGP effect, are also observed in pp collisions at the LHC, which may indicate that the dynamics at play in these two types of collision systems are in fact very similar. Two typical examples are enhanced strangeness [7] and the formation of a "ridge" [8]. This immediately raises a challenge for the general purpose pp event generators and their underlying models. If a QGP is indeed formed even in pp collisions, then the effects of such a formation should be included. On the other hand, if the flow-like effects in pp collisions have a different, non-thermal, origin, then it might be possible to capture the general features of nuclear collisions by adding a nuclear structure "on top" of existing pp models.

In the present paper we will primarily address the second of these possibilities, presenting a

model, henceforth called "Angantyr", which is an extrapolation of pp dynamics to collisions with nuclei with a minimum of adjustable parameters. In this way it forms a bridge between heavy ion and high energy hadron phenomenology. Angantyr is a generalisation to AA collisions of the model for pA scattering in ref. [9], which was able to reproduce general features in pA collisions, like multiplicity as a function of (measured) centrality, rapidity distributions, and to a certain degree also $p_\perp$ distributions. Like PYTHIA8 and the model in ref. [9], Angantyr does not include an assumption of a hot thermalised medium. The model can therefore serve as a baseline for understanding the non-collective background to observables sensitive to collective behaviour.

Before discussing the generalisation to heavy ion collisions, we want to discuss some features of *high energy pp scattering*, which are important for this generalisation.

First, as will be discussed in more detail below, diffractive excitation is important. At high energies the real part of the pp amplitude is small, and usually neglected in applications to collisions with nuclei. Diffraction (elastic scattering and diffractive excitation) is then the shadow of absorption into inelastic (non-diffractive) channels. Absorption is here specified by colour exchange between projectile and target, while diffraction corresponds to colour neutral (Pomeron) exchange. In the Good–Walker formalism diffractive excitation is then part of the diffractive beam, when the projectile mass eigenstate (the proton) is a (coherent) linear combination of scattering eigenstates with different absorption probability. These eigenstates have in refs. [10, 11] been interpreted as different parton cascades.

Secondly multiple partonic sub-collisions are very important at high energies. Here we use the scheme from ref. [12], as implemented in PYTHIA8, to describe inelastic non-diffractive events. Hard scattering is also seen in diffractive events, and here we use the Ingelman–Schlein formalism [13], which is also included in the PYTHIA8 package.

A generalisation of the formalism for pp collisions to an event generator for pA *and* AA *collisions* will have *four separate components*:

(i) It is necessary to determine nucleon positions within the colliding nuclei. Here a number of MCs are already available to generate nucleon distributions, see *e.g.* refs. [14–17].

(ii) One has to calculate the number of interacting nucleons and binary NN collisions. This is generally performed using the Glauber formalism [18, 19]. This formalism is based on the eikonal approximation in impact parameter space, where the projectile nucleon(s) are assumed to travel along straight lines and undergo multiple sub-collisions with nucleons in the target. The importance of including diffractive excitation was early pointed out by Gribov [20], but has often been neglected also in recent applications (see *e.g.* the review by Miller *et al.* [19])[1]. As mentioned above, diffractive excitation is a consequence

[1]As an example, in many analyses the NN interaction has been approximated by a "black disk model", where diffractive excitation of individual nucleons is completely neglected.

of fluctuations in the nucleon substructure. An important point is then that a nucleon in the projectile is fixed in the same state during its passage through the target nucleus. (And similarly the state of a target nucleon is fixed through the projectile nucleus.)

Fluctuations in the *projectile* proton in pA collisions was studied by Heiselberg *et al.* [21], for estimates of the number of individual NN sub-collisions. This formalism was further developed in several papers (see refs. [22–25] and further references in there). It is often referred to as the "Glauber–Gribov" colour fluctuation model (GGCF or just GG), and is used in several experimental analyses, *e.g.* in refs. [26, 27].

As discussed in ref. [9], taking averages over target nucleon states is enough for calculations of cross sections and the number of wounded nucleons in pA collisions, *provided* diffractively excited nucleons are also counted as wounded nucleons. For a generalisation to AA collisions it is, however, necessary to take into account individual fluctuations in both projectile and target nucleons. As far as we know, Angantyr is the first model where this condition is satisfied.

(iii) One must estimate the contribution to the final state from each interacting nucleon. The Angantyr model is here inspired by the old Fritiof model for pA and AA collisions [28, 29] and the notion of "wounded" nucleons [2]. Białas, Bleszyński, and Czyż [30] showed that the production of soft particles is determined by the number of wounded (or participant) nucleons, rather than the number of individual NN sub-collisions. (The latter was later seen to be correlated to hard processes, like production of high $p_\perp$ particles or vector bosons.) In the early Fritiof model [28] it was assumed that an interacting nucleon suffers a longitudinal momentum exchange with a distribution $\sim dQ/Q$, leading to an excited mass $\sim dM^2/M^2$. When hadronising like a colour string this gives *on average* a triangular distribution in rapidity. This behaviour was also later obtained by Białas and Czyż in an analysis of dAu collisions at RHIC [31].

The Fritiof model did not explicitly include diffractive excitation. We note, however, that if the mass distribution for diffractive excitation can be approximated by $dP \propto dM^2/M^2$, then the contribution from a diffractively excited nucleon is very similar to the contribution from an average wounded nucleon in the Fritiof model or from the analysis in ref. [31]. The wounded nucleons in Fritiof can therefore effectively represent both non-diffractively and diffractively wounded nucleons.

(iv) At high energies, the *hard* partonic sub–collisions (scaling with NN sub-collisions rather than wounded nucleons) play a very essential role. It is therefore necessary to account for those specifically in events with multiple NN collisions, *e.g.* when one projectile nucleon interacts with several target nucleons (or *vice versa*). In ref. [9] we introduced the concept of *primary* and *secondary absorptive* interactions, when a projectile nucleon is in-

[2] This is also the case for the HIJING model.

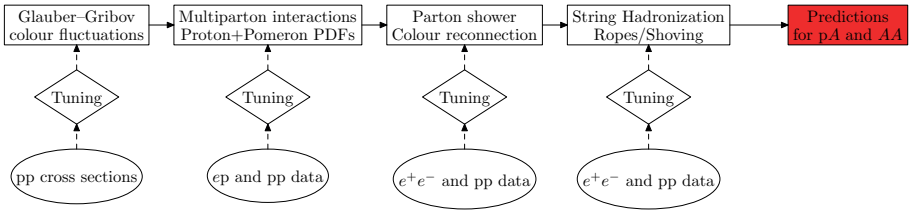


Figure I.1: Flowchart showing the programmatic structure of Angantyr. In order to make predictions for heavy ion collisions, several parts of a normal PYTHIA8 simulation needs to be modified, and tuned accordingly. In the flowchart we illustrate how each separate part is tuned to either e^+e^-, ep or pp data, while no tuning is done to heavy ion data.

teracting absorptively with more than one target nucleon. The corresponding NN parton-level event could be generated using the full multi-parton interaction (MPI) machinery in PYTHIA8, for both absorptive and diffractive interactions. To generate fully exclusive final states in AA collisions, we then have to calculate all sub-collisions between a nucleon μ in the projectile and nucleon ν in the target, study the number of multiple sub-interactions for all nucleons μ and ν, and here separate diffractive from non-diffractive (absorptive) interactions. This process is fully described in section 3.

We now return to the question of QGP formation. In the current version of Angantyr the generated partonic states are hadronised using the string fragmentation model in PYTHIA8, without including any final-state collective effects. In this way the model can be used as a starting point for implementing and analysing new models for collectivity. As an example we showed [32] that an enhanced strangeness production can be expected in (high multiplicity) pp collisions, due to overlapping colour strings forming "ropes", in agreement with experimental observations [7]. Furthermore we demonstrated in refs. [33, 34] that the enhanced density also ought to give an outward pressure, which may explain the observed flow-like effects in pp scattering.

In the present version of the model we limit ourselves to general features like distributions of particle density in rapidity and $p_\perp$, postponing a discussion of flow-like effect to a coming publication. We would like to emphasise, however, that the model can still be used as an important tool for understanding non-flow effects on experimental observables designed to measure flow and other collective behaviours.

In Figure I.1 we show how the structure described above is put together and tuned in the concrete simulation. Since all parts of the simulation; GG colour fluctuations to generate the number of sub-collisions, the PYTHIA8 MPI model, the parton shower and the hadronisation model rely on a number of parameters, these parameters need to be tuned, and a large part of this paper describes how this procedure is carried out. We want to emphasise

from the beginning that all parts are tuned to data from collisions of smaller systems, e^+e^-, ep and pp, and no tuning is done to heavy ion data. The results can thus be regarded as real predictions depending only on the chosen extrapolation procedure, and not a specific choice of parameters.

The layout of the paper follows the workflow of the generation procedure as shown in Figure I.1, and implemented in PYTHIA8. In section 2 we discuss how to calculate the number of wounded nucleons and the number of individual NN sub-collisions. Here we include fluctuations both in the distribution of nucleons in the nuclei and in the individual nucleon states, both for nucleons in the projectile nucleus and in the target nucleus. We note that a projectile (target) nucleon is fixed in the same diffractive eigenstate through the passage through the target (projectile) nucleus. If it is then not absorbed, it may end as diffractively excited, when projected to the system of mass eigenstates. Then in section 3 we discuss how to generate the parton-level sub-events for the different kinds of sub-collisions, and in section 4 we describe the procedure for stacking these sub-events together into complete exclusive hadronic final-states in AA. In section 5 we then make a digression to discuss the details of the generation of secondary absorptive sub-collisions, before we present some sample results in section 6. In section 7 we discuss model uncertainties, especially related to our treatment of secondary absorptive sub-collisions. Finally we discuss differences and similarities between our approach and other heavy ion event generators in section 8, before presenting some conclusions and an outlook in section 9.

2 Nucleon-nucleon sub-collisions in pA and AA

In high energy pp collisions the real part of the amplitude A is small. If this can be neglected, we can define the real quantity

$$T \equiv \mathbb{I}\mathrm{m}\{A\} = 1 - S. \tag{I.1}$$

If diffractive excitation also can be neglected, the elastic cross section is just the shadow of the absorption, which in impact parameter space is determined by the probability $1 - S^2$. The inelastic cross section is then simply the difference between the two. The elastic, total, and inelastic pp cross sections are then given by $d\sigma^{NN}_{\mathrm{el}}/d^2b = T(\mathbf{b})^2$, $d\sigma^{NN}_{\mathrm{tot}}/d^2b = 2T(\mathbf{b})$, and $d\sigma^{NN}_{\mathrm{abs}}/d^2b = 2T(\mathbf{b}) - T^2(\mathbf{b})$ respectively.

The formulations of high energy nucleus collisions in terms of individual nucleon–nucleon interactions was carried out by Glauber in a pioneering paper in ref. [18]. In this paper several kinds of fluctuations were neglected. As pointed out by Gribov, and discussed in the introduction, diffractive excitation of individual nucleons is essential, both for cross sections and for final state properties. The Glauber theory is formulated in impact parameter space, where cross sections can be directly interpreted as probabilities. It is then most convenient

to include diffractive excitation using the Good–Walker formalism [35], as the result of fluctuations in the nucleon wave functions. In this section we shortly discuss the Glauber and Good–Walker formalisms for estimating scattering cross sections and distributions of wounded nucleons and NN sub-collisions. The discussion of effects on the properties of exclusive final states will be presented in section 3.

2.1 Glauber formalism

The Glauber formalism is based on the eikonal approximation in transverse coordinate space. Here the projectile nucleon(s) travel along straight lines, and undergo multiple sub-collisions with small transverse momenta. Multiple interactions correspond to a convolution of the individual S-matrices in transverse momentum space, which in transverse coordinate space simplifies to a product.

We let $\mathbf{b}_\mu$ and $\mathbf{b}_\nu$ denote the set of positions in impact parameter space for the nucleons in the projectile and target nucleus respectively, and $\mathbf{b}$ the separation between the centres of the colliding nuclei. The S-matrix for scattering between nucleus A and nucleus B is the given by

$$S^{AB}(b) = \prod_{\mu=1}^{A} \prod_{\nu=1}^{B} S^{(N_\mu, N_\nu)}(b_{\mu\nu}). \tag{I.2}$$

Here $\mathbf{b}_{\mu\nu} = \mathbf{b}_\mu + \mathbf{b} - \mathbf{b}_\nu$ is the relative separation between the two colliding nucleons N_μ and N_ν. For pA collisions the product over μ contains only the projectile proton with $b_\mu = 0$.

As mentioned above, fluctuations were neglected in Glauber's original paper. In the *optical limit*, with a smooth distribution of nucleons in the nuclei, and where the size of the nuclei are large compared to the range of the NN interaction, the resulting nucleus-nucleus cross sections can be calculated analytically.

2.2 Fluctuations

2.2.1 Nucleus geometry

The simplest way to include fluctuations in the nucleon positions, $\mathbf{b}_\mu$, within a nucleus, is to randomly distribute the A nucleons in three-dimensional space according to a Woods–Saxon distribution. More advanced models include correlations in form of a hard repulsive core (*e.g.* [14, 15]), or a more sophisticated description of the two- (or three-) particle correlations between the nucleons within the nucleus [16, 17]. Fluctuations in the geometry is taken into account, when new nucleus states are generated for each new event.

2.2.2 Fluctuations in the individual NN interactions, and the Good–Walker formalism

We here shortly describe the Good–Walker formalism for diffractive excitation, assuming for simplicity first that a fluctuating projectile collides with a non-fluctuating target. For a projectile particle with an internal substructure, it is possible that the mass eigenstates differ from the elastic scattering eigenstates. We denote the mass eigenstates Ψ_i, with the projectile in the ground state (*e.g.* a proton) denoted Ψ_0, while Φ_l are the eigenstates to the scattering amplitude T, with $T\Phi_l = t_l\Phi_l$. The mass eigenstates are linear combinations of the scattering eigenstates, $\Psi_i = \sum_l a_{il}\Phi_l$. The scattering can now be regarded as a measurement, where the projectile "has to choose" one of the eigenvalues t_l, with probability $|a_{0l}|^2$.

The elastic amplitude for the ground state projectile is then given by $\langle\Psi_0|T|\Psi_0\rangle = \sum_l |a_{0l}|^2 t_l \equiv \langle T\rangle$, where $\langle T\rangle$ is the expectation value for the amplitude T for the projectile. The elastic cross section is then given by

$$d\sigma_{el}/d^2b = \langle T(b)\rangle^2. \tag{I.3}$$

We here work in impact parameter space, and the amplitude depends on b. The total diffractive scattering $\sigma_{\rm diff}$ (including the elastic) is the sum of transitions to all states Φ_l:

$$d\sigma_{\rm diff}/d^2b = \sum_l \langle\Psi_0|T|\Phi_l\rangle\langle\Phi_l|T|\Psi_0\rangle = \langle\Psi_0|T^2|\Psi_0\rangle, \tag{I.4}$$

where we have used the fact that Φ_l form a complete set of states. Subtracting the elastic cross section we then get the cross section for diffractive excitation, which thus is given by the fluctuations in the scattering amplitude:

$$d\sigma_{\rm D}/d^2b = \langle T^2\rangle - \langle T\rangle^2. \tag{I.5}$$

In a nucleon-nucleon collision both the projectile and the target are fluctuating, leading to single diffractive excitation of the projectile or the target, as well as to double diffraction. The different cross sections are then given by

$$
\begin{aligned}
d\sigma^{NN}_{\rm tot}/d^2b &= \langle 2T(\mathbf{b})\rangle_{p,t} \\
d\sigma^{NN}_{\rm abs}/d^2b &= \left\langle 2T(\mathbf{b}) - T^2(\mathbf{b})\right\rangle_{p,t} \\
d\sigma^{NN}_{\rm el}/d^2b &= \langle T(\mathbf{b})\rangle^2_{p,t} \\
d\sigma^{NN}_{\rm Dt}/d^2b &= \left\langle \langle T(\mathbf{b})\rangle^2_p \right\rangle_t - \langle T(\mathbf{b})\rangle^2_{p,t} \\
d\sigma^{NN}_{\rm Dp}/d^2b &= \left\langle \langle T(\mathbf{b})\rangle^2_t \right\rangle_p - \langle T(\mathbf{b})\rangle^2_{p,t} \\
d\sigma^{NN}_{\rm DD}/d^2b &= \left\langle T^2(\mathbf{b})\right\rangle_{p,t} - \left\langle \langle T(\mathbf{b})\rangle^2_p \right\rangle_t - \left\langle \langle T(\mathbf{b})\rangle^2_t \right\rangle_p + \langle T(\mathbf{b})\rangle^2_{p,t}\,. \qquad \text{(I.6)}
\end{aligned}
$$

Here $\langle\cdots\rangle_p$ and $\langle\cdots\rangle_t$ are averages over projectile and target states respectively, and subscripts Dt, Dp and DD stand for single diffractive excitation of the target, the projectile, and double diffraction respectively. We note here that while the total cross section depends only on the average of $T(b)$, all other cross sections include also average of T^2 over projectile and/or target states. However, if wounded target nucleons include also diffractively excited nucleons, we see that the corresponding cross section for a wounded target nucleon, $\sigma^{NN}_{\rm Wt} \equiv \sigma^{NN}_{\rm abs} + \sigma^{NN}_{\rm Dt} + \sigma^{NN}_{\rm DD}$, can be written

$$
d\sigma^{NN}_{\rm Wt}/d^2b = \left\langle 2\,\langle T(\mathbf{b})\rangle_t - \langle T(\mathbf{b})\rangle^2_t \right\rangle_p = 1 - \left\langle \langle S(\mathbf{b})\rangle^2_t \right\rangle_p . \qquad \text{(I.7)}
$$

2.2.3 Fluctuations in collisions with nuclei

The expression for the amplitude $T(\mathbf{b}) = (1 - S(\mathbf{b}))$ in eq. (I.6) can be directly inserted into the amplitude for collisions with nuclei in eq. (I.2) (as before we neglect the real part of the amplitudes). The scattering probability can be regarded as a measurement, after which a projectile nucleon is in one of the eigenstates to the amplitude T, and thus also to the probability for colour connection (the absorption probability) $2T - T^2$. Thus all nucleons are frozen in the same state during the scattering process. (We here neglect the modification when one or a few partons have changed colour in the first encounter.) As a consequence the average of the AA amplitude in eq. (I.2) will include also higher powers of T. However, for pA collisions the multiple sub-collisions imply that the total and wounded nucleon cross sections contain higher moments with respect to *projectile* fluctuations, but still only the average over the uncorrelated *target nucleon* states. We also note that these moments should be taken for fixed impact parameters. Thus, to calculate the ratios of *e.g.* the *integrated* elastic and total cross sections, it is also necessary to know the b-distribution of the amplitude.

To visualise the effects of fluctuations and diffractive excitation we can study a simple example with a proton colliding with two target nucleons, with and without fluctuations. We

assume in both cases that the inelastic NN cross section (including diffractive excitation) is $d\sigma_{\rm inel}^{NN}/d^2b = 3/4$.

Case 1: No fluctuations. The NN amplitude and S matrix are $T^{\rm NN} = 1/2$ and $S^{\rm NN} = 1 - T^{\rm NN} = 1/2$. The inelastic cross section when hitting *two* target nucleons is then from eqs. (I.2), (I.7) given by $d\sigma_{\rm inel}/d^2b = 15/16$, ($\sigma_{\rm D} = 0$).

Case 2: With fluctuations. We neglect the fluctuations in the target, and assume that the projectile state is given by $\Psi_0 = (1/\sqrt{2})(\Phi_1 + \Phi_2)$. The states Φ_1 and Φ_2 are here diffractive eigenstates with eigenvalues $t_1 = 0$ and $t_2 = 1$. From eq. (I.6) we get for collision with *one* target nucleon $d\sigma_{abs}/d^2b = 1/2$ and $d\sigma_D/d^2b = 1/4$. For *two* target nucleons we get actually the same result. If the projectile is in state Φ_1 it misses both targets, and if in state Φ_2, it is absorbed already in the first one. Thus the inelastic cross section is only 3/4 (1/2 for absorption and 1/4 for diffractive excitation) compared to 15/16 in the non-fluctuating case.[3]

2.3 From cross sections to probabilities

The absorptive cross section in impact parameter space shown in eq. (I.6) is the average of the expression $2T_{i,k}(\mathbf{b}) - T_{i,k}^2(\mathbf{b}) \equiv 1 - S_{i,k}^2(\mathbf{b})$, where $T_{i,k}$ is the scattering amplitude (and $S_{i,k}$ the S-matrix) for a projectile proton in state i colliding with a target in state k. This expression is always ≤ 1, and it can be directly interpreted as the probability for an absorptive interaction between the projectile and the target. (Such an interpretation is not possible in transverse momentum space, where the cross section has the dimension of momentum to the fourth power.)

We note, however, that neither the elastic cross section nor diffractive excitation is the average of an expression depending on only i and j. (The elastic cross section can be written $\sum_{i,j,k,l} T_{i,k}T_{j,l}$.) When the interaction is driven by absorption, elastic scattering and diffractive excitation is the result of interference between waves, which missed the absorbing target. The cross section for this diffractive scattering is also bounded by 1, and together with absorption it gives a total cross section bounded by 2. A consequence of this feature is that to properly generate events including diffractive excitation for AA collisions in an event generator, it is necessary to, for every projectile nucleon, μ, in state i calculate the average of the amplitude $T_{i,k}(b_{\mu\nu})$ over all states of each target nucleon, ν, for all impact parameters $b_{\mu\nu}$ (and similarly all averages over projectile states i for every target state j). This would give a very slow program, and in section 2.5 we show how to obtain a good approximation.

[3]This case is actually essentially the "fluctuating gray disk model" discussed in section 2.4.2 and used in analyses of RHIC data by PHENIX.

In pA collisions the picture is, however, much simplified. From eq. (I.7) we note that although the wounded nucleon cross section $d\sigma_{\rm Wt}^{NN}/d^2b$ contains one piece from absorption and one piece from diffraction, the sum is always bounded by 1. The question whether a target nucleon ν will be a wounded target (with this definition) in a sub-collision with a projectile in state i can only be answered by yes or no. Therefore the answer yes must have the probability given by the cross section in eq. (I.7). This is used *e.g.* in applications of the Glauber–Gribov model described in section 2.4.2.

2.4 NN scattering models used in Glauber calculation Monte Carlos

2.4.1 Non-fluctuating models

The simplest approximation for the NN amplitude is the "**black disk model**", where the target acts as a black absorber. This model has been frequently used in experimental analyses (see *e.g.* the review in ref. [19]). It is then assumed that two colliding nucleons are interacting, if their separation in impact-parameter is smaller than some radius R. The cross sections are here given by $\sigma_{inel} = \sigma_{el} = \sigma_{tot}/2 = \pi R^2$. As there are no fluctuations, the cross section for diffractive excitation is zero. It is then obvious that the model cannot reproduce the experimental results, which satisfy $\sigma_{\rm el}^{NN} \approx \sigma_{\rm D}^{NN} \approx \sigma_{\rm tot}^{NN}/4$. (Here $\sigma_{\rm D}^{NN}$ denotes the sum of single and double diffractive excitation.) In the literature it is common to set $2\pi R^2 = \sigma_{\rm tot}^{(exp)}$, which reproduces the experimental total cross section, but neither the elastic nor the inelastic cross section (when the latter includes diffractive excitation). In later studies it has become more common to choose $\pi R^2 = \sigma_{\rm inel}^{(exp)}$, which reproduces the total inelastic cross section, but gives $\sigma_{\rm tot}^{(model)} = 2\,\sigma_{\rm inel}^{(model)} \approx 1.5\,\sigma_{\rm tot}^{(exp)}$.

For most applications in pA and AA, the elastic cross section is not very important, but we note that it could still be reproduced by introducing a grayness or opacity of the collision, assuming that within a radius R the scattering amplitude is a constant a between 0 and 1. R and a can then always be adjusted to reproduce both the total and the elastic cross sections (and thus also the total inelastic cross section). Diffractive excitation would, however, still be absent.

2.4.2 Models including fluctuations

In a variation of the opacity model above, the projectile is instead fully absorbed with probability a. This obviously includes fluctuations and thus also diffraction. With the value $a = 1/2$ we get the cross section ratios $\sigma_{\rm el}^{NN} = \sigma_{\rm D}^{NN} = \sigma_{\rm tot}^{NN}/4$, in reasonable agreement with experiments. As the model describes the combined fluctuations of the projectile and the target, it is here not possible to separate diffractive excitation of the projectile from that

of the target or from double diffraction.

In the introduction we mentioned the "Glauber–Gribov" model for pA collisions, developed by Strikman and coworkers [21–25]. It is there assumed that the fluctuations in the *projectile* can be described by a distribution in the quantity $\sigma \equiv \int d^2b \, \langle 2T(\mathbf{b})\rangle_t$ of the form

$$\begin{aligned} P_{\rm tot}(\sigma) &= \rho \frac{\sigma}{\sigma+\sigma_0} \exp\left\{-\frac{(\sigma/\sigma_0-1)^2}{\Omega^2}\right\}, \\ \sigma_{\rm tot}^{NN} &= \int \sigma P_{\rm tot}(\sigma) d\sigma. \end{aligned} \tag{I.8}$$

(The second relation follows from eq. (I.6).) This formalism, has been used in analyses of pPb data from LHC, *e.g.* by ref. [26], to estimate the number of wounded or interacting nucleons, which in turn has been used to estimate the centrality for the collision. The quantity σ is then normally rescaled so that the integral in eq. (I.8) gives the inelastic rather than the total cross section. We note that, as the fluctuating quantity σ includes the fluctuations over *projectile* states, but averages over *target nucleon* states, we see from eq. (I.7) that what is counted as wounded nucleons includes diffractively excited nucleons.

As discussed in section 2.3, the cross section in eq. (I.7) also determines the probability distribution for wounded nucleons, but we want to emphasise that the differential cross section $\langle T(\mathbf{b})\rangle_t$ is needed for all values of the impact parameter b. In ref. [23] this is assumed to be Gaussian $\propto \exp(-b^2/2B(\sigma))$, with a slope parameter $B(\sigma)$ proportional to σ, in order to satisfy the unitarity constraint $T(b) \leq 1$.

In ref. [9] we investigated the fluctuations in the nucleon cross sections using Mueller's dipole approach to BFKL evolution [36, 37] as implemented in the DIPSY Monte Carlo program [11, 38, 39]. The model is formulated in impact parameter space, and includes also a set of sub-leading corrections beyond the leading-log BFKL approximation. Non-linear effects are introduced by the "colour swing" mechanism, which suppresses large dipoles, corresponding to $k_\perp$ below a saturation scale. BFKL evolution is a stochastic process, and the result was here that the fluctuations have a longer tail out to large cross sections compared to the distribution in eq. (I.8). Rather than the Gaussian suppression assumed in [21], we found a distribution more similar to a Log-normal for the b-integrated and target-averaged σ:

$$P_{\rm tot}(\ln\sigma) = \frac{1}{\Omega\sqrt{2\pi}} \exp\left(-\frac{\ln^2(\sigma/\sigma_0)}{2\Omega^2}\right). \tag{I.9}$$

To also describe the b-dependence of $\langle T(b)\rangle_t$, we used a semi-transparent disk approximation with the elastic amplitude

$$\langle T(\mathbf{b},\sigma)\rangle_t = T_0 \Theta\left(\sqrt{\frac{\sigma}{2\pi T_0}} - b\right). \tag{I.10}$$

The parameters (Ω and σ_0) in $P_{\rm tot}(\sigma)$ and T_0 in eq. (I.10) could here be fitted to $\sigma_{\rm tot}^{NN}$, $\sigma_{\rm el}^{NN}$ and $\sigma_{\rm Wt}^{NN}$ taken from experimental data, to obtain a Glauber-like calculation for pA. Together with the parton-level stacking also proposed in [9] we then also obtained a fair description of *e.g.* the observable used by ATLAS in [26] for estimating centrality, as well as the corresponding pseudo-rapidity distributions as a function of that centrality.

We note that the stochastic nature of BFKL evolution has also been studied by Iancu, Mueller and Munier in ref. [40]. When the probability for a dipole splitting is small, the mean field approximation in the Balitsky–Kovchegov equation does not properly describe the probability for rare events with large cross section. In ref. [40] they studied the fluctuations in the saturation scale, Q_s, and showed that for asymptotic energies the width of the distribution in $\ln(Q_s)$ is growing proportional to $\sqrt{\bar{\alpha}\ln(s)}$, with a tail to large Q_s-values in qualitative agreement with eq. (I.9).

2.5 Nucleon fluctuations in AA collisions

As mentioned in section 2.2.2, to study pA collisions also higher moments over projectile fluctuations are needed. When we now want to generalise the formalism to AA collisions, both projectile and target nucleons are frozen under the collision (but still uncorrelated). This implies that we must be able to calculate not only $\langle\langle T(b)\rangle_t^n\rangle_p$, but any moment $\langle\langle T(b)^{n_p}\rangle_p^{n_t}\rangle_t$. To cope with this situation we need a formalism which can give the amplitude $T_{ik}(b)$ for any combination of projectile state i and target state k.

We noted that the Log-normal distribution in eq. (I.9) is quite similar to a Gamma-function, and for technical reasons and the fact that the sum of two Gamma distributed random variables is also Gamma distributed, we will use that instead to model fluctuations in the radius, r, of a nucleon:

$$P(r) = \frac{r^{k-1}e^{-r/r_0}}{\Gamma(k)r_0^k}. \tag{I.11}$$

We then also use a slightly different elastic amplitude

$$T(\mathbf{b}, r_p, r_t) = T_0(r_p + r_t)\Theta\left(\sqrt{\frac{(r_p+r_t)^2}{2T_0}} - b\right). \tag{I.12}$$

where the opacity of the semi-transparent disk now depends on r_p and r_t:

$$T_0(r_p + r_t) = \left(1 - \exp\left(-\pi(r_p+r_t)^2/\sigma_t\right)\right)^\alpha. \tag{I.13}$$

This introduces two more parameter, σ_t and α, (besides k and r_0 in eq. (I.11)) and this varying opacity makes it possible to get a reasonable fit to all the cross sections in eq. (I.6), as well as the elastic slope parameter $B = -d\ln\,\sigma_{\rm el}^{NN}/dt|_{t=0}$, for a wide range of energies. The result for $\sqrt{s_{NN}} = 5$ TeV is shown in table I.1.

2.5.1 Determining the interaction of nucleon sub-collisions

We now want to take all pairs of colliding nucleons in an AA collision, and for each of these select which kinds of interactions are possible. At high energies all nucleons are frozen in their (random) states during the passage through the opposite nucleus. The probability for an absorptive interaction between nucleon μ (in a state i with radius $r_{i\mu}$) in the projectile and nucleon ν (in state k with radius $r_{k\nu}$) in the target, is then directly given by eq. (I.6) as $P_{abs} = 2T_{ik} - T_{ik}^2$, with $T_{ik} = T(\mathbf{b}, r_{i\mu}, r_{k\nu})$ given by eq. (I.12). To estimate the probability for a diffractive excitation of a given nucleon is more difficult, as diffractive excitation is part of the shadow scattering caused by absorption, to which all encountered nucleons contribute.

We showed in ref. [9] that for a given state of a projectile nucleon the probability that a given target nucleon is absorptively or diffractively wounded in the interaction is given by the average over the possible states of the target (*c.f.* eq. (I.7)) and that this probability factorises for all nucleons in the target nucleus. However, in AA the symmetry between projectile and target complicates things further, as we need both a specific state and the average over all states for all nucleons.

In Angantyr this is handled by generating two states (one primary, r, and one auxiliary, r') for each nucleon in the nuclei. The primary one is used to calculate the probability of an absorptive NN interaction, while the secondary is used to statistically sample the average state of each nucleon. The algorithm ensures that on average (over the four possible combinations of states in an NN interaction) we get the correct probability of the projectile and target nucleon being absorptively and diffractively wounded.

The technical details of this algorithm is presented in appendix 10, while here we will only show that it works as expected. In Table I.1 we give an example where we have fitted the parameterisation of the fluctuations according to eqs. (I.11) – (I.13) to the default parameterisation of the semi-inclusive cross sections in PYTHIA8. This default parametrisation [41, 42] does not necessarily agree well with cross section measurements from LHC [43], and it is possible for a user to easily supply their own cross sections as input to the fit. The last line denoted "generated" shows the results from generating NN collisions in Angantyr for $\sqrt{s} = 5$ TeV. We see that the absorptive cross section comes out close to the input one, and also the wounded cross sections, σ_{Wp} and σ_{Wt} are reasonably well reproduced. However, we see that the individual diffractive excitation cross sections are not reproduced, nor is the elastic ones. However, for the final states in AA collisions, we are mainly interested in getting the absorptive and wounded cross section right, so even if our procedure probably can be improved, we are quite satisfied with this result.

Table I.1: Fitting the values of input cross sections for pp collisions at $\sqrt{s} = 5$ TeV and using the resulting fluctuations in a generation and different collision types. B is the elastic slope $-d \log \sigma_{\rm el}/dt|_{t=0}$. The cross sections used as "input" were taken from the default parameterisation in PYTHIA8. The line "model" shows the results of a fit to the model in eqs. (I.11) – (I.13). The line "generated" finally shows the result of the approximation discussed in this subsection and in the appendix. The fitting procedure assumed a 5-10% uncertainty on the input values, and the statistical uncertainty on the presented output values are around and below 0.5%. The resulting parameters values in eqs. (I.11) and (I.13) were $k = 1.80$, $r_0 = 0.407$ fm, $\sigma_t = 13.88$ fm^2, and $\alpha = 0.22$.

	$\sigma_{\rm abs}$ (mb)	$\sigma_{\rm Wp}$ (mb)	$\sigma_{\rm Wt}$ (mb)	$\sigma_{\rm Dp}$ (mb)	$\sigma_{\rm Dt}$ (mb)	$\sigma_{\rm DD}$ (mb)	$\sigma_{\rm el}$ (mb)	B (GeV^{-2})
input	47.7	61.5	61.5	6.1	6.1	7.7	18.4	20.8
model	47.8	61.4	61.5	5.7	5.8	7.9	18.7	24.1
generated	47.8	61.3	61.3	11.4	11.4	2.2	-	-

3 From wounded nucleons to exclusive final states

In the wounded nucleon model, as formulated by Białas and Czyz [31], each wounded nucleon contributes to the final state multiplicity distribution, according to a single nucleus emission function $F(\eta)$, giving a total multiplicity of:

$$\frac{dN_{ch}}{d\eta} = w_p F(\eta) + w_t F(-\eta). \tag{I.14}$$

Here $w_{p|t}$ denotes the number of wounded nucleons from left and right respectively, calculated for a given centrality class, defined by impact parameter. In the wounded nucleon model, $F(\eta)$ must be extracted from data, and depends on centrality class [44], but a crucial feature of the model is that eq. (I.14) reduces to the pp multiplicity distribution for $w_p = w_t = 1$.

The Angantyr prescription for generating exclusive final states has conceptual similarities with the wounded nucleon model. But instead of extracting an emission function from data, MPI events from PYTHIA8 are used. We will in this section briefly review the PYTHIA8 MPI model, and motivate the addition of additional MPIs from multiple wounded nucleons to the model.

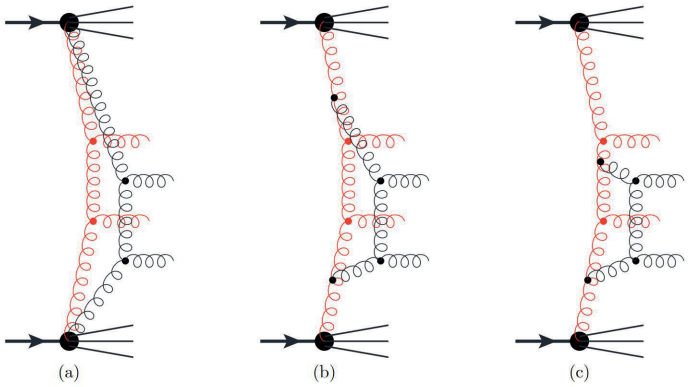


Figure I.2: Schematic pictures of multi-parton interactions in a pp collision. The y-axis should be interpreted as rapidity. All initial- and final-state radiation has been removed to avoid cluttering. Each gluon should be interpreted as having two colour lines associated with it, which in the subsequent string hadronisation will contribute to the soft multiplicity. In (a) the colour lines for both sub scatterings stretches all the way out to the proton remnants, while in (b) and (c) the secondary scattering is colour-connected to the primary one.

3.1 Multiparton interactions in pp collisions

In the PYTHIA8 MPI model [12], all partonic sub-collisions are to a first approximation treated as separate QCD $2 \to 2$ scatterings[4]. Since the cross section diverges at low $p_\perp$, it is regularised using a parameter $p_{\perp 0}$ which depends on the collision energy, giving:

$$\frac{d\sigma_{2\to 2}}{dp_\perp^2} \propto \frac{\alpha_s^2(p_\perp^2)}{p_\perp^4} \to \frac{\alpha_s^2(p_\perp^2 + p_{\perp 0}^2)}{(p_\perp^2 + p_{\perp 0}^2)^2}. \tag{I.15}$$

This cross section is then folded with parton densities to get a relative probability for each additional sub-scattering. The densities are rescaled according to an overlap function using some assumption about the matter distribution in the colliding protons and an assumed impact parameter.

In Figure I.2a there is an illustration of an event with two sub-scatterings (in red and black) which we have assumed are both of the type $gg \to gg$. Note that in the PYTHIA MPI model all incoming and outgoing partons would be dressed up with initial- and final-state radiation, but these have been left out of the figure to avoid cluttering. With completely uncorrelated sub scattering, one would assume the colours of the incoming gluons would also

[4]The MPIs are not fully uncorrelated, as momentum conservation needs to be obeyed, and the parton density corresponding to the extracted parton, is rescaled by a factor $(1 - x)$.

be uncorrelated, and since each gluon carries both colour and anti-colour one would naively think that in the subsequent hadronisation phase, there would be four strings stretched between the proton remnants and giving rise to particle production over the whole available rapidity range. Again to avoid cluttering of the figures, we ask the reader to simply imagine two colour lines (strings) stretched along each gluon and that the vertical axis can be loosely interpreted as rapidity.

Already in the original paper [12] it was realised that it was basically impossible to reproduce data if each sub-scattering was allowed to add particles in the whole available rapidity range. Especially sensitive to this was the multiplicity dependence of the average particle transverse momenta, and to rectify this the MPI model in PYTHIA was modified so that additional sub-scatterings almost always was colour connected to outgoing partons in previous sub-scatterings. This is illustrated in Figure I.2b and c, where the colour correlation between the two sub-scatterings gives rise to a colour flow *as if* they were (perturbatively) connected. In this way the multiple scatterings can give rise to increased average transverse momentum from the partons coming from extra sub-scattering, without increasing the multiplicity of soft particles due to the strings stretched all the way out to the proton remnants.

3.2 Multi-parton interactions in a $\mathrm{p}A$ collision

We now turn to the case of a $\mathrm{p}A$ collision and imagine the projectile proton interacting absorptively with two nucleons in the nuclei. To be true to the PYTHIA MPI model we should simply redefine the overlap function using the matter distribution of the two target nucleons. In principle this can surely be done, however, technically we found it almost forbiddingly difficult.

Instead we note that the handling of colour correlations in the pp model would typically result in string topologies corresponding to the sketch in Figure I.3a. The primary scattering looks like normal scattering between the projectile and one of the target nucleons, while the secondary scattering is now between the projectile and the other target nucleon. Since both target nucleons have been found to be absorptively wounded, the secondary scattering must be colour connected to the second target nucleon, while in the direction of the projectile it looks like a normal secondary scattering.

We also note that we would get the same colour topology, and hence the same distribution of particles, if the second sub-scattering was a separate single (high-mass) diffractive excitation event, which in PYTHIA8 is handled as a Pomeron-proton collision. This is illustrated in Figure I.3b, where the Pomeron is shown as a green zigzag line. A secondary absorptive wounded nucleon thus contributes to the final state *as if* the final state particles were produced in a single diffractive excitation. This similarity is what we, in the following, will exploit to build up a final state from primary absorptive interactions and secondary

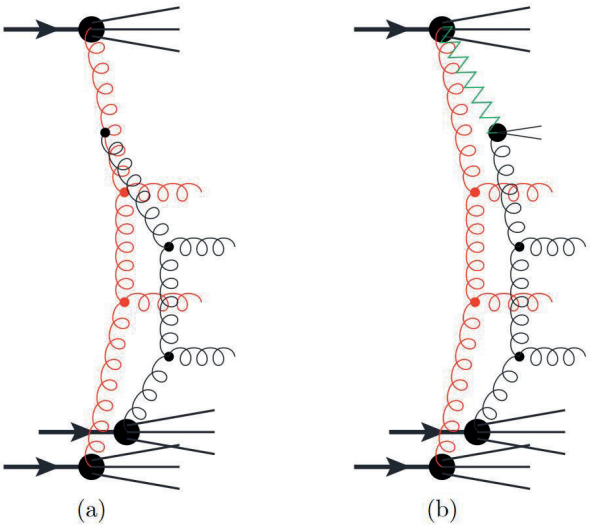


Figure I.3: A schematic picture (*c.f.* Figure I.2) of multiple scattering between one projectile and two target nucleons (*e.g.* in a pd collisions). In (a) the second interaction is directly colour connected to the first one, while in (b) the second nucleon is only diffractively excited by a Pomeron exchange. Both cases give rise to final string configurations that will contribute in the same way to the final state hadron distribution.

absorptive interactions, the latter being modelled as single diffractive excitation.

The procedure will therefore be to decide which of the two absorptive interactions is to be considered the primary one, and treat this as a completely normal non-diffractive multiple scattering event in PYTHIA. The secondary scattering will be generated as a single diffractive excitation event in PYTHIA. Also here there may be additional multiple parton scatterings, but they will be treated as multiple scatterings in the Pomeron–proton system, which is standard in the high-mass diffraction machinery in PYTHIA.

Referring back to eq. (I.14), this means that we are modelling the single nucleus emission function $F(\eta)$ using high-mass diffractive excitation events. We do not expect them to necessarily look like ordinary diffractive event, but we nevertheless use the diffractive machinery in PYTHIA8. In section 5 we will describe how we modify this machinery in order to try to fulfil the requirement that $F(\eta) + F(-\eta)$ (*i.e.* $w_p = w_t = 1$ in eq. (I.14)) would reproduce the distribution in a normal non-diffractive pp event in PYTHIA8.

The two different sub-events are then merged together so that the elastically scattered proton in the diffractive event is discarded, and the momentum of the Pomeron is instead taken

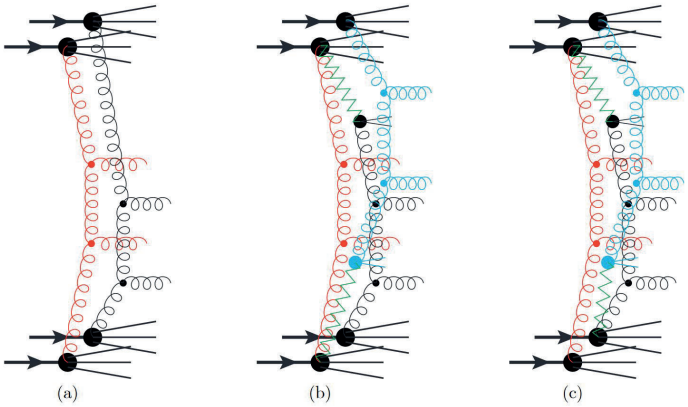


Figure I.4: A schematic picture (*c.f.* Figure I.3) of multiple scatterings between two projectile and two target nucleons in an AA collision. In (a) there are two separate NN collisions, while in (b) and (c) there is one primary sub-collision and two secondary ones.

from remnants of the projectile proton.

The assumption in [9] was that the momentum fraction of the Pomeron in such diffractive events can be taken to be distributed approximately as $dx_{\mathbb{P}}/x_{\mathbb{P}}$, which means that the mass of the diffractive system is given by dM_X^2/M_x^2. This is approximately what one has found for normal high-mass diffractive events and it is the same assumption as in the old Fritiof model. We do not have a solid explanation why this should be the case. In [9] we gave some handwaving arguments based on AGK cutting rules and the similarity between triple-Pomeron diagrams in diffractive NN scatterings and (doubly) non-diffractive proton–deuterium scattering, but in the end the best argument for this choice is that it seems to work very well.

3.3 Multi-parton interactions in an AA collision

Going one step further in complexity we now consider AA collisions. In Figure I.4 we illustrate the situation when two nucleons in one nucleus collides with two nucleons in the other, and all four possible NN interactions are absorptive. We find that there is three ways of doing this which are consistent with our pA model. Either, as in Figure I.4a, we can model it as two primary absorptive interactions, or as one primary and two secondary interactions, where the second of these can either be coupled to the primary interaction (b) or to the first secondary one (c).

All three cases will give us four absorptively wounded nucleons, and in Fritiof and the original wounded nucleon model there would be no distinction between the cases. In the Angantyr model we do, however, want to differentiate between these, and in the following section we will describe a procedure to classify all NN interactions in a AA collisions.

4 Generating and combining parton-level NN events

In general, each nucleon in the projectile nucleus may interact with several nucleons in the target nucleus an vice versa. When building up the final state by stacking parton level nucleon–nucleon events we need to concentrate on the most important ones first. This is in line with the general philosophy in PYTHIA8, that harder processes always are considered before softer ones. After having gone through all pairs of projectile–target nucleons and determined their interactions as outlined in the previous section, we therefore order all these interactions in increasing nucleon–nucleon impact parameter, $b_{\mu\nu}$.

We will then go through this list several times, treating one kind of interaction at the time, starting with the absorptive interactions, as they will give the back bones around which we will build up the full event. As soon as an NN interaction has been selected a corresponding sub-event will be generated with the standard PYTHIA8 minimum bias model, and the corresponding nucleons are marked as already interacted. If an NN interaction is found in the list where one of the nucleons has already interacted, this will be labelled *secondary* and the generated sub-event will be added to the sub-event to which the already interacted nucleon belongs, as described in [9] and detailed below.

4.1 Selecting primary absorptive collisions

The first pass over the potential NN interactions, we will only look at absorptive interactions. This will give us a set of $N^{'}_{\mathrm{abs}}$ primary absorptive collisions and a set of $N^{''}_{\mathrm{abs}}$ secondary ones (where one of the nucleons already has already been absorptively wounded in another interaction).

For each of the primary ones we now generate an inelastic non-diffractive minimum bias event in PYTHIA8, each of which will give a separate sub-events. However, since the procedure takes sub-collisions with small $b_{\mu\nu}$ first, the primary absorptive events should typically be a bit harder and have higher multiplicity than the secondary ones. In [9] this was handled by telling PYTHIA8 to generate $N^{'}_{\mathrm{abs}} + N^{''}_{\mathrm{abs}}$ events, but only keeping the $N^{'}_{\mathrm{abs}}$ ones with smallest impact parameter (as reported by PYTHIA8). For the method described here we have instead implemented directly in PYTHIA8 a way to specify by hand which impact parameter you want a given minimum bias event to have, which makes thing a bit more

efficient, and also gives a noticeable improvement on the description of some observables, as discussed below in section 6.

Just as in standard PYTHIA8 it is easy to specify signal processes rather than only consider minimum bias events. This may be used to simulate triggers on hard jets more efficiently, or to *e.g.* produce Z-tagged jets in central AA collisions [45] or top events in pA collisions [46] or AA. The way this is done is simply to substitute the hardest absorptive primary event with a corresponding signal event, and reweighting the event with a factor

$$w_{\mathrm{signal}} = \frac{(N'_{\mathrm{abs}} + N''_{\mathrm{abs}})\sigma^{NN}_{\mathrm{signal}}}{\sigma^{NN}_{\mathrm{abs}}} \tag{I.16}$$

to get the correct cross section. For signal processes with a large cross section the possibility to have additional signal processes in the same event is also taken into account, however for technical reasons at most N'_{abs} signal sub-events can be included in each event[5].

In the current implementation we assume that minimum bias processes are basically iso-spin invariant, and all such sub-events are generated as pp events in PYTHIA8, flipping by hand the iso-spin of a remnant quark or di-quark afterwards in case the corresponding nucleon was actually a neutron, to conserve total charge. Signal processes are, however, not necessarily isospin invariant. To account for this, we generate pp, pn, np, and nn collisions separately for all signal processes. To decide what type of collision should be generated, all nucleons in the colliding nuclei are marked as either protons or neutrons, under the assumption that neutrons and protons are distributed evenly in the nucleus.

One should note that measurements of proton and neutron distributions in *e.g.* lead at low energies [47] have indicated that the neutron distribution reaches further out than the proton distribution, giving rise to a "neutron skin" effect. It has been pointed out [48] that this could give rise to effects at the 10% level in selected observables in peripheral PbPb collisions. It has also been pointed out [49] that one could in principle use this effect to design different centrality measures, especially in the case of asymmetrical collision systems. Currently we know of only one very recent Glauber calculation including such effects [50], and in the present version we have left them out entirely[6].

4.2 Adding secondary absorptive interactions.

Once the back-bone sub-events have been generated we go through the list again, this time only looking at the secondary absorptive interactions, in which one of the participating nucleons has already been included in a generated primary absorptive sub-event. As described

[5]For most use cases this should be adequate, as $\sigma^{NN}_{\mathrm{signal}} \ll \sigma^{NN}_{\mathrm{abs}}$ for most processes of interest

[6]An interested user can, however, plug in their own Glauber MC including neutron skin effects.

in [9] we will generate these secondary absorptive sub-collisions as if they were single diffractive excitation events. We here use the standard PYTHIA8 diffraction machinery, but with important modifications detailed in section 5 below.

The final state generated for a given secondary absorptive interaction is then added to a primary absorptive sub-event. The elastically scattered proton is removed and the energy and momentum it had given to the excited nucleon is instead taken from the remnants of the nucleon in the primary sub-events.

It may very well happen that there is not enough energy left in the remnants in the primary sub-event to allow for the addition of a diffractively excited state. In that case it is possible to try again and maybe generate a diffractive event with lower M_X. There is a parameter in the program the limits the number of tries allowed, and if the maximum is passed, the corresponding secondary absorptive interaction is simply discarded (although the corresponding nucleon still has the chance to become wounded in another secondary interaction).

The way secondary nucleon interactions are selected according to the NN cross sections, does not take into account possible effects of energy-momentum conservation, therefore it makes sense to try to take such effects into account in this *a posteriori* way. The parameter we introduced should not be taken as the final word in the matter, but at least it allows us to investigate the effects of energy-momentum conservation.

4.3 Adding diffractive interactions

Having taken care of all absorptive interactions we continue with diffractive interactions in much the same way. For each type we again go through the impact-parameter-ordered list of NN interactions twice. In the first round, we only consider *primary* interactions, *i.e.* where neither of the nucleons have previously been included in a sub-event, and generate a sub-event which could be a single or a double diffractive excitation. These are treated as (soft) diffractive events in PYTHIA8, as discussed in section 5.

In the second round we also consider *secondary* interactions, where one of the nucleons has already been treated, and an appropriate contribution from the other nucleon (which we will here call a *half* event) is generated and added to the corresponding previous sub-event.

As an example consider an already wounded nucleon in the projectile nucleus, which interacts with a previously unwounded nucleon in the target. The wounded nucleon is already connected to another target nucleon, and cannot be further excited. There are three possibilities for the diffractive interaction:

1. The new interaction is a single diffractive excitation of the *target* nucleon. The inter-

action is then treated as a normal single diffractive excitation of the target nucleon.

2. The new interaction is a single diffractive excitation of the *projectile* (already wounded) nucleon. In this case the target nucleon is elastically scattered.

3. The new interaction is a double diffractive excitation. In this case the already wounded projectile nucleon is not modified, and the interaction is again treated as a single diffractive excitation of the *target* nucleon.

In a final iteration[7] also purely elastic interactions are considered, and here again the half events are single elastically scattered nucleons. In each case energy and momentum conservation is handled in the same way as for secondary absorptive interaction.

Modulo the effects of secondary interactions being discarded due to energy-momentum conservation, this procedure will correctly handle the probability that a given nucleon is wounded in some way. Note however that, as discussed in section 2.5, although some nucleons in the program are classified as elastically scattered, elastic scattering is not included properly. As elastic scattering is a coherent effect of shadowing due to absorption, the Good–Walker formalism can be used to calculate the cross section for elastic scattering of the incoming *nuclei*, but not for individual nucleons in a nucleus. [8] Diffractive excitation of individual nucleons can, however, be calculated via the trick described in section 2.5.

In the end we have generated a set of parton-level sub-events, which we now can join together in a single parton-level AA event. This event is then handed back to PYTHIA8 for hadronisation and decay of unstable hadrons. Finally the non-interacting projectile and target nucleons are bunched together in two remnant nuclei.[9]

5 Modifications of single diffractive to secondary absorptive

In section 3 and in ref. [9] we argued that secondary absorptive interactions will contribute to particle production in the same way as a single diffractive (SD) excitation event (*c.f.* Figure I.3). Assuming that such SD events produce a simple flat string with mass distributed as dM_X^2/M_X^2, this would naively give a triangular shape of the $F(\eta)$ wounded nucleon emission function in eq. (I.14).

We will use the SD excitation machinery in PYTHIA8, where at high energies the diffractive systems are much more complicated than a single string. As described in more detail below,

[7] Note that central diffraction is not handled properly in the current version of the program.

[8] Naturally electromagnetic interaction, not included here, is responsible for most of the coherent elastic nucleus scattering.

[9] The nucleus remnants are in the event record given the name *NucRem* and PDG-id codes on the form $100ZZZAAA9$, which in the PDG standard corresponds to a highly excited nucleus.

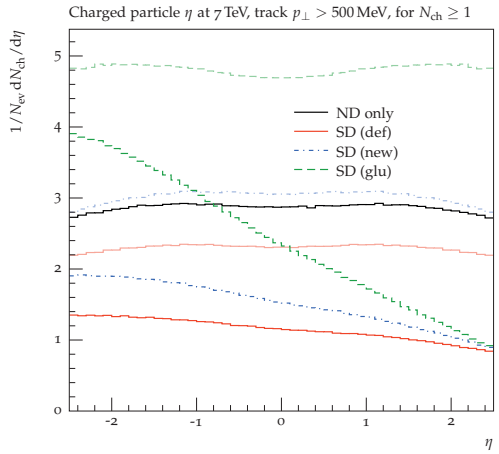


Figure I.5: Illustration of the shape of the multiplicity function in eq. (I.14) using the η distribution as measured by ATLAS in [51]. The black and red lines are the shapes of standard non-diffractive and single diffractive events from PYTHIA8 respectively. The green dashed and blue dash-dotted lines are single diffractive events generated by PYTHIA8 using the modifications presented in [9] and the modifications presented in this article respectively. For each single diffractive line there is also a pale line corresponding to adding the mirror image to emulate a non-diffractive distribution à la Fritiof.

it models the diffractive excitation as a non-diffractive (ND) interaction between the target nucleon and a Pomeron emitted from the projectile (in the spirit of Ingelman and Schlein [13]), and this is then treated with the full MPI machinery as if the Pomeron was a hadronic object with parton densities. In Figure I.5 we show the average multiplicity as a function of pseudo-rapidity for ND events, and compare it to SD events from PYTHIA8 using the default settings. Clearly we get a somewhat triangular shape for the SD events (SD(def) in the figure), and adding the multiplicity from target and projectile excitation, we get a shape similar to the ND shape, fully in accordance with eq. (I.14) in the case of $w_p = w_t = 1$.

In ref. [9] we noticed that using the default PYTHIA8 SD machinery for secondary absorptive collisions resulted in too low activity in pA and tried different modifications to increase the multiplicity. One of these modifications included increasing the gluon density in the Pomeron, which is also shown in Figure I.5 (SD(glu)).

Here we will try to be more systematic in our approach to modify the default PYTHIA8 SD machinery. Looking at Figure I.3b, it is clear that the rapidity region close to the one close to the direction of the two target nucleons will be our main focus. Here we note that we could equally well have chosen the second nucleon to be in the primary interaction and the first nucleon to be in the secondary, and would then want to have the same distribution of

particles. This means that we want the single diffractive event to look as much as possible as a non-diffractive event close to the direction of the two nucleons. We have therefore investigated several different modifications of the SD model and for different diffractive masses we have studied particle distributions in different pseudo-rapidity intervals and compared these with the corresponding particle distributions in the same intervals for ND events.

In the end we settled for a new modification (labelled SD(new) in Figure I.5), which is the default way of generating secondary absorptive interactions as of version 8.235 of PYTHIA8[10]. To motivate this, we first need to take a closer look at the SD machinery in PYTHIA.

5.1 High-mass diffractive excitation and secondary absorptive

There are more than one way of generating diffractive events in PYTHIA8, but here we will only concern ourselves with the *soft* diffraction used for minimum bias events. Also here there are two treatments depending on the mass, M_X. For low masses, $\lesssim 10$ GeV, the excited system is modelled as a simple longitudinally stretched string. In an AA collision, such small excitations will typically be mixed up with the nucleus remnants in the very forward and backward regions and we will here mainly concentrate on high-mass diffraction, which contributes also in the central rapidity region as seen in Figure I.5.

For high-mass diffraction, PYTHIA treats a proton-Pomeron collision as a normal non-diffractive (ND) hadron-hadron collision and uses the whole MPI machinery with initial- and final-state parton showers. This means that there will be multiple $2 \to 2$ semi-hard partonic scatterings given by

$$d\sigma_{ij}^{\mathrm{p}\mathbb{P}}(p_\perp^2) = \frac{dx_{\mathbb{P}}}{x_{\mathbb{P}}}\frac{dx_1}{x_1}\frac{d\beta}{\beta}F(x_{\mathbb{P}})x_1 f_i^{\mathrm{p}}(x_1,p_\perp^2)\beta f_j^{\mathbb{P}}(\beta,p_\perp^2)d\hat{\sigma}_{ij}(p_\perp^2). \tag{I.17}$$

Here $x_{\mathbb{P}}$ denotes the fraction of the target proton momentum taken by the Pomeron; β is the fraction of the Pomeron momentum taken by the parton j; and x_1 is the fraction of the projectile proton momentum taken by parton i. Furthermore we have the parton densities in the proton, f_i^{p}, and the corresponding densities in the Pomeron, $f_j^{\mathbb{P}}$. Finally we have the flux factor $F(x_{\mathbb{P}})$ controlling the diffractive mass given by $M_X^2 = x_{\mathbb{P}}s$. In the following we will assume a flat distribution in $\log(M_X^2)$, in which case $F(x_{\mathbb{P}})$ is just a constant.

The partonic cross section $d\hat{\sigma}_{ij}(p_\perp^2)$ diverges for small $p_\perp^2$, and although it is regularised as in eq. (I.15) the integrated partonic cross section may still exceed the total non-diffractive p$\mathbb{P}$ cross section for a given M_X. In the PYTHIA MPI model this is then interpreted as the

[10]Normal diffractive interactions between projectile and target nucleons are treated by the usual PYTHIA8 diffraction set-up.

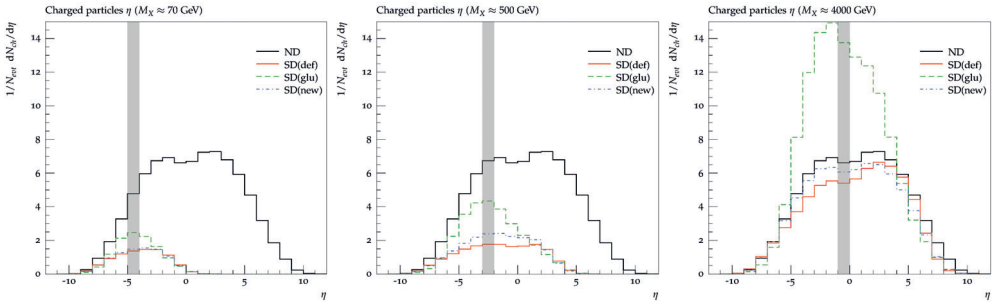


Figure I.6: Pseudo-rapidity distribution of charged particles for different diffractive masses for the default single diffraction in PYTHIA8 (red solid lines), the modifications made in [9] (green dashed lines) and the new modifications presented here (blue dash-dotted lines). Left, centre and right histograms correspond to M_X values of $\approx$ 70, 500, and 4000 GeV respectively. For comparison the results from non-diffractive PYTHIA events at $\sqrt{s} = 5$ TeV is shown as the solid black line. The shaded areas in the figure indicate the pseudo rapidity intervals where the comparisons between SD and ND particle distributions in section 5.2 were studied.

possibility of having several sub-scatterings in each collision, with the average number of sub-scatterings given by

$$\langle N_{sc}^{\mathrm{p}\mathbb{P}}(M_X)\rangle = \frac{1}{\sigma_{\mathrm{ND}}^{\mathrm{p}\mathbb{P}}(M_X)} \int \frac{dx_1}{x_1} \int \frac{d\beta}{\beta} \int dp_\perp^2 \sum_{ij} x_1 f_i^{\mathrm{p}}(x_1, p_\perp^2) \beta f_j^{\mathbb{P}}(\beta, p_\perp^2) \frac{d\hat{\sigma}_{ij}}{dp_\perp^2}. \tag{I.18}$$

Here the default value of the of the non-diffractive p$\mathbb{P}$ cross section, $\sigma_{\mathrm{ND}}^{\mathrm{p}\mathbb{P}}(M_X)$, is just set to a constant 10 mb. The $p_\perp^2$ integral is over the full available phase space, all the way down to zero, but with the $\hat{\sigma}_{ij}$ regulated as in eq. (I.15). The parameter $p_{\perp 0}$ here varies as a small power of M_X^2, in the same way as the $p_{\perp 0}$ in normal pp scatterings varies with s.

In Figure I.6 we show the resulting pseudo-rapidity distribution of charged particles for different values of M_X for diffractive events from PYTHIA8 with $\sqrt{s} = 5\,TeV$.[11] Here we see the expected behaviour with a large rapidity gap for smaller M_X, typical for diffraction. When we want to use the diffractive excitation in PYTHIA to model the secondary absorptive interactions, we want to make the event in the target proton direction to look as much as a normal non-diffractive pp event as possible, and in particular we want the whole event to look approximately the same in the limit $M_X^2 \to s$. From the figure we see that this is not quite the case for the default diffraction parameters in PYTHIA8. We also see that the modifications we presented in [9] seems to be a bit too forceful.

[11] The kinematics is given by the LHC pPb run, giving a slightly tilted distribution in η.

Looking at eqs. (I.17) and (I.18) it is easy to see that we can increase the multiplicity by either increasing the general activity by modifying the Pomeron parton densities (as is done in SD(glu) in figures I.5 and I.6), or we can try to increase the number of sub-scatterings by *e.g.* adjusting the free parameter $\sigma_{\mathrm{ND}}^{\mathrm{pI\!P}}(M_X)$. We will here look at both these options by studying eq. (I.18) more closely. Studying the average number of sub-scatterings for a fixed rapidity, $y = \log(x_1/\beta x_{\mathrm{I\!P}})/2$, we get

$$\frac{d\langle N_{sc}^{\mathrm{pI\!P}}\rangle}{dy} = \frac{1}{\sigma_{\mathrm{ND}}^{\mathrm{pI\!P}}(M_X^2)} \int \frac{dx_1}{x_1} \int \frac{d\beta}{\beta} \int dp_\perp^2 \sum_{ij} x_1 f_i^{\mathrm{p}}(x_1, p_\perp^2) \beta f_j^{\mathrm{I\!P}}(\beta, p_\perp^2) \times \frac{d\hat{\sigma}_{ij}}{dp_\perp^2} \delta\left(y - \log\frac{x_1}{\beta x_{\mathrm{I\!P}}}\right) \quad \text{(I.19)}$$

If we now compare this to the same for standard non-diffractive pp events,

$$\frac{d\langle N_{sc}^{\mathrm{pp}}\rangle}{dy} = \frac{1}{\sigma_{\mathrm{ND}}^{\mathrm{pp}}(s)} \int \frac{dx_1}{x_1} \int \frac{dx_2}{x_2} \int dp_\perp^2 \sum_{ij} x_1 f_i^{\mathrm{p}}(x_1, p_\perp^2) x_2 f_j^{\mathrm{p}}(x_2, p_\perp^2) \times \frac{d\hat{\sigma}_{ij}}{dp_\perp^2} \delta\left(y - \log\frac{x_1}{x_2}\right) \quad \text{(I.20)}$$

we see immediately that if we modify the Pomeron parton density and make it $x_{\mathrm{I\!P}}$-dependent, $\beta f_j^{\mathrm{I\!P}}(\beta, p_\perp^2) \to x_{\mathrm{I\!P}} \beta f_j^{\mathrm{p}}(x_{\mathrm{I\!P}}\beta, p_\perp^2)$, and at the same time make the total non-diffractive pI\!P cross section as well as the soft regulator, $p_{\perp 0}$, independent of M_X, *i.e.*, $\sigma_{\mathrm{ND}}^{\mathrm{pI\!P}}(M_X^2) \to \sigma_{\mathrm{ND}}^{\mathrm{pp}}(s)$ and $p_{\perp 0}(M_X^2) \to p_{\perp 0}(s)$, we will get very similar expressions. They will not be exactly the same, since the kinematical limits $p_\perp$ will differ, especially for small M_X. Also, for technical reasons, PYTHIA8 will adjust the selected $p_{\perp 0}$ for each M_X value to ensure that the average number of scatterings is always larger than one, effectively making low M_X events softer.

The resulting modification is shown in Figure I.6 as the lines labelled SD(new), and we see that the multiplicity in the proton direction is not much improved at small M_X, but at large M_X it traces the non-diffractive quite well.

In the next section we will look in more detail on the particle distributions in the rapidity regions where we want the secondary absorptive sub-events to resemble normal non-diffractive events in PYTHIA.

5.2 Comparing primary and secondary absorptive sub-events

From Figure I.6 we see that SD final state particles only populate the rapidity region corresponding to the colour exchange between the Pomeron and the proton (*c.f.* Figure I.3b).

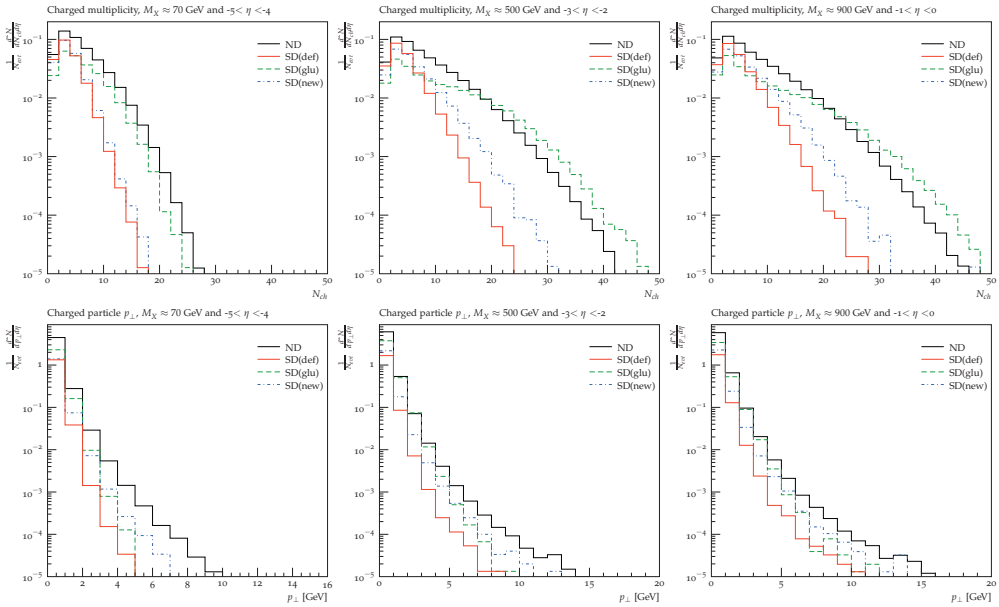


Figure I.7: Charged particle distributions in non-diffractive events (black lines marked ND) compared to different options (SD(def): red lines, SD(glu): green dashed lines, and SD(new): blue dash-dotted lines) for single diffractive excitation events in different rapidity slices and different excitation masses, M_X. The top panel shows the multiplicity of charged particles, and the bottom panel their transverse momentum distribution.

We will here investigate further to what extent the SD events generated by PYTHIA (with or without modifications) look the same as the ND events in this region. To do this we will study the distribution of particles in different pseudo-rapidity slices for different values of the diffractive mass, M_X. In these slices we have looked at standard minimum bias observables based on charged particles, such as average multiplicity (shown in Figure I.6), the distribution in multiplicity (N_{ch}), the transverse momentum distribution ($p_\perp$), the distribution in summed ($\sum p_\perp$) and average ($\langle p_\perp \rangle$) transverse momentum for particles within one unit of η, and average transverse momentum as a function of multiplicity ($\langle p_\perp(N_{ch}) \rangle$).

Naturally, we do not expect these observables to look the same for a diffractively excited system and a full non-diffractive event. Close to the rapidity gap, we are in the fragmentation region of the Pomeron remnant, and here the transverse momentum of final state particles are severely restricted by the kinematics. Also close to the proton fragmentation region, the transverse momenta are limited by kinematics, but here we expect the SD and ND events to look very similar, and indeed we find that they do.

Here we will concentrate on the rapidity regions around the plateau of each M_X, and

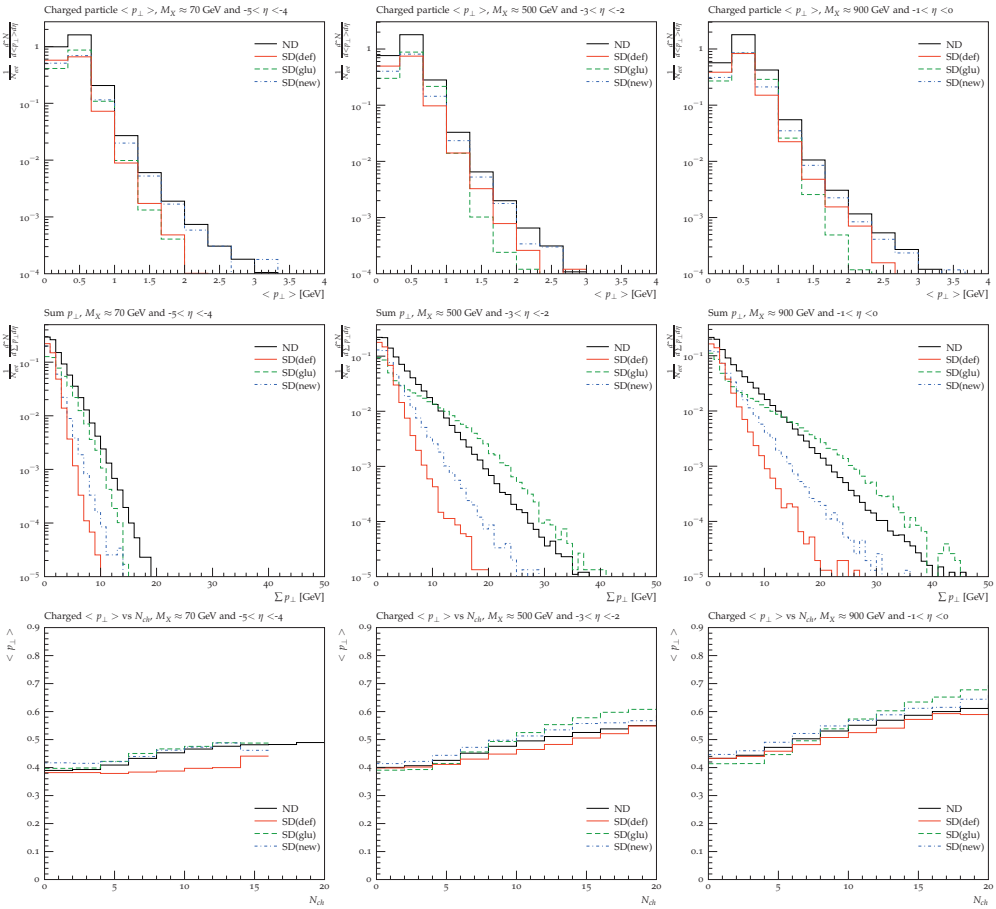


Figure I.8: As Figure I.7, but different observables. The top panel shows the average transverse momentum, while the middle one shows the summed transverse momentum and the bottom panel the average transverse momentum as a function of the multiplicity.

in figures I.7 and I.8 we show some distributions in the slices $\eta \in [-5, -4]$, $[-3, -2]$ and $[-1, -0]$ (the shaded regions in Figure I.6) for mass bins with $M_X \approx 70$, 500 and 900 GeV respectively. As for the overall multiplicity we find that the default SD machinery, (SD(def)), is quite far from the ND observables in the same rapidity slice. The SD(glu) modification is much closer, but overshoots quite significantly at large M_X in the multiplicity distribution (Figure I.7) and $\sum p_\perp$ (Figure I.8). The SD(new) curve gives a slightly better description of $p_\perp$ in Figure I.7 and the average $p_\perp$ observables in Figure I.8, but no improvement – or even a slightly worse performance – in the two remaining observables. The choice of which option to use can therefore only be based on an assessment

of what types of observables are deemed most important to reproduce correctly. In particular the dependence of the average transverse momentum on the multiplicity is known to be very sensitive to the handling of the multi-parton interactions [12], and here we see that SD(new) is quite close to the ND curves here, as may be expected from comparing eqs. (I.20) and (I.19).

The fact that the $\sum p_\perp$ distributions in SD(glu) in Figure I.8 is much harder than in standard ND events would be a problem for the description of the centrality observables used in pA and AA, which are often based on the total transverse activity in the forward/backward region (see section 6.2).

It is, however, clear that we could have put more emphasis on charged multiplicity and $\sum p_\perp$ in the regions where the SD(glu) option outperforms SD(new), and thereby made another choice of recommended option. In section 7 we will compare the three different choices against each other for pA results.

In section 4.1 we explained how the impact parameter obtained for each NN sub-collision is used as input to PYTHIA8. Here small impact parameters will lead to more multiple scatterings for primary absorptive sub-events. The same impact parameter dependence is also used for secondary absorptive sub-events. It is therefore interesting to compare the SD events with ND events for a specific impact parameter. In Figure I.9 we show typical examples of such comparisons for impact parameters slightly smaller and larger than average. Comparing with the corresponding distributions in figures I.7 and I.8, we see that the difference between the SD and ND curves tend to diminish with increasing impact parameter, which is good, since by construction the secondary absorptive interactions are at larger impact parameter than the primary ones.

In Figure I.9 we did not show curves for SD(glu), and in the following we will disregard this option completely. The modifications there are too severe and somewhat ad-hoc, resulting in far too large effects especially on particle production at high M_X (Figure I.6). We will also disregard the SD(def) option, as it produces too few particles in the nucleus' fragmentation region in pA collisions [9]. SD(new) does not give a perfect reproduction of the ND distributions, and we do not expect any SD model to do that, due to phase space constraints.

The conclusion from the analyses in this section is that SD(new) provides an overall fair description, as well as being more theoretically appealing than the other variations. The SD(new) is therefore, since version 8.235 of PYTHIA, the default model for secondary absorptive sub-events in Angantyr.

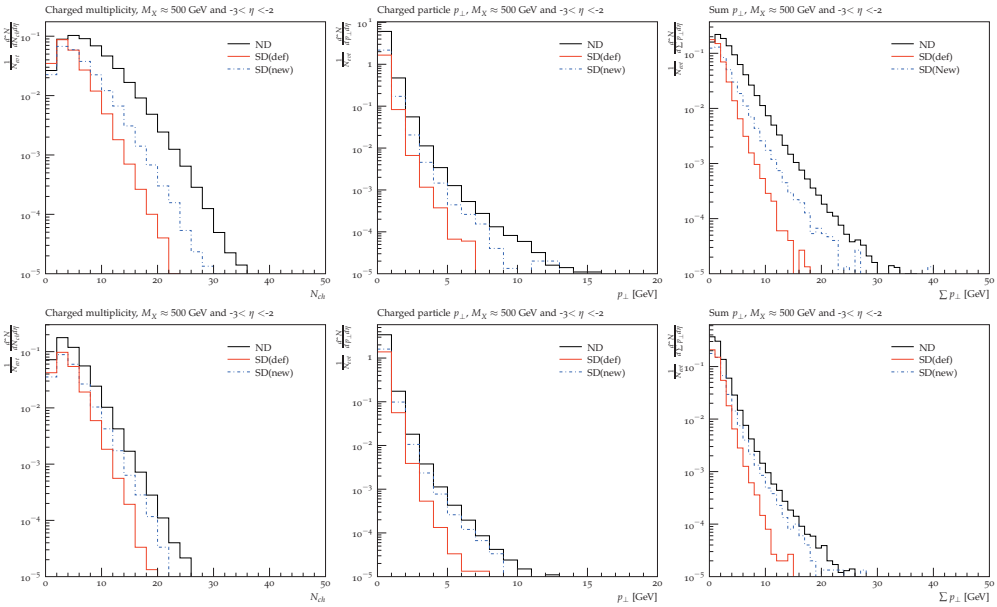


Figure I.9: Multiplicity, $p_\perp$ and $\sum p_\perp$ of charged particles for different modifications of SD events with $M_X \approx 500$ GeV compared to ND events in the pseudo-rapidity interval $-3 < \eta < -2$. All events were generated at fixed impact parameter, $b/\langle b \rangle = 0.9$ (top panel) and 1.3 (bottom panel). The lines are as in Figure I.7.

6 Sample results

All results presented here are generated with PYTHIA8 version 8.235 using default settings[12]. This means in particular that:

- the nucleon distributions in the nuclei are generated according to the formulae in [15] using the hard-core option, where parameters are tuned to low-energy eA data;
- the impact parameter is sampled using a Gaussian distribution with a width large enough to have fairly uniform weights;
- the fluctuations in the nucleons were modelled according eqs. (I.11) – (I.13), fitted the default parameterisation of semi-inclusive cross sections in PYTHIA8;
- the different NN interactions were classified using the procedure described in sections 2.5 and 4;

[12]Since Angantyr is the default heavy-ion model in PYTHIA, it suffices to specify suitable nuclei as beam particles to reproduce the results presented here.

- the sub-events were generated with the default PYTHIA8 minimum-bias machinery, except for the secondary absorptive ones, where the modifications in section 5 was used.

As with most things in PYTHIA8, there are many options beyond the default behaviour in Angantyr, and there are also so-called user hooks where the user can implement alternative models for *e.g.* the nucleon distribution, impact-parameter sampling and modelling of fluctuations. There are also a number of parameters in Angantyr that influences the generation of collisions involving nuclei, but most of these can be fitted to pp data. In fact, there are only two parameters that clearly influences the results presented here, which cannot be tuned to pp data. One is the distribution of diffractive masses used in the generation of secondary absorptive sub-events. Here we have assumed a distribution $\propto dM_X^2/M_X^{2(1+\Delta)}$ where we have simply chosen $\Delta = 0$ as in the original wounded nucleon model as implemented in Fritiof. The other was mentioned in section 4.2 and is related to energy-momentum conservation when adding secondary sub-events. The default is to simply veto a secondary NN interaction if there is not enough energy left in the corresponding remnant nucleon in the primary sub-event. An alternative is to instead generate a new secondary sub-event (regenerating M_X) to see if that one can be included.[13] Below in section 7 we will study the effects of these choices.

6.1 pp results

We begin by using the Angantyr generation for the simplest of nuclei, *i.e.* for pp collisions. Since we actually use the PYTHIA8 minimum bias machinery, we need to make sure that typical minimum-bias observables are reproduced as well when using Angantyr. We expect some differences since all semi-inclusive cross sections are not exactly reproduced in the generation, as explained in section 2.5. Furthermore the distribution in impact parameter is not the same, and since this directly affects the amount of MPI it is important to make sure that the translation between the two works, at least on average.

In PYTHIA8, the impact parameter is by default chosen according to an exponentially falling overlap function, while in Angantyr it is determined by the fluctuations and opacity functions in eqs. (I.11) – (I.13), and it is not straight forward to translate directly between the two. In principle one could try to implement the Angantyr distribution as an option in the PYTHIA8 MPI machinery, which then would require a full retuning to pp data. Here we have decided to instead implement a simple scaling factor, b_{scale}, so that for absorptive (non-diffractive) events,

$$b_{\text{Pyt}} = \frac{\langle b_{\text{Pyt}} \rangle}{\langle b_{\text{Ang}} \rangle} \frac{b_{\text{Ang}}}{b_{\text{scale}}}, \tag{I.21}$$

[13] The number of attempts allowed for this is governed by the parameter `Angantyr:SDTries`.

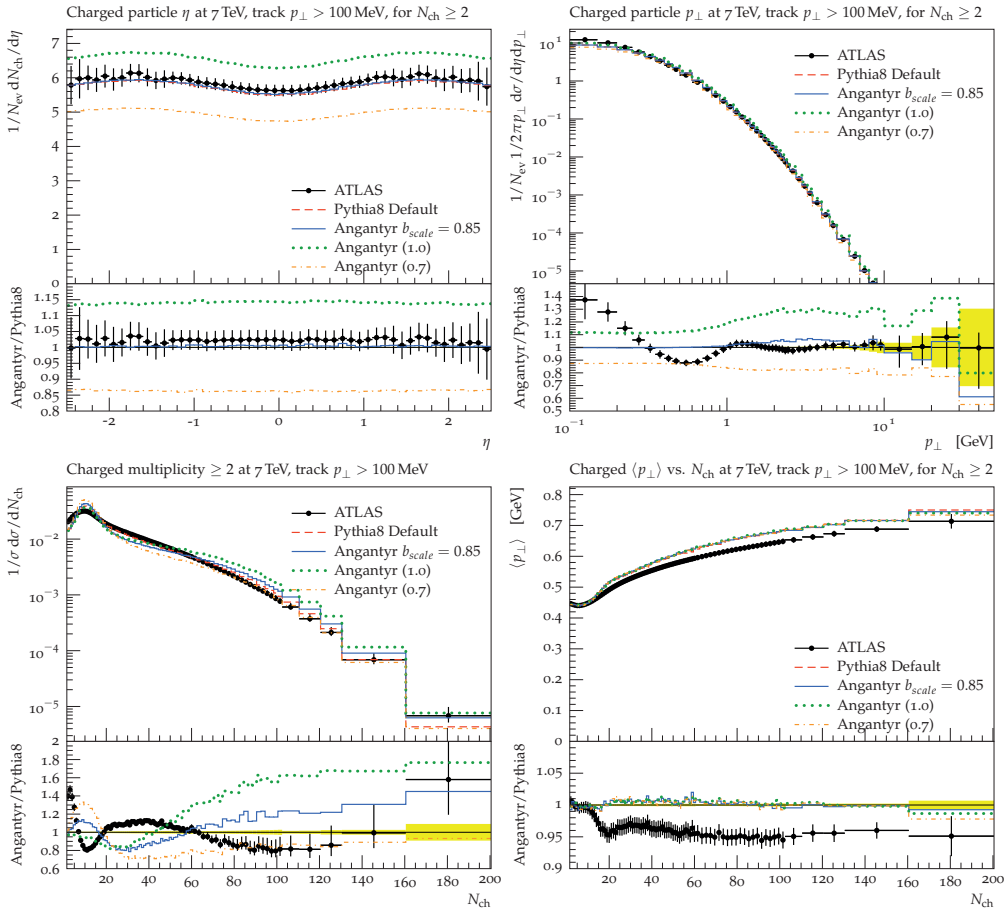
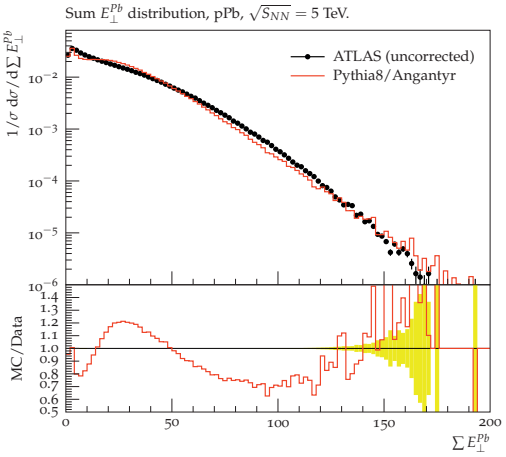


Figure I.10: The default PYTHIA8 description of some typical minimum-bias observables in pp, compared to the description using the Angantyr machinery. The latter is given for a range of values of b_{scale} (as quoted in parenthesis in the figure legend). For comparison we show data from ATLAS [51] as implemented in Rivet [52].

which is set to a value ensuring that Angantyr gives approximately the same results as PYTHIA8 for typical pp minimum-bias observables. In Figure I.10 we see that our tuned value of $b_{\text{scale}} = 0.85$ fairly well reproduces the PYTHIA8 results and gives approximately the same level of agreement with data. For comparison, the figure also shows the effect of varying this scale to $b_{\text{scale}} = 1.0$ and 0.7, as indicated in the parenthesis in the figure legend.



percentile (%)	ATLAS (GeV)	Angantyr (GeV)
0 – 1	90	87
1 – 5	66	62
5 – 10	53	51
10 – 20	41	39
20 – 30	32	32
30 – 40	24	26
40 – 60	13	16
60 – 90	6	3

Figure I.11: The summed transverse energy in the lead direction ($-4.9 < \eta < -3.2$) for pPb collisions at $\sqrt{s_{NN}}$=5 TeV. Data from ATLAS [26] is compared to results from Angantyr. The table shows the resulting bin edges when dividing up in percentiles for the experimental and generated data respectively.

6.2 p*A* results

Comparing to pA data means that we need to consider the concept of *centrality*, which is used in almost all published experimental heavy ion results. Centrality is based on a final-state observable that is assumed to be correlated with the overall impact parameter of a collision. Typically, this observable involves the activity (multiplicity, transverse energy) close to the direction of the nuclei, and other observables are then conventionally presented in bins of percentiles of this centrality observable.

We will here use the centrality observable defined by ATLAS in [26], which is based on the summed transverse energy in the pseudo-rapidity interval $[-4.9, -3.2]$. As seen in Figure I.11, Angantyr is able to reproduce fairly well the measured distribution. However, it should be noted that the experimental distribution has not been corrected for detector effects, so it is difficult to draw firm conclusions about the performance of the model.

When we want to use this centrality measure we now have the option to divide it into percentile bins using the measured distribution or the generated distribution, and since they do not exactly agree we will get somewhat different bins, as is shown in table in Figure I.11.

In Figure I.12(a) we show the average charged particle multiplicity as a function of pseudo-rapidity measured in the centrality bins defined in Figure I.11. It is important to remember that even if this is presented as the centrality dependence of the pseudo-rapidity distribution, what is in fact measured is the correlation between the transverse energy flow in the

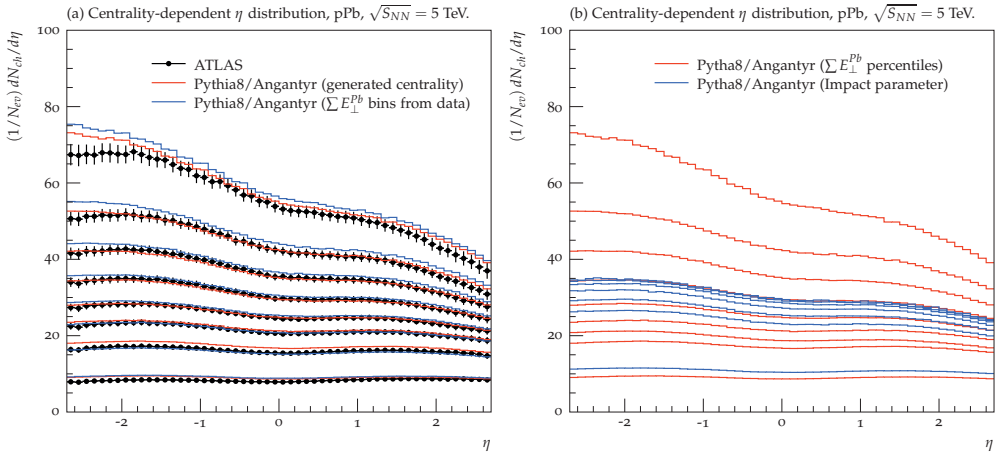


Figure I.12: Comparison between the average charged multiplicity as a function of pseudo rapidity in percentile bins of centrality for pPb collisions at $\sqrt{s_{NN}} = 5$ TeV. In (a) data from ATLAS [26] is compared to results from Angantyr. The lines correspond to the percentile bins in Figure I.11 (from top to bottom: 0–1%, 1–5%, …, 60–90%). The red line is binned using percentiles of the generated $\sum E_{\perp}^{Pb}$, and the blue line according to the experimental distribution (*c.f.* the table in Figure I.11). In (b) the red line is the same as in (a), but here the blue line uses percentile bins based on the generated impact parameter in Angantyr.

direction of the nuclei and the central multiplicity. In the figure we therefore show two sets of lines generated with Angantyr with the two different binnings presented in Figure I.11. Clearly the difference between the two is not significant, which is an indication that Angantyr fairly well reproduces the centrality measure. And the fact that neither curve is far from the experimental data[14] gives a strong indication that the Angantyr is a reasonable way of extrapolating pp final states to pA.

Comparing to the results we presented in [9], the description of data has been much improved. The main reason for this is the more careful treatment of secondary absorptive sub-events, but the new handling of the impact-parameter dependence in the primary absorptive events has also somewhat improved the description of data.

Within our model it is possible to look at the actual centrality of an event in terms of the generated impact parameter, and in Figure I.12(b) we show a comparison between the pseudo-rapidity distribution when binned in percentiles of the generated impact parameter and when binned in the generated $\sum E_{\perp}^{Pb}$ distribution. Clearly, in the Angantyr model, the binning in $\sum E_{\perp}^{Pb}$ is not very strongly correlated with the actual centrality in impact

[14] The η-distributions in Figure I.12(a) has been corrected for detector effects.

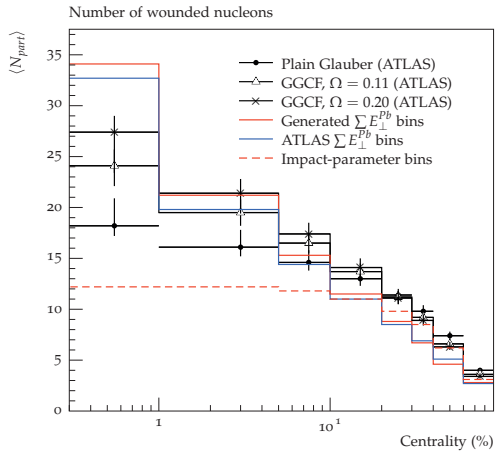


Figure I.13: Average number of wounded nucleons as a function centrality for pPb collisions at $\sqrt{s_{NN}} = 5$ TeV. The point are taken from [26] where the numbers were calculated using three different Glauber calculations: filled circles used a standard calculation without fluctuations, while triangles and crosses used the model in eq. (I.8) with fluctuations controlled by $\Omega = 0.11$ and $\Omega = 0.20$ respectively. The solid lines are generated with Angantyr binned in generated (red) and experimental (blue) $\sum E_{\perp}^{Pb}$ percentiles. The dashed line is also from Angantyr, but binned in impact-parameter percentiles.

parameter. This is especially the case for the most central collisions. The reason for this is the fluctuations modelled in Angantyr, both in the number of wounded nucleons and in the correlation between the number of wounded nucleons and the activity in the direction of the nucleus.

To study the fluctuations further we show in Figure I.13 the average number of wounded nucleons as a function of $\sum E_{\perp}^{Pb}$-centrality, both for Angantyr and for three Glauber-model fits performed by ATLAS in [26]: one using standard calculation without fluctuations, and two using the fluctuating cross sections in eq. (I.8) with different Ω-parameters. Clearly we see that Angantyr has larger fluctuations than these standard calculations. In Figure I.13 we also show the number of wounded nucleons in percentile bins of generated impact parameter. As expected the dependence is very weak for the most central bins ($0 - 30\%$), confirming here that the ATLAS centrality measure mainly picks up the fluctuations in the number of wounded nucleons in this region, and does not correlate very well with the actual impact parameter. The number of participant nucleons is a thus highly model dependent quantity, especially considering pA collisions.

Another way of studying possible nuclear effects in pA is to study particle production as a function of $p_{\perp}$. In Figure I.14 we show a comparison to CMS data. The model is clearly

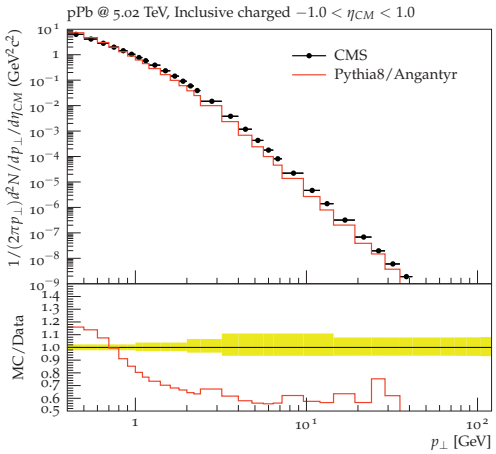


Figure I.14: The transverse momentum distribution of charged particles in the central pseudo-rapidity region in inclusive pPb events.

not perfect, but nevertheless gives a fair description of the shape over the ten orders of magnitudes shown. Comparing to the results in ref. [9] we again see an increased agreement due to the more careful treatment of secondary absorptive sub-events.

6.3 *AA* results

When we now turn to AA collisions, we expect the fluctuations to have less influence on the centrality measure, since at small impact parameters there are so many NN sub-collisions that most fluctuations will average out. It is therefore reasonable to assume that basically any centrality observable based on multiplicity or energy flow in the nuclei directions will be well correlated with the number of wounded nucleons and the actual impact parameter. Since we will now compare simulation to results from the ALICE experiment, we must in principle use the ALICE experimental definition of centrality, rather than the one from ATLAS used in the previous chapter. In ALICE centrality is defined as percentiles of the amplitude distribution obtained in the two V0 detectors, placed at $-3.7 < \eta < -1.7$ and $2.8 < \eta < 5.1$. Since this amplitude is not unfolded to particle level, and cannot be reproduced by Angantyr without realistic detector simulation, we instead construct a reasonable particle level substitute for this measure. We assume that the V0 amplitude is proportional to the total $\sum E_\perp$ from charged particles with $p_\perp > 100$ MeV in that region.

In Figure I.15 we compare the measured V0 amplitude [53] with the substitute observable, scaled to match the bin just before the distribution drops sharply at high amplitudes. The shape of the distribution is described quite well, while the normalisation is a bit off. This is likely due to difficulties extracting the data for very low amplitudes. We will throughout

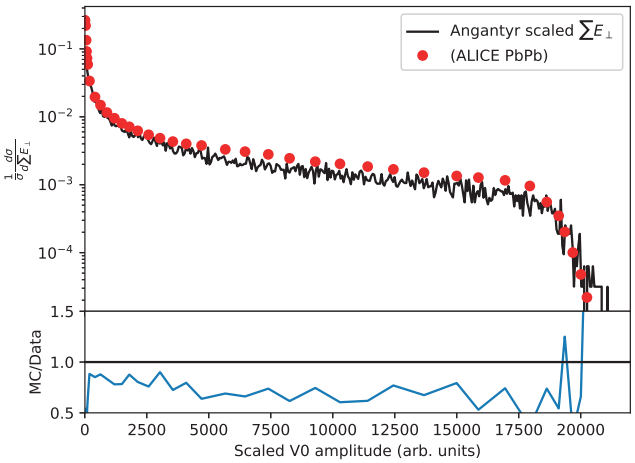


Figure I.15: Scaled $\sum E_\perp$ of charged particles at $-3.7 < \eta < -1.7$ and $2.8 < \eta < 5.1$ from Angantyr, compared with the ALICE V0 amplitude, data taken from ref. [53].

this section use this as a centrality observable, combined with the trigger setup described in ref. [53]. Furthermore, all experiments have some definition of what a primary particle is. In Figure I.12 we used the ATLAS definition where all particles with $c\tau > 10$ mm are considered as primary[15]. The ALICE definition is at its heart very similar, but has been described in more detail in ref. [54]. This definition has been conveniently implemented in Rivet [54], and we use this definition instead of a cut on $c\tau$.

In order to finish the discussion on the centrality measure, we show in Figure I.16(a) the ALICE results on the centrality dependence of the average charged multiplicity in the central pseudo-rapidity bin for PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV [53] using the measured centrality, and in Figure I.16(b) with impact parameter bins. The agreement between these two results are clearly much better in PbPb than for pPb, confirming the initial statement in this section.

In Figure I.16(a) we also show our predictions[16] for Xenon–Xenon collisions at $\sqrt{s_{NN}} = 5.44$ TeV compared to the ALICE data that were published in [55].

In Figure I.17 we show the charged multiplicity compared to ALICE data [56–58] over a

[15]This means *e.g.*, that a pair of $\pi^+\pi^-$ which comes from the decay of a K^0_S, will not be included in the charged multiplicity

[16]Although we present this after the data was published we still consider it a prediction, as the program was released before the data was analysed.

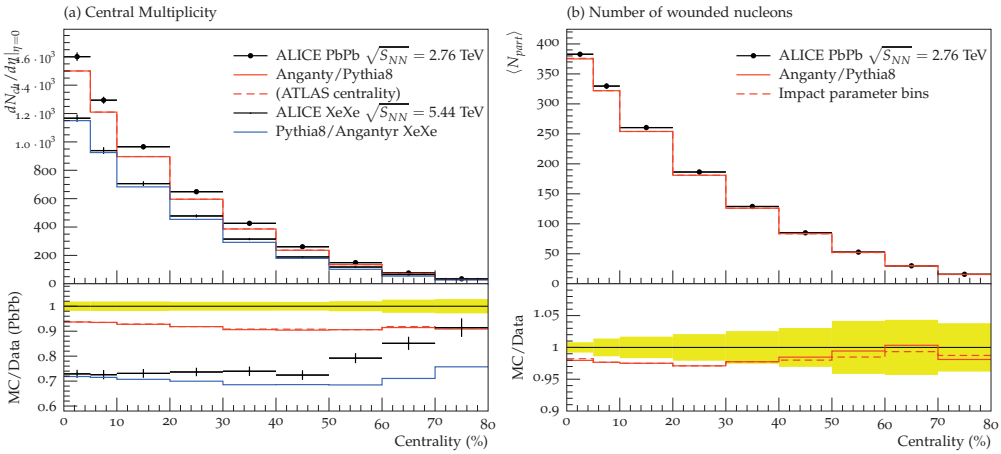


Figure I.16: (a) The centrality dependence of the average charged multiplicity in the central pseudo-rapidity bin for PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV and XeXe collisions at $\sqrt{s_{NN}} = 5.44$ TeV. Data points (for PbPb) are from ALICE [53], while red (PbPb) and blue (XeXe) lines are from Angantyr. (b) Shows the averaged number of wounded nucleons as a function of centrality. The points are from a Glauber-model calculations from ALICE [53], while the red line is the result from Angantyr. For comparison the dashed line shows the number of wounded nucleons as a function of percentiles in generated impact parameter in Angantyr.

much wider η range, for both $\sqrt{s_{NN}} = 2.76$ TeV and $\sqrt{s_{NN}} = 5.02$ TeV.

The trend, also visible in Figure I.16, is that Angantyr produces somewhat too few particles at central η; the multiplicity is systematically 5-10% too low. We regard this as surprisingly good, considered that no tuning of any kind to AA data has been done.

We now turn to transverse momentum spectra in AA collisions. In Figure I.18 we show results from ATLAS [59] compared to our model. The published $p_\perp$ spectra was scaled with the average number of wounded nucleons, calculated using a black disk Glauber model. We have not used the number of wounded nucleons as input to Angantyr, just scaled our result with the same number (as published in the article) to obtain comparable spectra. Hence, the results are not scaled to match, as both are simply scaled with the same number.

Finally we want to add a comment about the low multiplicity in the central region, shown in figs. I.16(a) and I.17. One of the main features of Angantyr is that tuning of MPI model, shower and hadronisation should only be carried out using e^+e^-, ep and pp data. However, looking at the comparison to pp in Figure I.10, we see that even the pp model undershoots the multiplicity at very low $p_\perp$ (below 500 MeV). Since ALICE measures

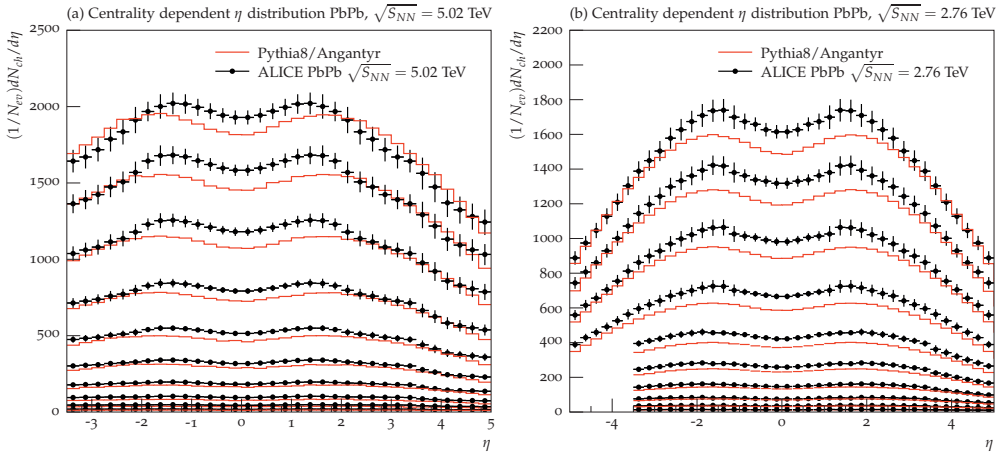


Figure I.17: The centrality dependence charged multiplicity over a wide η range in PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV (a) and $\sqrt{s_{NN}} = 2.76$ TeV (b). Both for centralities 0-5%, 5-10%, 10-20%, 20-30%...80-90%. Data from ALICE [56–58].

charged particle multiplicity all the way down to zero transverse momentum[17], it is not clear if the default PYTHIA8 behaviour should even be applicable here. The transverse momentum of such low-$p_\perp$ particles does not origin in the (perturbative) parton shower, but rather in the dynamics of string breakings. As seen from the comparison to pp this is not yet fully understood. The validity of this point is underlined by comparing to the ATLAS data shown in Figure I.18, where multiplicity is measured with low-$p_\perp$ cut–off of 500 MeV. In Figure I.19 we show the multiplicity distribution obtained by integrating the distributions measured by ATLAS, and see that the description improves.

We want to make clear that (part of) this discrepancy could of course be due to a faulty comparison to data, where triggers, centrality measure *etc.* is not implemented in exactly the way as it is done by experiments. But if it is not, it points to an interesting point for improvement of the underlying model for soft particle production, also in pp. We will return to this subject in a future paper, but meanwhile we note that it would be interesting if experiments like ALICE, who can measure very near zero $p_\perp$, will extend their publications to also include data with a minimum $p_\perp$ cut–off, which could serve as an important aide in further understanding.

[17]The multiplicity below 50 MeV is extrapolated, but this does not contribute to the total multiplicity by more than a few percent.

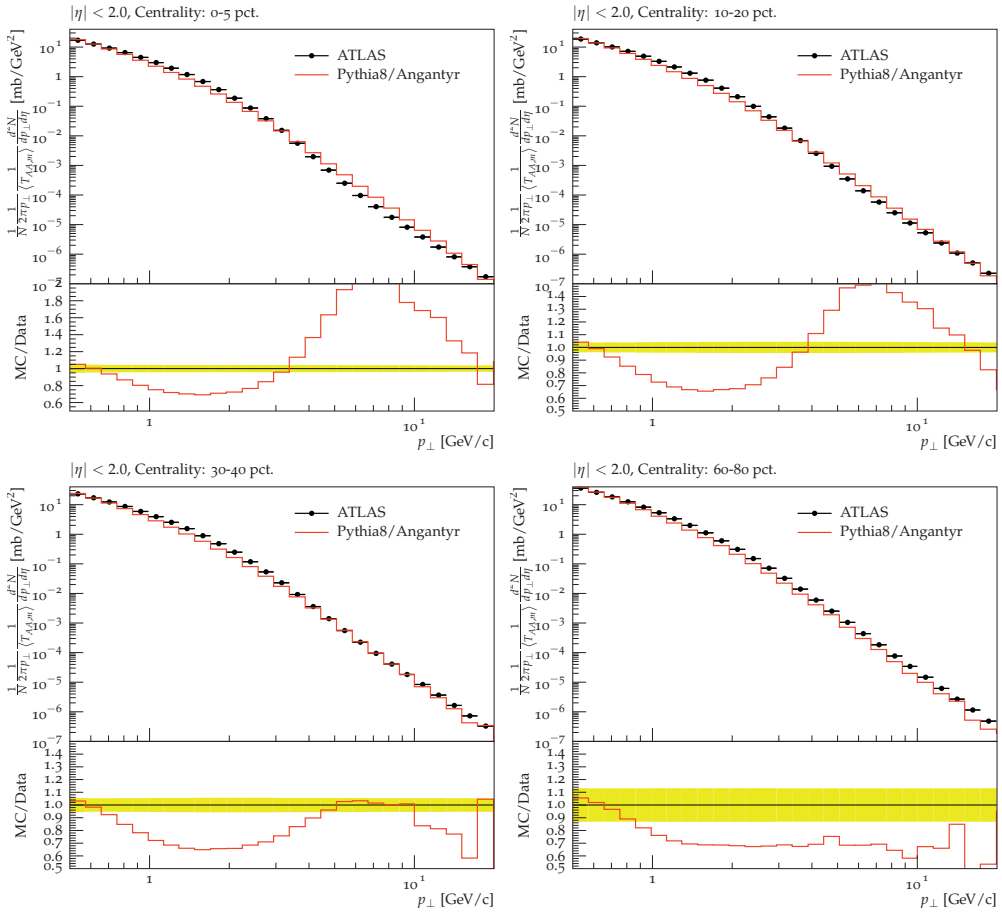


Figure I.18: Transverse momentum distributions of charged particles in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV in four centrality bins, compared to Angantyr. Data from ATLAS [59].

6.4 Collectivity and non-flow estimation

One of the primary goals of the heavy ion programs at RHIC and LHC, is to investigate the collective behaviour of final state particles produced in collisions of nuclei accelerated to relativistic energies. The anisotropic flow measures the momentum anisotropy of the final state particles. As such, it is sensitive to both the initial geometry of the nuclear overlap region, as well as the transport properties of the final state before hadronisation.

The anisotropic flow is quantified in flow coefficients v_n and corresponding symmetry planes Ψ_n, defined by a Fourier series decomposition of the azimuthal distribution of final

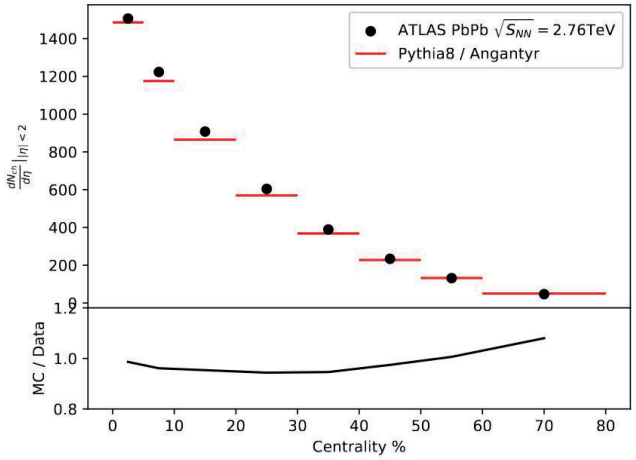


Figure I.19: Comparison to total multiplicity at mid–rapidity in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, with a minimum $p_\perp$ cut of 500 MeV, obtained by integrating the $p_\perp$ distributions measured by ATLAS [59].

state particles:

$$\frac{dN}{d\phi} \propto 1 + 2\sum_{n=1}^{\infty} v_n \cos\left[n(\phi - \Psi_n)\right]. \tag{I.22}$$

In practise, the flow coefficients are calculated using cumulants [60–62], which we also employ here. When flow coefficients are calculated using two-particle cumulants, the calculated coefficient also picks up azimuthal correlations not related to collectivity, but from *e.g.* resonance decays and intra-jet correlations. Such "non-flow" effects can be suppressed by requiring a gap in η between particle pairs.

In Figure I.20 we show $v_2\{2\}$ as function of centrality[18] measured with and without a $\Delta\eta$ gap of 1.0, by ALICE [64, 65] and CMS [63] respectively. Since Angantyr produces a full final state, it allows for the construction of the same observable, even in the absence of collective effects, giving an estimate of the non-flow present. We see that the non-flow contribution in the most central collisions is negligible (as one would expect), but rise to about 40% of the measured result for $v_2\{2\}$ without gap for peripheral collisions. This number falls to 20% when a gap is included, indicating that the method of applying a gap can remove some non-flow effects, but not all.

We want to emphasise that at this point, Angantyr does not make any attempt at modelling collective effects, and can therefore be used to estimate the contribution of non-flow. It is

[18] Using the aforementioned adapted version of ALICE centrality.

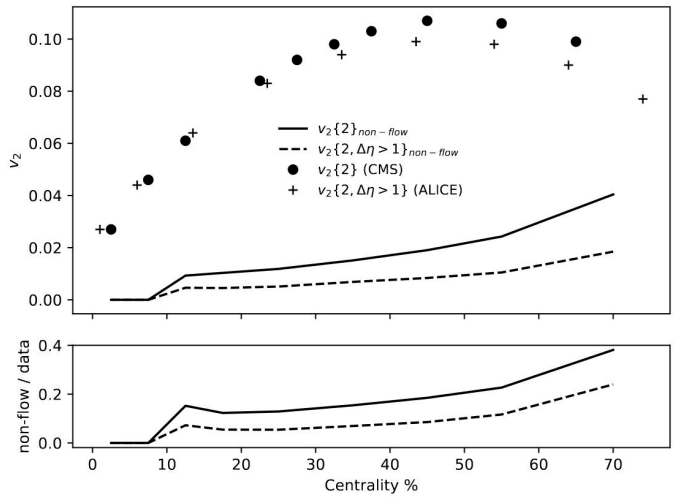


Figure I.20: The elliptic flow coefficient $v_2\{2\}$ at $\sqrt{s_{NN}} = 2.76$ TeV, as measured by CMS [63] (without $\Delta\eta$-gap) and ALICE [64, 65] (with $\Delta\eta = 1$), compared to the non-flow contribution calculated by Angantyr. In the ratio plot it is seen that the non-flow contribution without $\Delta\eta$-gap is nearly 40%. This is reduced to 20% when applying a gap.

our plan to introduce a microscopic model for collectivity, based on string–string interactions to Angantyr, which has shown promising results in pp. The increased energy density from overlapping strings would here give a transverse pressure, leading to strings "shoving" each other before hadronisation [34, 66].

7 Model uncertainties

The main idea behind Angantyr is to extrapolate pp dynamics, as described by the model for MPIs/underlying event in the PYTHIA8 MC, to heavy ion collisions, retaining as much as possible from pp. This principle was outlined already in the introduction, especially Figure I.1, but as the model has now been presented, as well as results from pA and AA collisions, we will here also discuss the model uncertainties related to this extrapolation procedure.

Primary interactions correspond directly to inelastic non-diffractive pp collisions. Here PYTHIA8, is known to reproduce most features of both soft and hard pp collisions at LHC fairly well, and the extrapolation to primary interactions in a heavy ion collision is therefore mainly a source of model uncertainty up to PYTHIA8's shortcomings in describing such

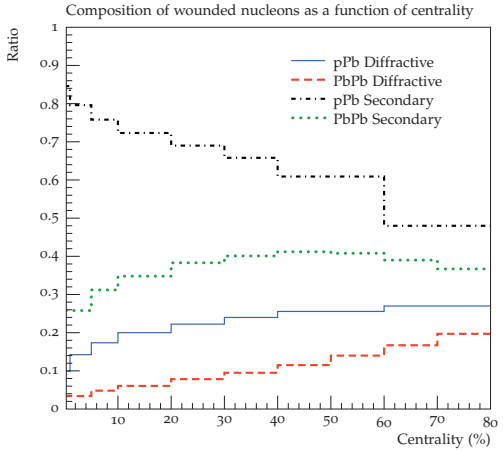


Figure I.21: The fraction of the wounded nucleons in the Angantyr model that are diffractively excited as a function of centrality for pPb at $\sqrt{s_{NN}} = 5$ TeV (blue line) and PbPb at $\sqrt{s_{NN}} = 2.76$ TeV (red dashed line). Also shown is the fraction of wounded nucleons that come from secondary absorptive interactions in pPb (black dash-dotted line) and PbPb (green dotted line).

collisions in pp. We already discussed some of those shortcomings in the previous section, but as they are not uncertainties directly related to the Angantyr model (but rather the underlying PYTHIA8 model) we will not discuss them further here.

The largest uncertainty comes instead from our treatment of secondary absorbed nucleons. The main reason is that secondary absorption has no pp equivalent. In section 5 we outlined the procedure of modifying single diffractive collisions to describe secondary absorbed nucleons, and we will investigate uncertainties related to this treatment in section 7.1.

Diffractively excited nucleons give a comparatively small contribution in collisions with nuclei, especially in central AA collisions, as illustrated in Figure I.21. Diffractive excitation of nucleons can in principle be determined from pp collisions, but as we will discuss in section 7.2, this is not straight forward.

7.1 Uncertainties in treating secondary wounded nucleons

A core feature of the Angantyr model, is that the contribution from a secondary absorbed nucleon is similar to the contribution from an excited nucleon in a single diffraction event. This corresponds to the black pieces in figures I.3a and I.3b respectively. This assumption has two components:

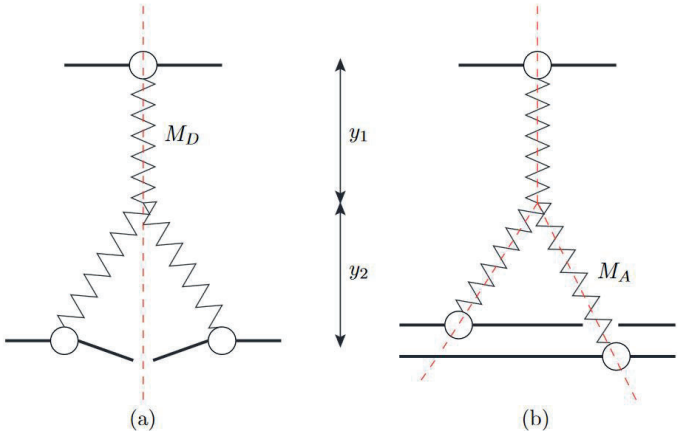


Figure I.22: Pomeron diagrams with cuts indicated for (a) single diffractive excitation in proton–proton and (b) doubly absorptive proton–deuteron scattering.

1. The distributions in the rapidity range covered, Δy, and the corresponding mass, $M \approx \exp(\Delta y/2) \times (1\mathrm{GeV})$, are similar.

2. The distribution of partons from the projectile nucleon, involved in the interaction with the secondary absorbed nucleon in Figure I.3a, is similar to the partons in the Pomeron in Figure I.3b.

Naturally none of these assumed similarities can be exact. Extracting the relevant properties in diffractive excitation in pp collisions from data at LHC has also large uncertainties, as we will discuss further in section 7.2. We also note that:

3. Energy–momentum conservation has generally important effects in high energy reactions, and has to be satisfied when nucleons suffer multiple NN sub-collisions.

Also this point is associated with some model uncertainty, as discussed in section 4.2 and in section 7.1.3 below.

In the following we will discuss the uncertainties associated with all three choices in the treatment of secondary wounded nucleons, and their impact on model predictions. We will focus on pA collisions, where there can at most be a single primary interaction, and the treatment of secondary interactions consequently has a relatively larger effect. This is illustrated in Figure I.21, where we see that secondary absorbed nucleons correspond to about 80% of all wounded nucleons in central pPb collisions, but only about 25% in central PbPb collisions.

7.1.1 Mass distribution

We begin by discussing point (i), the mass distribution of secondary wounded nucleons. The picture in Figure I.3a has the structure of a triple-Pomeron diagram. This similarity is somewhat symbolic, as each chain in this figure includes the multiple parton scatterings in Figure I.2, which correspond to Pomeron loops in a Reggeon field theoretical approach (see *e.g.* refs. [67–69]). The triple-Pomeron diagrams shown in Figure I.22 would have a weight proportional to:

$$\begin{aligned} dy_1 dy_2 \delta(y_1 + y_2 - Y) \exp(\Delta(y_1 + 2y_2)) \quad &= \\ &= \frac{ds}{s^{(1-2\Delta)}} \frac{dM_D^2}{(M_D^2)^{(1+\Delta)}} \qquad \text{for diffractive excitation,} \\ &= \frac{ds}{s^{(1-\Delta)}} \frac{dM_A^2}{(M_A^2)^{(1-\Delta)}} \qquad \text{for secondary absorption.} \end{aligned} \tag{I.23}$$

Here y_1 and y_2 are the rapidities indicated in the figure, and $Y = y_1 + y_2 \propto \ln(s)$ is the total allowed rapidity range. The quantity $M_D \propto \exp(y_1/2)$ is the diffractively excited mass to the left, and $M_A \propto \exp(y_2/2)$ is the mass of the secondary absorbed nucleon to the right. Finally the expression $1 + \Delta = \alpha_{\mathbb{P}}(0)$ is the intercept of the Pomeron trajectory.

As discussed above, in the default version of Angantyr we assume a mass distribution $\propto dM^2/M^2$ for both diffractively excited and secondary absorbed nucleons, corresponding to a critical Pomeron with $\Delta = 0$. With a hard BFKL-like Pomeron one could imagine a positive Δ in the range $0 < \Delta < 0.2$. In Figure I.23 we show the result of generating the secondary absorptive sub-events with $\Delta = 0$, 0.1, and 0.2. From the $\sum E_{\perp}^{Pb}$ distribution in pPb at $\sqrt{s_{NN}} = 5.02$ TeV (used by ATLAS as centrality measure) shown in Figure I.23a, we see a noticeable effect already below 50 GeV. The effect follows the expectation that a larger Δ will give larger M_A values and thus more activity. However, we also see that above 50 GeV the distributions for larger Δ seem to run out of steam, which we attribute to the fact that higher M_A values mean that the energy available from the projectile proton is used up faster. This means that fewer secondary absorptive interactions are accepted. In Figure I.23b we also show the resulting pseudo-rapidity distribution of charged particles for two centrality bins (using the experimentally determined bin edges in $\sum E_{\perp}^{Pb}$). The larger values of M_A are also reflected in the η-distributions, where the effect is that the distribution becomes too flat to describe data, especially for central events.

7.1.2 Parton distribution in the projectile

As discussed in section 5, the secondary absorptive interaction in Figure I.3a may involve several partons coming from the projectile nucleon, in a way similar to how diffractive excitation is described by a Pomeron PDF in the Ingelman–Schlein model. Point (ii)

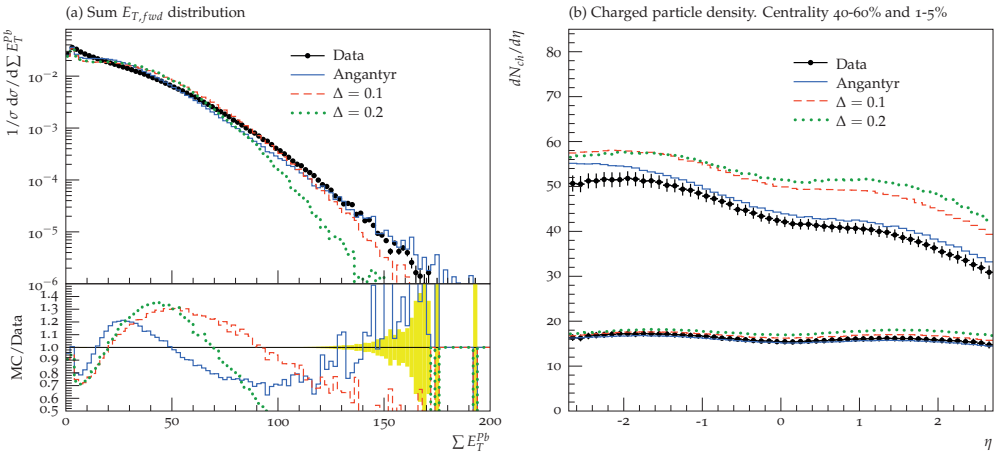


Figure I.23: Comparison between different choices of Δ for generation of secondary absorptive sub-events. Variations shown for (a) the summed transverse energy in the Pb direction ($-4.9 < \eta < -3.2$) and (b) the average charged multiplicity as a function of pseudo–rapidity for pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. Data points are from ATLAS [26]. Full blue line is the default choice of $\Delta = 0$, while the red dashed and green dotted lines corresponds to $\Delta = 0.1$ and 0.2 respectively. In (b) the lines on the bottom and top corresponds to the 40-60% and 1-5% centrality bins respectively, using the experimentally defined bin limits in $\sum E_{\perp}^{Pb}$.

concerns the distribution of these partons. In section 5 we studied three different distributions, SD(new) (which is the default for secondary absorption), SD(def) (which is the PYTHIA8 default for diffractive excitation), and SD(glu) (which is the modified PDF for increased gluon activity introduced in ref. [9]). In Figure I.24a we show the effect on the $\sum E_{\perp}^{Pb}$ distribution. Below 50 GeV, where the bulk of the events are found, all three options are reasonably close to each other, but the tail of the distributions diverges considerably, in a way consistent with the differences found in section 5. The resulting pseudo–rapidity distributions shown in Figure I.24b do not show so dramatic differences. It is, however, clear that our default choice gives the best description of data. As discussed in section 5 our default choice is the one that makes most sense on theoretical grounds, and it is satisfying to see that it also makes sense in comparison to data.

7.1.3 Energy-momentum conservation

Energy-momentum conservation is frequently seen to have a very large impact in high energy reactions. Here its effect could be seen in Figure I.23a. It is not clear from first

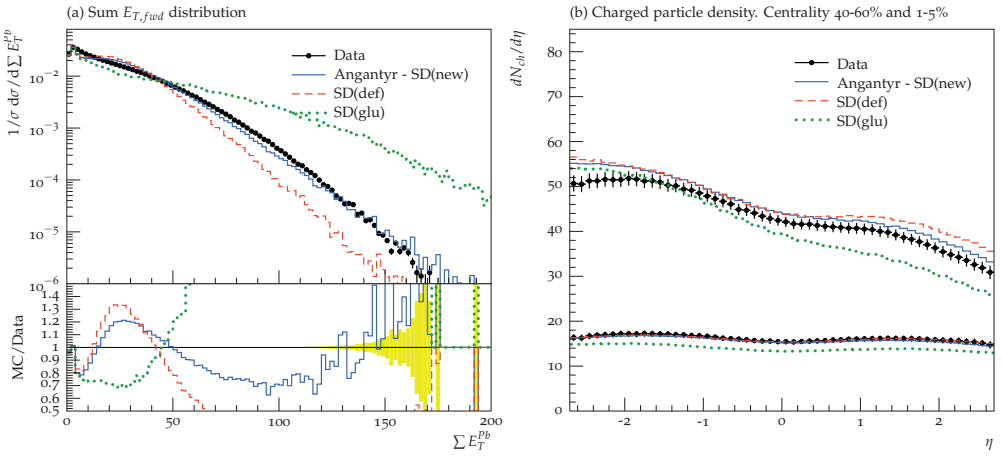


Figure I.24: Same as Figure I.23, but comparing different choices in the treatment of secondary absorptive interactions. The lines corresponds to the models in Figure I.6.

principles if energy–momentum conservation should prohibit a sub–collision, if a single sampling of the M_A distribution turns out to require more than what is available, or if it is possible to simply try again. To further study the effects of this ambiguity, we show in Figure I.25, what happens if we allow Angantyr to retry adding secondary sub-events, which fail due to energy-momentum conservation (as discussed in section 4.2). We see that it does have an impact on the most central collisions in the $\sum E_{\perp}^{Pb}$ centrality measure, while the effect on the resulting η-distribution is barely visible. It is interesting to note that the effect of allowing more attempts seem to saturate quickly, and going from 2 to 4 attempts makes a much smaller change than allowing two attempts instead of one (which is the default).

7.2 Diffractively excited nucleons

In contrast to the secondary absorbed nucleons, a positive Δ in eq. (I.23) means lower masses for diffractively excited nucleons. In principle the M_D-distribution could be measured in pp collisions at LHC, but it is quite challenging to isolate single diffraction from the experimental distribution in the size of a gap in rapidity (see refs. [70, 71]).

In collisions with nuclei, multiple NN interactions imply that, the probability for absorption is enhanced and, as a consequence, the probability for diffractive excitation is reduced. From Figure I.21 we see that in pA collisions about 20% of the wounded nucleons are diffractively excited, dropping to 10% in central pPb collisions. In AA collisions this fraction is further reduced to an average about 10%, and below 4% for central PbPb collisions.

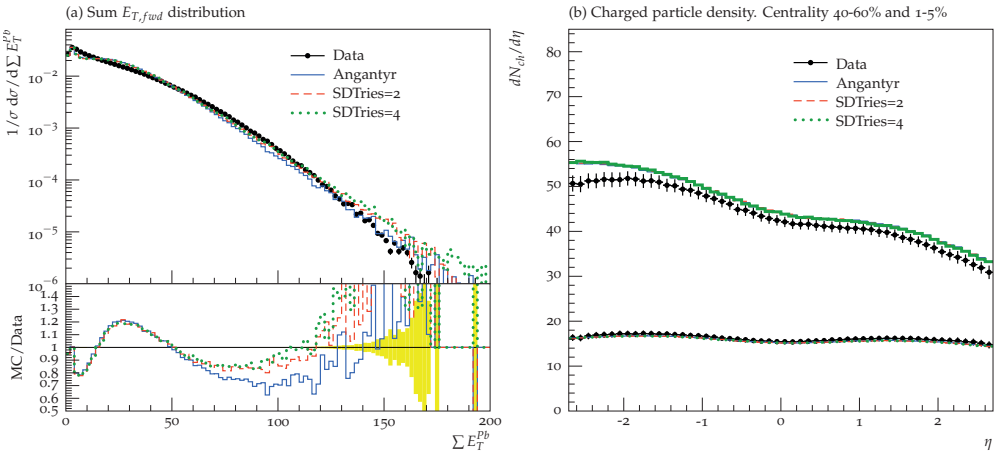


Figure I.25: Same as Figure I.23, but now varying the number of attempts (parameter `Angantyr:SDTries`) allowed to generate secondary sub-events that can be added without violating energy–momentum conservation before giving up and vetoing the secondary interaction. The default version allows only a single attempt and is shown as the blue lines, while allowing two or four attempts is shown as dashed red and dotted green lines respectively.

This implies that a reasonable variation of the diffractive component will have comparatively small effect. For this reason we have here chosen to keep the default setting in the PYTHIA8 MC, with a distribution $\propto dM_D^2/M_D^2$.

One could here also imagine including a Reggeon contribution $\propto dM_D^2/(M_D^2)^{1.5}$. This contribution is concentrated to low masses, and would not affect the results in most of the rapidity range, including the forward detectors used to measure the centrality. It could, however, give a contribution in the very forward region, and thus it might be of importance *e.g.* for interactions with cosmic rays.

7.3 Uncertainties in AA collisions

Above we have discussed model uncertainties in pA collisions. We have also pointed out that the corresponding uncertainties are significantly smaller in AA collisions, in particular in central AA collisions. In Figure I.21 we showed that the fraction of wounded nucleons which are secondary absorbed is about 70% in pPb but about 35% in PbPb collisions. For central collisions these ratios are about 80% in pPb and only about 25% in PbPb. We have checked that a corresponding reduction of the uncertainties is obtained in the MC results for AA collisions.

8 Relation to other models

As Angantyr is a new model, it is instructive to compare it to existing models, and we here discuss the most commonly used ones, also mentioned in the introduction, HIJING [6], AMPT [5], and EPOS-LHC [4]. Here HIJING is most similar to Angantyr. Like Angantyr it is constructed as an extrapolation of pp dynamics, with the explicit motivation that differences between the model and experimental results may indicate effects of collective behaviour. In contrast AMPT and EPOS are both assuming collective expansion of a thermalised medium.

The HIJING generator is built with a similar starting point as Angantyr, thus it is inspired by the Fritiof model, using PYTHIA for generating multiple hard partonic sub-collisions and the Lund string model in PYTHIA for the hadronisation. Similarly to Angantyr, HIJING relies on a Glauber calculation to determine the number of inelastic sub-collisions, which are of two types: soft nucleon-nucleon collisions treated as in Fritiof, and hard parton-parton collisions treated as in PYTHIA. A new version written in C++ was recently presented [72].

In contrast to Fritiof the interacting nucleons are in HIJING excited to higher masses, covering most of the available rapidity range, but just as in the later Fritiof version [29, 73], gluon radiation is added using the soft radiation scheme [74] implemented in Ariadne [75]. The hard partonic scatterings are determined via nucleus PDFs, where the parton density is suppressed by a shadowing factor $R_{a/A}$, compared to A independent nucleon PDFs. To avoid double counting, emitted gluons in the soft component is allowed only for $p_\perp$ below a scale p_0 (chosen to be ≈ 2 GeV), while the hard partonic collisions have a lower cut at $p_\perp = p_0$.

Another difference between Angantyr and HIJING is that in HIJING fluctuations are neglected both in the initial states of the individual nucleons and in the position of nucleons within the nuclei. The soft NN amplitude is then chosen to reproduce the inelastic cross section including diffraction. The probability for multiple scattering is determined by the nuclear overlap function in impact parameter space. In Angantyr we find that fluctuations plus the distinction between primary and secondary absorptively wounded nucleons, have a quite significant effect for the final state multiplicity. In HIJING, the same effect may partly be due to the introduction of the shadowing factor $R_{a/A}$. The shadowing factor is a geometry dependent "k-factor", which accounts for nucleons shadowing for each other during the collision, thus reducing to nucleon–nucleon cross section from the result obtained from pp collisions, to a lower, effective cross section. This suppresses the hard partonic cross section with up to 50% in AuAu collisions at $\sqrt{s_{NN}} = 200$ GeV [6]. In the end all partons are in HIJING connected by strings, and hadronised with PYTHIA. As an option it is possible to include a model for jet quenching, and also a jet trigger, enhancing

the rate for events with high-$p_\perp$ jets.

As mentioned above, AMPT presumes that a hot dense medium is formed. It uses the parton state obtained in HIJING as initial conditions. The partons then evolve in a partonic cascade up to freeze-out. After freeze-out the partons are connected in strings, which hadronise according to the Lund model in PYTHIA. Finally the obtained hadrons form a secondary cascade until the density is low enough, when they continue as free particles. As an option the hadronisation can also be calculated via quark-antiquark coalescence.

Finally the EPOS model works on different principles than the other two, as no explicit Glauber calculation is performed. Instead partonic sub-collisions are calculated using parton-based Gribov–Regge theory [76]. An elementary scattering is here represented by a cut Pomeron or "parton ladder". This ladder is interpreted as a flux tube, or a string, where the intermediate gluons provide a transverse motion. The strings then break up into segments by quark-antiquark pair production. In the central region with high density, the "core", the segments within a bin in η form a cluster, which expands longitudinally and radially until freeze-out. In regions of low density, called the "corona", the strings fragment instead directly to hadrons. This is mainly the case in the fragmentation regions. In a recent version, called EPOS LHC [4], a new flow parametrisation is introduced, which does not take advantage of the complete hydrodynamical calculation followed by the hadronic cascade as in EPOS2 [77] or EPOS3 [78]. One consequence is here that the time for one PbPb event is reduced from one hour to a few tenths of a second. According to the authors, this also implies that this version should not be used for a precise study of $p_\perp$ distributions or particle correlations in HI collisions.

In Figure I.26 we compare the multiplicity spectra at $\sqrt{s_{NN}} = 2.76$ TeV from Figure I.17(b) with Angantyr and the three generators discussed above. (For HIJING jet quenching is disabled in Figure I.26, but this should not have a major impact on the result.)

We note that with all differences mentioned above, all four generators produce quite similar results for the centrality dependence of the charged particle distribution.

Comparing first HIJING to Angantyr, we see that while Angantyr undershoots at mid-η, HIJING overshoots on the full interval, and produces a too wide shape for the distribution. The likely source of this difference is the different way of handling secondary absorptive events, as described in section 3. HIJING treats all absorptive events on a similar footing, but the nuclear shadowing included in HIJING implies that it produces an overall lower amount of hard sub-collisions.

AMPT uses HIJING for initial conditions, but compared to HIJING the overall multiplicity is reduced by the partonic and the hadronic cascades. However, although the central density agrees with data, the distribution is too wide. We also note that AMPT reproduces multiplicity at mid-η better than Angantyr, and refer to our discussion about possible re-

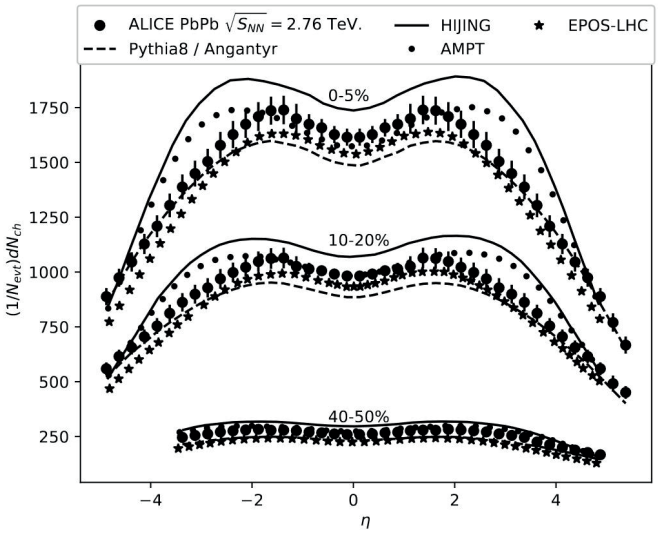


Figure I.26: Comparison of Angantyr to the generators EPOS-LHC, AMPT and HIJING. The figure shows charged particle production as function of pseudo-rapidity in PbPb at $\sqrt{s_{NN}} = 2.76$ TeV as measured by ALICE [57].

tuning of Angantyr to low-$p_\perp$ pp data in section 6.3. Finally EPOS-LHC also does a better job than Angantyr at mid-η, but worse away from the central region. We note that AMPT and EPOS-LHC, which both include the hydrodynamic expansion of a hot medium, do not describe data better than Angantyr over the full η-range.

It is, however, clear that if one wants to pin down the physics of a possible plasma phase, more exclusive observables than particle production must be used. This is indeed also the case in contemporary studies at the LHC and RHIC. Considering the precision obtained by the current tools, we see that there is a need for improved tools for comparing theory to data in heavy ion physics. To account for the final 10% discrepancy shown by all four generators, analysis specific effects like choice of centrality measure, trigger selection, primary particle definition *etc.* all play a major role. In the present paper all comparisons of Angantyr to data are carried out using the Rivet tool [52], which has proved highly successful for this task in pp. This has, however, been done using our own implementation of the experimental analyses. It is crucial for the further development of Monte Carlo event generators for heavy ion physics, that present and future heavy ion data is released using Rivet (or a similar tool), and we are pleased to note that experiments are now starting to commit to this task, also in heavy ion physics.

9 Conclusion and Outlook

We have introduced a new model called Angantyr for generating exclusive final states in proton-nucleus and nucleus-nucleus collisions. It extrapolates pp dynamics with a minimum of free parameters, and in this way it bridges the gap between heavy ion and high energy physics phenomenology. It does not assume a hot thermalised medium, and the aim is to see how well such an extrapolation can reproduce experimental data, thus exposing effects of collective behaviour. The model is a generalisation of the model for pA collisions in ref. [9], and is based on the following points:

- The basic pp interaction is described by the PYTHIA8 event generator, based on multiple partonic sub-collisions and string hadronisation.
- The generalisation to nucleus collisions is inspired by the Fritiof model, and the notion of "wounded" or "participating" nucleons.
- The number of wounded nucleons is calculated from the Glauber model in impact parameter space, including "Gribov corrections" due to diffractive excitation of individual nucleons.
- The Glauber model is formulated in impact parameter space. Diffractive excitation is then most conveniently described by the Good–Walker formalism, as the result of fluctuations in the nucleon substructure. We here for the fist time account for fluctuations in both the projectile and the target nucleons, in a Glauber calculation. (As frequently in MC simulations, fluctuations in the position of nucleons in the nuclei are also included.)

The model is implemented in an event generator, which generates exclusive final states. It is included in the PYTHIA8 package, where the user simply specifies a nucleus instead of a hadron as projectile and/or target. The possibility to add a signal process (of electroweak or other origin) is also included, enabling the user to study every process one could normally study in a pp collision.

We have shown that Angantyr gives a good description of general final state properties. This includes not only multiplicity and transverse momentum distributions both in pPb and PbPb collisions, but also its dependence on centrality. We note, however, that this dependence is very sensitive to the experimental definition of centrality. Thus we see that for low centrality the correlation between central multiplicity and "centrality" is more a correlation between central and forward activity, rather than between central activity and impact parameter. The model predictions for XeXe collisions are also in good agreement with ALICE data published later.

The model underestimates somewhat central particle production, when $p_\perp$ is integrated down to $p_\perp = 0$. This may be not surprising, as it is an extrapolation of PYTHIA's description of pp dynamics, which is too low for small $p_\perp$ below 200 MeV. Future work is needed to improve the hadronisation models in this region, including their interface to the perturbative shower.

The description of data is quite sensitive to the handling of, in particular, secondary absorptive sub-events. We have investigated several different choices relating to this treatment, relating to (i) the distributions in the covered rapidity ranges, (ii) distributions of partons in the projectile nucleon, and (iii) energy–momentum conservation. For visualization we performed this investigation in pPb collisions, noting that they will be significantly smaller in PbPb collisions. Although our final choices may not be based on completely solid theoretical grounds, the fact that alternatives investigated give a poorer description of data tells us, that the choices are reasonable. Certainly there are other variations to investigate, but we postpone such studies to a future publication.

In PYTHIA8 all strings decay into hadrons independently. Thus it does not include a mechanism to reproduce the collective effects seen in pp collisions. Such effects are therefore also not reproduced by the present version of Angantyr, and the model should be thought of as a baseline for understanding the non-collective background to observables sensitive to collective behaviour.

Also in high energy pp collisions the number of strings is quite large, in particular in events with high multiplicity. In ref. [33] we showed that overlapping strings forming "ropes" can qualitatively reproduce the increased strangeness in pp [7], as well as in pPb and PbPb [79] collisions. In ref. [34] we further showed that the transverse pressure due to the increased energy density provides a transverse expansion and a qualitative description of the "ridge" observed in pp collisions. An important future direction will be to fully include these models in Angantyr, and test to what degree they provide a description of the observed collective effects in nucleus collisions. Besides the angular correlations, the transverse expansion may affect the $p_\perp$ distributions, which are less accurately reproduced in pPb and PbPb collisions.

To conclude we think that it is notable that a direct extrapolation of pp dynamics can reproduce general features of inclusive particle production in AA collisions to better than 10%. This emphasises the importance of correlation studies, and in a future version of Angantyr we plan to include the collective effects from string-string interactions in the description of collisions with nuclei. In the future we also want to find observables sensitive to the fluctuations related to diffractive excitation and the internal substructure of nucleons. This is an essential feature which distinguishes Angantyr from other event generators available for nucleus-nucleus collisions.

Acknowledgments

CB wants to thank C. H. Christensen for assistance in comparing to ALICE data. This work was funded in part by the Swedish Research Council, contracts number 2016-03291, 2016-05996 and 2017-0034, in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No 668679, and in part by the MCnetITN3 H2020 Marie Curie Initial Training Network, contract 722104.

10 Appendix: Generating absorptively and diffractively wounded nucleons

Here we will go through the technicalities of choosing the interactions between projectile and target nucleons. In [9] we showed that for a fixed nucleon–nucleon impact parameter, b, and a fixed projectile state, the cross section for the target nucleon to be wounded is given by the average of the fluctuations in the target nucleon. Writing the imaginary part of the scattering amplitude for given projectile and target states, p and t, in terms of the corresponding S-matrix, $T_{pt}(b) \equiv 1 - S_{pt}(b)$, we have

$$d\sigma_{\mathrm{Wt}} = \left(1 - \left\langle \langle S_{pt} \rangle_t^2 \right\rangle_p\right) d^2b. \tag{I.24}$$

This works well for pA collisions, but for AA we also want to look at the probability for the projectile nucleon being wounded, and on top of this we want to be able to separate between absorptively and diffractively wounded nucleons.

10.1 Absorptively wounded nucleons

We expect the absorptively wounded nucleons will give the most important contributions to the final state particle production, and we therefore want to take special care to capture cross section fluctuations in this case and at the same time make sure we correctly reproduce the absorptive nucleon–nucleon cross section,

$$d\sigma_{\mathrm{abs}} = \left(1 - \left\langle S_{pt}^2(b) \right\rangle_{pt}\right) d^2b. \tag{I.25}$$

The procedure will therefore be to generate one state for each nucleon in the projectile and target nuclei and for each pair of nucleons calculate

$$P_{\mathrm{abs}} = 1 - S_{pt}^2(b), \tag{I.26}$$

and declare the nucleon-nucleon interaction absorptive with this probability. This will clearly give the correct absorptive nucleon–nucleon cross section.

If we find the interaction is not absorptive we want to go on and check if either the target or the projectile or both are diffractively wounded, but this will then require us to consider averages over the possible states of the projectile or target or both. In the following we will consider a diffractively wounded target, but the corresponding treatment of the projectile is completely analogous.

10.2 Diffractively wounded nucleons

In general it is not necessarily straight forward to analytically calculate the average $\langle S_{pt}^2(b)\rangle_t$ needed to get the correct cross section for diffractive excitation. Instead we will estimate the fluctuations by generating a secondary, or auxiliary state for each projectile (p') and target (t') nucleon. We will still calculate the probability for absorptive interaction using only the primary states, but to get the probability of the target nucleon to be wounded we note that the product $S_{pt}(b)S_{pt'}(b)$ will on average yield the correct value for $\langle\langle S_{pt}^2(b)\rangle_t\rangle_p$, so naively we could use the probability $P_{\rm Wt} = 1 - S_{pt}(b)S_{pt'}$. However, it is clear that we will then have a negative probability for having a diffractively wounded target $P_{\rm Wt} - P_{\rm abs} < 0$, for $S_{pt} < S_{pt'}$. Therefore we also need to consider the statistically equivalent situation where the absorptive interaction probability is given by

$$P'_{\rm abs} = 1 - S_{pt'}^2(b), \tag{I.27}$$

while the corresponding wounded probability is still

$$P'_{\rm Wt} = 1 - S_{pt}(b)S_{pt'}(b) = P_{\rm Wt}, \tag{I.28}$$

where the probability for a diffractively wounded target is then positive.

The procedure we have chosen to handle this, is to shuffle probabilities between the two situations so that we always get non-negative probabilities for diffractively wounded nucleons according to

$$\tilde{P}_{\rm Wt} = \begin{cases} S_{tp} < S_{pt'}: & 0 \\ S_{pt} > S_{pt'}: & P_{\rm Wt} + P'_{\rm Wt} - P'_{\rm abs} = 1 - 2S_{pt}(b)S_{pt'}(b) + S_{pt'}^2(b) \end{cases} \tag{I.29}$$

$$\tilde{P}'_{\rm Wt} = \begin{cases} S_{tp} < S_{pt'}: & P'_{\rm Wt} + P_{\rm Wt} - P_{\rm abs} = 1 - 2S_{pt}(b)S_{pt'}(b) + S_{pt}^2(b) \\ S_{pt} > S_{pt'}: & 0 \end{cases} \tag{I.30}$$

which will give the correct cross section for the target nucleon being wounded.

By considering the auxiliary state for the projectile, p', we can then also find the probability for the projectile being diffractively wounded. And if both are wounded we say that the interaction is a double diffractive excitation[19].

References

[1] Gleisberg, T. *et al.* "Event generation with SHERPA 1.1". *JHEP* **02,** 007. arXiv: 0811.4622 [hep-ph] (2009).

[2] Bellm, J. *et al.* "Herwig 7.0/Herwig++ 3.0 release note". *Eur. Phys. J.* **C76,** 196. arXiv: 1512.01178 [hep-ph] (2016).

[3] Sjöstrand, T. *et al.* "An Introduction to PYTHIA 8.2". *Comput. Phys. Commun.* **191,** 159–177. arXiv: 1410.3012 [hep-ph] (2015).

[4] Pierog, T., Karpenko, I., Katzy, J. M., Yatsenko, E. & Werner, K. "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider". *Phys. Rev.* **C92,** 034906. arXiv: 1306.0121 [hep-ph] (2015).

[5] Lin, Z.-W., Ko, C. M., Li, B.-A., Zhang, B. & Pal, S. "A Multi-phase transport model for relativistic heavy ion collisions". *Phys. Rev.* **C72,** 064901. arXiv: nucl-th/0411110 [nucl-th] (2005).

[6] Wang, X.-N. & Gyulassy, M. "HIJING: A Monte Carlo model for multiple jet production in pp, pA and AA collisions". *Phys. Rev.* **D44,** 3501–3516 (1991).

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

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

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

[10] Hatta, Y., Iancu, E., Marquet, C., Soyez, G. & Triantafyllopoulos, D. N. "Diffusive scaling and the high-energy limit of deep inelastic scattering in QCD at large N(c)". *Nucl. Phys.* **A773,** 95–155. arXiv: hep-ph/0601150 [hep-ph] (2006).

[11] Avsar, E., Gustafson, G. & Lönnblad, L. "Small-x dipole evolution beyond the large-N(c) imit". *JHEP* **01,** 012. arXiv: hep-ph/0610157 [hep-ph] (2007).

[19] This will actually not give the correct double diffraction cross section, but as we are mainly concerned with the number of wounded nucleons, we postpone the detailed treatment of double diffraction to future improvements

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

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

[14] Broniowski, W., Rybczynski, M. & Bozek, P. "GLISSANDO: Glauber initial-state simulation and more.." *Comput. Phys. Commun.* **180,** 69–83. arXiv: 0710.5731 [nucl-th] (2009).

[15] Rybczynski, M., Stefanek, G., Broniowski, W. & Bozek, P. "GLISSANDO 2 : GLauber Initial-State Simulation AND mOre…, ver. 2". *Comput. Phys. Commun.* **185,** 1759–1772. arXiv: 1310.5475 [nucl-th] (2014).

[16] Alvioli, M., Drescher, H. .-J. & Strikman, M. "A Monte Carlo generator of nucleon configurations in complex nuclei including Nucleon-Nucleon correlations". *Phys. Lett.* **B680,** 225–230. arXiv: 0905.2670 [nucl-th] (2009).

[17] Alvioli, M., Holopainen, H., Eskola, K. J. & Strikman, M. "Initial state anisotropies and their uncertainties in ultrarelativistic heavy-ion collisions from the Monte Carlo Glauber model". *Phys. Rev.* **C85,** 034902. arXiv: 1112.5306 [hep-ph] (2012).

[18] Glauber, R. J. "Cross-sections in deuterium at high-energies". *Phys. Rev.* **100,** 242–248 (1955).

[19] Miller, M. L., Reygers, K., Sanders, S. J. & Steinberg, P. "Glauber modeling in high energy nuclear collisions". *Ann. Rev. Nucl. Part. Sci.* **57,** 205–243. arXiv: nucl-ex/0701025 [nucl-ex] (2007).

[20] Gribov, V. "Glauber corrections and the interaction between high-energy hadrons and nuclei". *Sov. Phys. JETP* **29,** 483–487 (1969).

[21] Heiselberg, H., Baym, G., Blaettel, B., Frankfurt, L. L. & Strikman, M. "Color transparency, color opacity, and fluctuations in nuclear collisions". *Phys. Rev. Lett.* **67,** 2946–2949 (1991).

[22] Blaettel, B., Baym, G., Frankfurt, L. L., Heiselberg, H. & Strikman, M. "Hadronic cross-section fluctuations". *Phys. Rev.* **D47,** 2761–2772 (1993).

[23] Alvioli, M. & Strikman, M. "Color fluctuation effects in proton-nucleus collisions". *Phys. Lett.* **B722,** 347–354. arXiv: 1301.0728 [hep-ph] (2013).

[24] Alvioli, M., Frankfurt, L., Guzey, V. & Strikman, M. "Revealing "flickering" of the interaction strength in pA collisions at the CERN LHC". *Phys. Rev.* **C90,** 034914. arXiv: 1402.2868 [hep-ph] (2014).

[25] Alvioli, M., Cole, B. A., Frankfurt, L., Perepelitsa, D. V. & Strikman, M. "Evidence for x-dependent proton color fluctuations in pA collisions at the CERN Large Hadron Collider". *Phys. Rev.* **C93,** 011902. arXiv: 1409.7381 [hep-ph] (2016).

[26] ATLAS, Aad, G. *et al.* "Measurement of the centrality dependence of the charged-particle pseudorapidity distribution in proton–lead collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV with the ATLAS detector". *Eur. Phys. J.* **C76,** 199. arXiv: 1508.00848 [hep-ex] (2016).

[27] ALICE, Adam, J. *et al.* "Centrality dependence of particle production in p-Pb collisions at $\sqrt{s_{\rm NN}}$= 5.02 TeV". *Phys. Rev.* **C91,** 064905. arXiv: 1412.6828 [nucl-ex] (2015).

[28] Andersson, B., Gustafson, G. & Nilsson-Almqvist, B. "A Model for Low p(t) Hadronic Reactions, with Generalizations to Hadron-Nucleus and Nucleus-Nucleus Collisions". *Nucl. Phys.* **B281,** 289–309 (1987).

[29] Pi, H. "An Event generator for interactions between hadrons and nuclei: FRITIOF version 7.0". *Comput. Phys. Commun.* **71,** 173–192 (1992).

[30] Białas, A., Bleszyński, M. & Czyż, W. "Multiplicity Distributions in Nucleus-Nucleus Collisions at High-Energies". *Nucl. Phys.* **B111,** 461–476 (1976).

[31] Białas, A. & Czyż, W. "Wounded nucleon model and Deuteron-Gold collisions at RHIC". *Acta Phys. Polon.* **B36,** 905–918. arXiv: hep-ph/0410265 [hep-ph] (2005).

[32] Bierlich, C. & Christiansen, J. R. "Effects of color reconnection on hadron flavor observables". *Phys. Rev.* **D92,** 094010. arXiv: 1507.02091 [hep-ph] (2015).

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

[34] Bierlich, C., Gustafson, G. & Lönnblad, L. "Collectivity without plasma in hadronic collisions". *Phys. Lett.* **B779,** 58–63. arXiv: 1710.09725 [hep-ph] (2018).

[35] Good, M. L. & Walker, W. D. "Diffraction disssociation of beam particles". *Phys. Rev.* **120,** 1857–1860 (1960).

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

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

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

[39] Flensburg, C., Gustafson, G. & Lönnblad, L. "Inclusive and Exclusive Observables from Dipoles in High Energy Collisions". *JHEP* **08,** 103. arXiv: 1103.4321 [hep-ph] (2011).

[40] Iancu, E., Mueller, A. H. & Munier, S. "Universal behavior of QCD amplitudes at high energy from general tools of statistical physics". *Phys. Lett.* **B606,** 342–350. arXiv: hep-ph/0410018 [hep-ph] (2005).

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

[42] Donnachie, A. & Landshoff, P. V. "Total cross-sections". *Phys. Lett.* **B296,** 227–232. arXiv: hep-ph/9209205 [hep-ph] (1992).

[43] Rasmussen, C. O. & Sjöstrand, T. "Models for total, elastic and diffractive cross sections". *Eur. Phys. J.* **C78,** 461. arXiv: 1804.10373 [hep-ph] (2018).

[44] Barej, M., Bzdak, A. & Gutowski, P. "Wounded-quark emission function at the top energy available at the BNL Relativistic Heavy Ion Collider". *Phys. Rev.* **C97,** 034901. arXiv: 1712.02618 [hep-ph] (2018).

[45] CMS, Sirunyan, A. M. *et al.* "Study of Jet Quenching with Z + jet Correlations in Pb-Pb and pp Collisions at $\sqrt{s}_{NN} = 5.02\text{TeV}$". *Phys. Rev. Lett.* **119,** 082301. arXiv: 1702.01060 [nucl-ex] (2017).

[46] CMS, Sirunyan, A. M. *et al.* "Observation of top quark production in proton-nucleus collisions". *Phys. Rev. Lett.* **119,** 242001. arXiv: 1709.07411 [nucl-ex] (2017).

[47] Tarbert, C. M. *et al.* "Neutron skin of ^{208}Pb from Coherent Pion Photoproduction". *Phys. Rev. Lett.* **112,** 242502. arXiv: 1311.0168 [nucl-ex] (2014).

[48] Helenius, I., Paukkunen, H. & Eskola, K. J. "Neutron-skin effect in direct-photon and charged hadron-production in Pb+Pb collisions at the LHC". *Eur. Phys. J.* **C77,** 148. arXiv: 1606.06910 [hep-ph] (2017).

[49] Paukkunen, H. "Neutron skin and centrality classification in high-energy heavy-ion collisions at the LHC". *Phys. Lett.* **B745,** 73–78. arXiv: 1503.02448 [hep-ph] (2015).

[50] Loizides, C., Kamin, J. & d'Enterria, D. "Precision Monte Carlo Glauber predictions at present and future nuclear colliders". arXiv: 1710.07098 [nucl-ex] (2017).

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

[52] Buckley, A. *et al.* "Rivet user manual". *Comput. Phys. Commun.* **184,** 2803–2819. arXiv: 1003.0694 [hep-ph] (2013).

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

[54] ALICE, Adam, J. *et al.* "The ALICE definition of primary particles". eprint: ALICE-PUBLIC-2017-005 (June 2017).

[55] ALICE, Acharya, S. *et al.* "Centrality and pseudorapidity dependence of the charged-particle multiplicity density in Xe-Xe collisions at $\sqrt{s_{\mathrm{NN}}}$ = 5.44 TeV". arXiv: 1805.04432 [nucl-ex] (2018).

[56] ALICE, Abbas, E. *et al.* "Centrality dependence of the pseudorapidity density distribution for charged particles in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}}$ = 2.76 TeV". *Phys. Lett.* **B726,** 610–622. arXiv: 1304.0347 [nucl-ex] (2013).

[57] ALICE, Adam, J. *et al.* "Centrality evolution of the charged-particle pseudorapidity density over a broad pseudorapidity range in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV". *Phys. Lett.* **B754,** 373–385. arXiv: 1509.07299 [nucl-ex] (2016).

[58] ALICE, Adam, J. *et al.* "Centrality dependence of the pseudorapidity density distribution for charged particles in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV". *Phys. Lett.* **B772,** 567–577. arXiv: 1612.08966 [nucl-ex] (2017).

[59] ATLAS, Aad, G. *et al.* "Measurement of charged-particle spectra in Pb+Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV with the ATLAS detector at the LHC". *JHEP* **09,** 050. arXiv: 1504.04337 [hep-ex] (2015).

[60] Zhou, Y., Zhu, X., Li, P. & Song, H. "Investigation of possible hadronic flow in $\sqrt{s_{NN}} = 5.02$ TeV $p-Pb$ collisions". *Phys. Rev.* **C91,** 064908. arXiv: 1503.06986 [nucl-th] (2015).

[61] Bilandzic, A., Snellings, R. & Voloshin, S. "Flow analysis with cumulants: Direct calculations". *Phys. Rev.* **C83,** 044913. arXiv: 1010.0233 [nucl-ex] (2011).

[62] Bilandzic, A., Christensen, C. H., Gulbrandsen, K., Hansen, A. & Zhou, Y. "Generic framework for anisotropic flow analyses with multiparticle azimuthal correlations". *Phys. Rev.* **C89,** 064904. arXiv: 1312.3572 [nucl-ex] (2014).

[63] CMS, Chatrchyan, S. *et al.* "Measurement of the elliptic anisotropy of charged particles produced in PbPb collisions at $\sqrt{s}_{NN}$=2.76 TeV". *Phys. Rev.* **C87,** 014902. arXiv: 1204.1409 [nucl-ex] (2013).

[64] ALICE, Aamodt, K. *et al.* "Higher harmonic anisotropic flow measurements of charged particles in Pb-Pb collisions at $\sqrt{s_{NN}}$=2.76 TeV". *Phys. Rev. Lett.* **107,** 032301. arXiv: 1105.3865 [nucl-ex] (2011).

[65] ALICE, Adam, J. *et al.* "Anisotropic flow of charged particles in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV". *Phys. Rev. Lett.* **116,** 132302. arXiv: 1602.01119 [nucl-ex] (2016).

[66] Bierlich, C., Gustafson, G. & Lönnblad, L. "A shoving model for collectivity in hadronic collisions". arXiv: 1612.05132 [hep-ph] (2016).

[67] Gotsman, E., Levin, E. & Maor, U. "A comprehensive model of soft interactions in the LHC era". *Int. J. Mod. Phys.* **A30,** 1542005. arXiv: 1403.4531 [hep-ph] (2015).

[68] Khoze, V. A., Martin, A. D. & Ryskin, M. G. "Elastic scattering and Diffractive dissociation in the light of LHC data". *Int. J. Mod. Phys.* **A30,** 1542004. arXiv: 1402.2778 [hep-ph] (2015).

[69] Ostapchenko, S. "Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: I. QGSJET-II model". *Phys. Rev.* **D83,** 014018. arXiv: 1010.1869 [hep-ph] (2011).

[70] ATLAS, Aad, G. *et al.* "Rapidity gap cross sections measured with the ATLAS detector in *pp* collisions at $\sqrt{s} = 7$ TeV". *Eur. Phys. J.* **C72,** 1926. arXiv: 1201.2808 [hep-ex] (2012).

[71] CMS, Khachatryan, V. *et al.* "Measurement of diffraction dissociation cross sections in pp collisions at $\sqrt{s}$ = 7 TeV". *Phys. Rev.* **D92,** 012003. arXiv: 1503.08689 [hep-ex] (2015).

[72] Barnaföldi, G. G. *et al.* "First Results with HIJING++ in High-Energy Heavy-Ion Collisions". *Nucl. Part. Phys. Proc.* **289-290,** 373–376. arXiv: 1701.08496 [hep-ph] (2017).

[73] Andersson, B., Gustafson, G. & Pi, H. *FRITIOF extended* in *Perturbative QCD and hadronic interactions. Proceedings, Hadronic Session of the 27th Rencontres de Moriond, Les Arcs, France, March 22-28, 1992* (1992), 429–432.

[74] Andersson, B., Gustafson, G., Lönnblad, L. & Pettersson, U. "Coherence Effects in Deep Inelastic Scattering". *Z. Phys.* **C43,** 625 (1989).

[75] Lönnblad, L. "ARIADNE version 4: A Program for simulation of QCD cascades implementing the color dipole model". *Comput. Phys. Commun.* **71,** 15–31 (1992).

[76] Drescher, H., Hladik, M., Ostapchenko, S., Pierog, T. & Werner, K. "Parton based Gribov-Regge theory". *Phys. Rept.* **350,** 93–289. arXiv: hep-ph/0007198 (2001).

[77] Werner, K., Karpenko, I., Pierog, T., Bleicher, M. & Mikhailov, K. "Event-by-Event Simulation of the Three-Dimensional Hydrodynamic Evolution from Flux Tube Initial Conditions in Ultrarelativistic Heavy Ion Collisions". *Phys. Rev.* **C82,** 044904. arXiv: 1004.0805 [nucl-th] (2010).

[78] Werner, K., Guiot, B., Karpenko, I. & Pierog, T. "Analysing radial flow features in p-Pb and p-p collisions at several TeV by studying identified particle production in EPOS3". *Phys. Rev.* **C89,** 064903. arXiv: 1312.1233 [nucl-th] (2014).

[79] Bierlich, C. "Rope Hadronization and Strange Particle Production". *EPJ Web Conf.* **171,** 14003. arXiv: 1710.04464 [nucl-th] (2018).

Paper II

II

A Spatially Constrained QCD Colour Reconnection in pp, pA, and AA Collisions in the PYTHIA8/Angantyr Model

Leif Lönnblad and Harsh Shah

Eur. Phys. J. C 83 (2023) 7, 575
DOI: 10.1140/epjc/s10052-023-11778-3, 10.1140/epjc/s10052-023-11816-0 (erratum)
e-print: arXiv:2303.11747 [hep-ph]
LU-TP 23-04, MCNET-23-06

Abstract: We present an updated version of the QCD-based colour reconnection model in PYTHIA, where we constrain the range in impact parameter for which reconnections are allowed. In this way, we can introduce more realistic colour reconnections in the Angantyr model for heavy ion collisions, where previously only reconnections within separate nucleon sub-collisions have been allowed. We investigate how the new impact parameter constraint influences final states in pp collisions, and retune parameters of the multi-parton interaction parameters in PYTHIA to compensate so that minimum bias data are reproduced. We also study multiplicity distributions in pA collisions and find that, in order to counteract the loss in multiplicity due to the introduction of global colour reconnections, we need to modify some parameters in the Angantyr model while keeping the parameters tuned to pp fixed. With Angantyr we can then extrapolate to AA collisions without further parameter tuning and retaining a reasonable description of the basic multiplicity distributions.

1 Introduction

The field of heavy-ion (HI) collisions is widely studied under assumption of the creation of a thermalised medium of strongly coupled partons; the Quark-Gluon-Plasma (QGP). Observables showing, *e.g.*, *strangeness enhancement*, long range *collectivity*, *quarkonia suppression*, and *jet quenching* in HI collisions are conventionally assumed to reflect the formation of such a QGP [1]. Two of these observables, namely *strangeness enhancement* [2], and long range *collectivity* [3], are, however, observed also in pp collisions, where the QGP formation is conventionally not assumed to be present. This has cast doubts on our understanding of pp collisions as well as on our understanding of these observables in HI collisions.

Pythia [4, 5] is a well known and widely used event generator for small collision systems such as $\mathrm{e^+e^-}$ and pp. In [6] we developed the Angantyr model to allow the use of Pythia's excellent description of pp collisions also for HI, by introducing a sophisticated stacking of multiple pp-like collisions to build up complete HI collision events. In this way, we have constructed a test bench for developing models for collective effects that can explain the behaviour of the above-mentioned observables without introducing a QGP.

Angantyr uses an advanced Glauber model [7, 8], where the Good–Walker picture [9] of diffraction provides a description of fluctuations (often referred to as Glauber–Gribov colour fluctuations) that influences both the number of nucleon-nucleon (NN) sub-collisions and the type of each individual sub-collision in a heavy-ion event. Each of these sub-collisions is then simulated using the standard Pythia minimum-bias machinery, producing non-diffractive, elastic, single and double diffractive sub-events according to the type determined in the Glauber simulation. The resulting sub-events are then simply stacked together into a full HI event.

In Angantyr there is a special treatment of situations where one nucleon collides non-diffractively with several others. In this situation only one such sub-collision is considered *primary* and is modelled by a full non-diffractive event in Pythia. The other, *secondary*, pp collisions are treated as diffractive excitations of the additional nucleon and modelled in Pythia using the standard Pomeron-based single diffractive model. In such a secondary non-diffractive (SND) sub-collision there is a special treatment of the Pomeron parton densities, to better mimic the particle production of a non-diffractive pp event in the direction of the excited nucleon.

The Angantyr model contains a number of new parameters, however most of these are tuned using pp observables, except the ones controlling the SND sub-events, which are tuned to pA observables. It is an important feature of the Angantyr model that there is then no further parameters to tune when generating AA collision events. The Angantyr model is nevertheless able to reproduce general features such as multiplicity distributions

in different centrality bins for, *e.g.*, PbPb and XeXe collisions at the LHC.

Currently, although the sub-events are generated on parton-level and stacked together to be hadronized together using PYTHIA string fragmentation, the colour dipoles that build up the string from different sub-collisions do not interact with each other in the Angantyr model. Therefore in the Angantyr all sub-collisions hadronize separately in HI events. Figure II.1 outlines the general scheme of the existing HI events simulation in the left part under PYTHIA/Angantyr (Default). This is, of course, a simplification and we don't believe that there is no cross-talk at all between the sub-events.

This work is aimed to further develop the Angantyr model to have interactions among partons produced in different sub-collisions in HI events. Recently two models based on *string interaction* have been investigated by the Lund group. One is the shoving model [10, 11], where the overlapping fields of nearby strings give a repulsive force that gives rise to an azimuthal flow. The other is the rope hadronization model [12, 13] where the increased tension in overlapping strings gives rise to strangeness enhancement. In addition there is now a model for hadron rescattering in PYTHIA [14, 15], that also works for HI collisions.

Here we will instead focus on how the strings are formed from the coloured partons produced in the scattering and after initial- and final-state parton showers. This is typically done using the $N_c \to \infty$ limit where any coloured parton is uniquely coupled to an anti-coloured one in a *dipole*. Gluons carry both colour and anti-colour, so we will get a set of dipoles connected together with gluons that form strings.

PYTHIA treats multi-parton interactions (MPIs) [16] as independent partonic interactions, where the initial- and final-state parton showers are also independent vacuum radiations. Before the produced strings are allowed to hadronize, however, it was early on clear that the strings in these different parton interactions needed to undergo a *colour reconnection* (CR) procedure in order to describe pp data.

Colour reconnections is the only step in PYTHIA where the produced partons from different sub-scatterings interact with each other before hadronization. In that spirit, we will here look at the effects of allowing CR to work also on partons from different NN sub-collisions in HI collisions as illustrated on the right side of Figure II.1 under PYTHIA/Angantyr (new).

In this work we use the QCD-CR model [17], which is different from the MPI-based CR model [16], which is the default CR model in PYTHIA. We give a short overview of the two CR models in section 2. We introduce a new parameter to determine the allowed spatial transverse separation between colour dipoles to be colour reconnected. This new parameter plays a key role in enabling CR among partons from different sub-collisions. We describe its importance and expected effects on the event final states in more detail in section 2.1.

The rest of this paper is organised as follows. In section 3 we discuss the re-tuning of

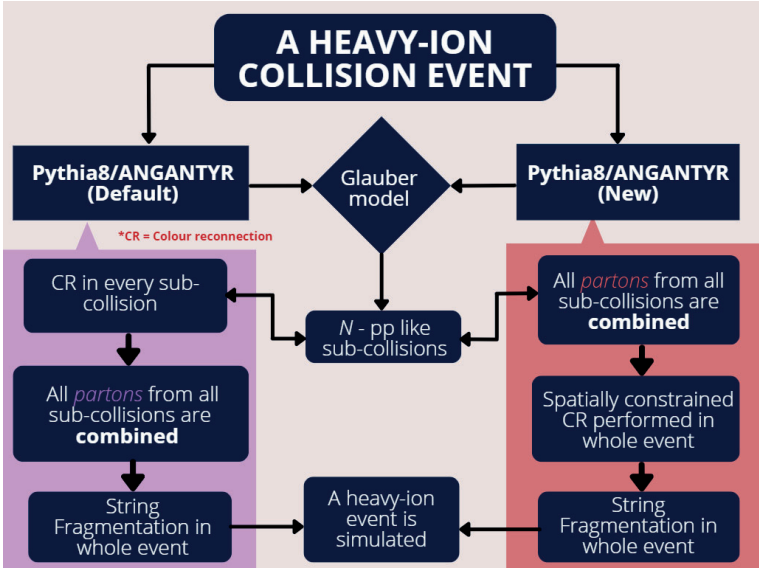


Figure II.1: Comparison of the new implementation against the default structure of the event simulation in the Angantyr model. So far we were treating a heavy-ion collision event as a sophisticated superposition of multiple pp like collisions stacked together at the parton level after the colour reconnection. In this work, we are treating a heavy-ion event as one event as early as possible, which is from the colour reconnection state onward.

selected parameters needed for the reproduction of pp collision data to remain intact when we introduce the transverse separation cut. There we also describe the subsequent re-tuning of SND parameters to reproduce multiplicity distributions in pPb data. We also introduce some modifications to the fragmentation of so-called junction strings in PYTHIA (explained in more detail in appendix 6), which were necessary to allow QCD-CR in HI collisions. In section 4 we present some outcomes of the new model in pp, pA, and AA, before we present our conclusions in section 5 together with an outlook.

2 The Colour Reconnection

In hadronic collisions there are coloured particles in both the initial and final state, and it is reasonable to assume that there will be multiple parton scattering in a single collision. Assuming that all scatterings are completely independent of each other, and also contribute equally to the momentum distribution and multiplicity of final state hadrons, one would expect that the multiplicity would grow with the number of scatterings, while observables

such as the average transverse momentum, $\langle p_\perp \rangle$, would be almost constant.

The fact that already the ISR [18] and UA1 [19] experiments found that $\langle p_\perp \rangle$ actually increases with the number of charged particles, N_{ch}, tells us that this picture of MPIs is too naive, and indicates some sort of collective behaviour in the hadronization stage, that correlates partons from different scatterings. In particular it would indicate that additional scatterings would contribute to the average transverse momentum but not so much to the multiplicity.

In the original MPI model [16] this was handled with a colour rearrangement. A single partonic scattering would produce colour connections between scattered partons and the hadron remnants, resulting in long strings that produce many (soft) hadrons. A secondary scattering would naively do the same, but the rearrangement allows the secondary scattered partons to instead be connected to the previous scatterings. This reduces the number of additional soft hadrons produced in additional scatterings. The scatterings will, however, still contribute to the average transverse momentum, giving a rise in $\langle p_\perp \rangle$ with multiplicity.

Later the effects of colour rearrangements were also studied in particle event generators in a series of papers [20–24] to investigate possible CR in $\mathrm{e^+e^-}$ annihilation at LEP. One of the problems was understanding the uncertainty in the W mass as measured in e^+e^- $\rightarrow W^+W^- \rightarrow q_1\bar{q_1}'q_2\bar{q_2}'$ events. The naive expectation is here that the $q\bar{q}$ from each W-decay are colour connected separately. However, it is possible to have an alternative configuration where quarks and anti-quarks originating from different W bosons are colour connected as the final colour configuration. The difference in colour connections will influence the jet shapes, and hence also the experimentally reconstructed W masses. The probability for such a rearrangement of the colour configuration is given by $1/\mathrm{N_c^2} = 1/9$. It is referred as *colour reconnection* in the event generators. Today, in event generators such as Pythia, CR is a generic name given to algorithms which decide a colour configuration to be used to colour connect the partons. The probability for CR in lepton collisions is further reduced by the limited space-time overlap between the produced parton systems from the W bosons decays in the case of the above example, as described in [21].

Due to the non-perturbative nature and limited understanding of the colour configuration at the parton level, there is some liberty in developing CR models. The common approach in all of them is minimizing the rapidity span of the produced hadrons from the strings. For more details, we refer to [16, 17] and references therein.

The possibility of a CR model based on $SU(3)$ colour algebra with a finite number of colours was first proposed in [20] for the W pair production and their purely hadronic decays in $\mathrm{e^+e^-}$ collisions. The QCD-CR model [17] is an extension of the default CR in Pythia introducing $SU(3)$ colour algebra in pp collisions. In addition to the standard reconnection of colour lines (sometimes referred to as a *swing*, see Figure II.2a), this model also includes the possibility for *junction* formation, where three string pieces are

connected to a single point. Junctions can be formed by two dipoles reconnecting into a junction–anti-junction pair connected by a new colour line as in Figure II.2b, or by three dipoles reconnecting to separate junction—anti-junction systems as in Figure II.2c. As the junctions carry baryon number, the QCD-CR model introduces a new baryon production mechanism, in addition to the normal formation of baryons in the string fragmentation through diquark production.

Before we discuss junctions and other possibilities for colour reconnection under QCD-CR, let us see what makes it beyond the leading colour (LC). Referring to the arguments in[17], consider a scenario when two gluons are extracted from a proton, here QCD gives several possibilities for the colour multiplets formed by these two gluons. Each of the gluons can have 8 possible colours, and from the colour algebra, these two gluons are in one of the colour multiplets,

$$8 \otimes 8 = 27 \oplus 10 \oplus \bar{10} \oplus 8 \oplus 8 \oplus 1. \tag{II.1}$$

The 27 (or a "viginti-septet") represents LC, where the two gluons extend four independent strings from the proton. For random gluon colours, this has the probability

$$P_{LC} = \frac{27}{64} \approx 0.4, \tag{II.2}$$

which is less than 50%. Hence it is evident that sub-leading colour topology has non-negligible effects. We note that the least probable colour configuration is 1 (singlet), where both gluons have exactly opposite colours, and it can occur with a probability of $1/64$. Similarly, other possible cases of two quarks, a quark and a gluon, or a quark and an anti-quark are shown in [17]. The conclusion is that sub-leading multiplets are ignored in a LC model, and for hadron collisions, sub-leading multiplets have a significant contribution. The colour algebra becomes more and more complex for cases where multiple partons are extracted from the beam particle.

The QCD-CR starts with LC ($N_c \to \infty$) connections after the parton showers and assigns $SU(3)$ weighted colours to the partons. The above-mentioned $SU(3)$ colour rules are applied only on the uncorrelated partons. The assumption of uncorrelated partons, and assigning them $SU(3)$ weighted colour compositions allows us to have an approximation of the colour configuration the partons may have. Once all the partons are assigned colours (one colour for a quark, and two colour indices for a gluon), the colour reconnection is performed based on the fundamental idea of minimising the so-called λ-measure [25]. The λ-measure gives an estimate of the number of hadrons produced in the string breakup.

Three different initial colour topologies and respectively allowed reconnections are shown in figure II.2, for the colour dipoles between a quark and an anti-quark. There is one more configuration for dipoles with a long gluon chain, which are colour reconnected as *zipper-style* junctions, but we are not going into details about such complex configurations here.

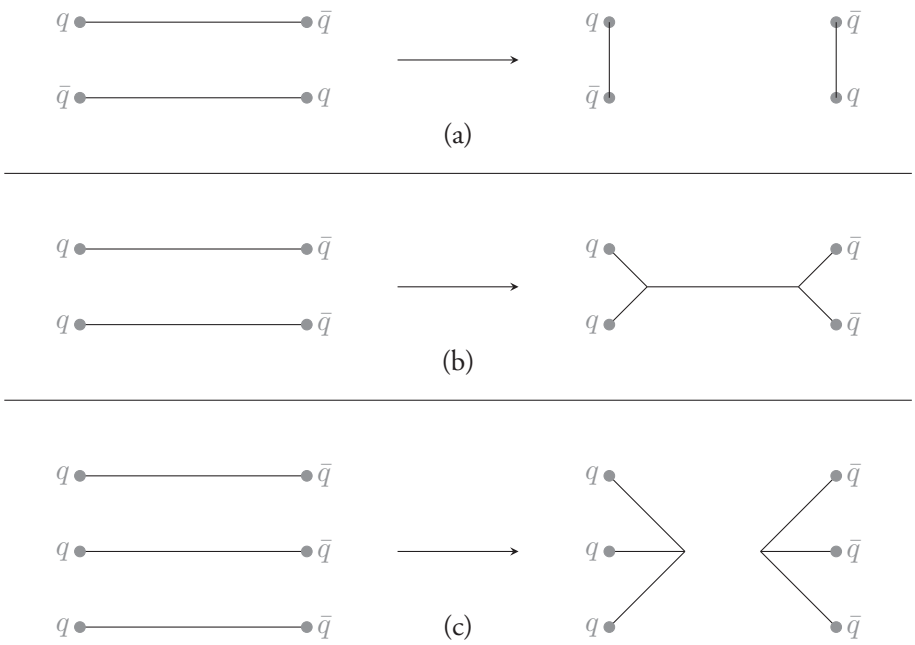


Figure II.2: Two dipoles and three dipoles CR possibilities. For two dipoles, they can either have (a) a simple reconnection (a.k.a. a *swing*) or (b) a formation of a connected junction and anti-junction system. Three dipoles can form (c) disconnected junction and anti-junction systems.

For two colour dipoles, there are two possible reconnection topologies:

(a) Ordinary style (Colour dipole swing): In this case, both dipoles exchange their colour connected partons. The QCD colour algebra will control the reconnection probability in addition to the λ measure for the new configuration. In this particular case, the reconnection probability from the QCD constraints is $1/9$, since both dipoles have to have the same colour configuration.

(b) Junction style: Instead of the reconnection of the dipole endpoints, a new string piece is created connecting the two quarks to one end of the string piece, and two anti-quarks to the other end. This configuration creates a junction and an anti-junction connected by a string piece. The QCD probability is here $1/3$, which is higher than in the previous case. However, the potential reduction in the λ-measure in this type of configuration is smaller, due to the creation of a new string piece. Hence, the algorithm suppresses such a configuration.

The QCD-CR model has also a possibility for reconnection for three colour dipoles:

(c) Junction style (Three dipoles): Two independent string systems with a junction and an anti-junction are formed after the reconnection. In this case, the QCD probability is $1/27$.

Similarly, colour reconnections are performed for dipoles containing $q-g, g-g$, and $\bar{q}-g$. For all types of colour dipoles, the reconnection probability is controlled by the $SU(3)$ colour algebra and the λ-measure.

In the model implementation, some simplifications are made. All dipoles are assigned colour indices from 1 to 9. For the ordinary swing reconnection to be allowed, two dipoles have to have the same index, which will provide a $1/9$ probability. For the case of junction style reconnection between two dipoles, the constraint is that the dipole indices have to be different but their value modulo three has to be the same, giving the probability $2/9$. Similarly, for the three dipoles case the junction style CR will require all three dipoles to have different indices, but the same value for the modulo three of the dipole indices, giving the probability $2/81$.

Clearly the junction formation reconnections are here suppressed compared to the pure colour algebra ($1/3 \to 2/9$ and $1/27 \to 2/81$ respectively), and to compensate for this a special parameter C_j is used to decrease the λ-measure for junction systems in the model to favour junction formation over the swing reconnection.

The above constraints will only decide if a certain colour configuration will be allowed or not. To determine if the allowed configuration is preferred or not, the model calculates the λ-measure. The model only allows the reconnection between the two or three dipoles if the λ-measure of the new configuration is lower than the original one.

2.1 Spatially constrained model

In the QCD-CR model, all colour dipoles are allowed to undergo CR in pp collisions in Pythia. For any two (or three) dipoles, whether they will be colour reconnected or not, is primarily decided based on whether the colour indices match and whether or not the colour reconnection will reduce the overall λ-measure.

Our aim is to treat super-positioned pp like events as a single HI event as early as possible in the Angantyr model. One way is to perform CR on all the colour dipoles from all the sub-collisions before the hadronization stage in the Angantyr model. But the spatial span of a HI collision can be as large as the diameters of the two colliding nuclei. The strong force is a short range force, and its range is approximately the size of a proton. Therefore, it is essential to introduce an additional spatial constraint between the colour dipoles to be colour reconnected.

The primary assumption in this work is that the majority of the colour dipoles are more or less parallel to the beam axis, and they are separated in the transverse plane. Hence we use the transverse positions of the partons to determine the separation between any two colour dipoles. For a more realistic constraint, one must take into account the full 3+1-dimensional space-time coordinates. Here we instead make a simplified assumption that the position of the dipole in the impact parameter can be represented by the midpoint between the two partons. If the distance between two such dipole's midpoints is larger than some parameter, δb_c, they should not be allowed to reconnect.

When we apply the spatial constraint in the QCD-CR model in pp collisions, the direct consequence of the constraint will be on the multiplicity distribution: fewer dipoles are now allowed to undergo CR compared to the default setup, the total multiplicity distribution will increase. But also other observables will be affected, such as $\langle p_\perp \rangle (N_{ch})$ and the $p_\perp$ distribution. Before applying the spatial constraint to HI collisions, we will therefore need to re-tune some of the parameters of the MPI and QCD-CR models to retain a good description of pp data.

In HI collisions, the spatial constraint will allow CR among the nearby colour dipoles independent of their original sub-collisions. This will increase CR in a HI event, especially in *central* collisions, and thus reduce the multiplicity. To counteract this we will need to retune some of the parameters in the Angantyr model.

In section 3, we will discuss the selection of the parameters we will retune and examine how they affect relevant observables.

2.2 Improved junction handling

Compared to the default reconnection model in PYTHIA, the QCD-CR results in an increase in simulation time. This could be expected due to the increase in complexity of the algorithm. We noticed, however, that the primary cause of the increase was the high failure rate in the hadronization of the junction systems. Many junction systems do not hadronize properly at the first attempt, and the algorithm has to repeat the process multiple times to succeed.

Moreover, when running the QCD-CR for HI collisions we noticed that a substantial number of events were thrown away because PYTHIA was not able to hadronize certain complicated junction systems. Normally such errors are unimportant, but since these errors were more frequent in high multiplicity events, they caused an artificial skewing of the overall multiplicity distribution. This is noticeable already in pp and became quite significant in AA. For this reason, we decided to improve the handling of the junction hadronization in PYTHIA. The changes we made are mainly technical improvements and will be included

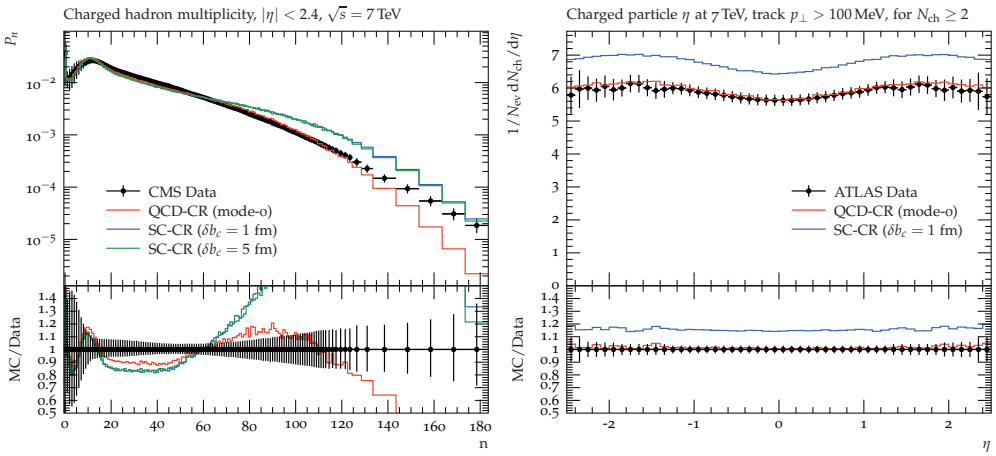


Figure II.3: The effect of improved junction hadronization in PYTHIA before retuning. Events are generated for $\sqrt{s} = 7$ TeV pp non-single-diffractive collisions and compared with CMS [26] and minimum bias collisions with ATLAS [27] data. *Left*: multiplicity in $|\eta| < 2.4$ with CMS data. *Right*: multiplicity distribution in η for $|\eta| < 2.4$ with ATLAS data. The red line shows the results for default QCD-CR (Mode-0) in PYTHIA, and the blue and green lines represent the results for SC-CR with improved junction hadronization and added spatial constraint set to large values: blue lines $\delta b_c = 1$ tm and green line $\delta b_c = 5$ fm.

in a future PYTHIA release. For completeness, we include the details of these changes in appendix 6.

The direct consequence of our modifications in the junction hadronization can be seen in pp collisions as an enhancement of high multiplicity events. In figure II.3 we show example multiplicity distributions in CMS and ATLAS minimum bias events, compared to the result from the QCD-CR(mode-0) predictions and the results of our modifications (labelled SC-CR for *Spatially Constrained* CR), but with an allowed dipole separation so large that only the effects of the modified junction hadronization are included. In the leftmost figure, we show two values of the allowed dipoles separation, $\delta b_c = 1$ and 5 fm, to show that already 1 fm is large enough to remove the effect of the spatial constraints.

Clearly, the QCD-CR needs to be returned when introducing the improved junction hadronization, but in the following, we will also want to tune the value of the allowed dipole separation in light of using the SC-CR model also for heavy ion collisions.

3 Selection of parameters and retuning strategy

There are many parameters within and outside the QCD-CR model, which can be re-tuned. Table II.1 shows the set of the parameters re-tuned when the QCD-CR model is introduced in PYTHIA.

Table II.1: The list of parameters and their values in the Monash tune, and in the QCD-CR (Mode-0) tune in PYTHIA.

Parameters	Monash	QCD-CR (Mode-0)
`StringPT:sigma`	0.335	0.335
`StringZ:aLund`	0.68	0.36
`StringZ:bLund`	0.98	0.56
`StringFlav:probQQtoQ`	0.081	0.078
`StringFlav:ProbStoUD`	0.217	0.2
`StringFlav:probQQ1toQQ0join`	0.5,0.7,0.9,1.0	0.0275,0.0275,0.0275,0.0275
`BeamRemnants:remnantMode`	0	1
`BeamRemnants:saturation`	-	5
`MultiPartonInteractions:pT0Ref` ($p_{\perp 0}^{\mathrm{ref}}$)	2.28	2.12
`ColourReconnection:mode`	0	1
`ColourReconnection:allowDoubleJunRem`	-	off
`ColourReconnection:m0` (m_0)	-	2.9
`ColourReconnection:allowJunctions`	-	on
`ColourReconnection:junctionCorrection` (C_j)	-	1.43
`ColourReconnection:timeDilationMode`	-	0

The parameters in the first part of Table II.1 are tuned outside of the QCD-CR model. Those parameters directly affect the flavour production under string breaking during the fragmentation stage in the hadronization framework. The primary reason for re-tuning those parameters was to adjust the flavour production under the new colour reconnection treatment. We decided not to modify the parameters associated with the string breaking and quark-antiquark pair creation with the parameters in this analysis, and instead mainly focus on the parameters governing the overall multiplicity.

We selected the following parameters to retune against pp data:

1. $p_{\perp 0}^{\mathrm{ref}}$, low-$p_\perp$ suppression for MPIs,
2. m_0, a scale parameter used in the λ-measure and a mass cut-off for pseudo-particles (see section 3.2),
3. C_j, a parameter reducing the λ-measure for junction systems,

4. δb_c, the allowed dipole separation.

The last one of these is the new parameter we introduce in PYTHIA[1] which gives maximum transverse distance between centres of a dipole pair to be considered for reconnection (in femtometres). δb_c also restricts the junction formation with three dipoles. If any of the three dipole pairs is separated with a transverse distance larger than δb_c, then those three dipoles are not allowed to form junctions.

The primary strategy for our retuning is to first modify these four parameters, and after an acceptable tune for pp observables has been obtained we turn to pA and try to adjust parameters in the Angantyr model to also get an acceptable description there. The main feature of the Angantyr model affecting the overall behaviour of the multiplicity is the treatment of secondary non-diffractive (SND) sub-collisions.

The SND interactions are introduced in the Angantyr model to treat situations, where a nucleon is tagged as a participant in multiple non-diffractive type collisions with other nucleons. While primary collisions are generated as normal non-diffractive pp events, these secondary ones are generated as a diffractive excitation of the additional nucleon. This is the main feature that allows Angantyr to reproduce general features of the final state multiplicity in pA and AA events. In [6] we modified the Pomeron parton distribution functions (PDFs) for SND events. In this work, we have decided to rather modify the Pomeron flux in SND events, and modify the so-called ϵ_{pom} parameter. Modification of this parameter is aimed to modify only the SND interactions. Since SND interactions do not occur in pp collisions, we can not tune the ϵ_{pom} parameter there.

After obtaining a reasonable fit to pA data, we can use the obtained tune to generate AA events, and at this stage, there are no further parameters to tune.

In the following we will study the effects of the selected parameters individually. We will restrict our study to typical minimum bias observables such as charged multiplicity distributions and $\langle p_\perp \rangle (N_{ch})$.

3.1 Low-$p_\perp$ suppression for MPIs

For minimum bias events, PYTHIA regularise the QCD $2 \to 2$ process by introducing a collision energy dependent parameter $p_{\perp 0}$ to regularise the divergences in the partonic cross section

$$\frac{d\sigma_{2\to 2}}{dp_\perp^2} \to \frac{d\sigma_{2\to 2}}{dp_\perp^2} \times \frac{p_\perp^4}{(p_\perp^2 + p_{\perp 0}^2)^2} \frac{\alpha_s^2(p_\perp^2 + p_{\perp 0}^2)}{\alpha_s^2(p_\perp^2)} \tag{II.3}$$

[1] In PYTHIA this parameter is now called `ColourReconnection:dipoleMaxDist`.

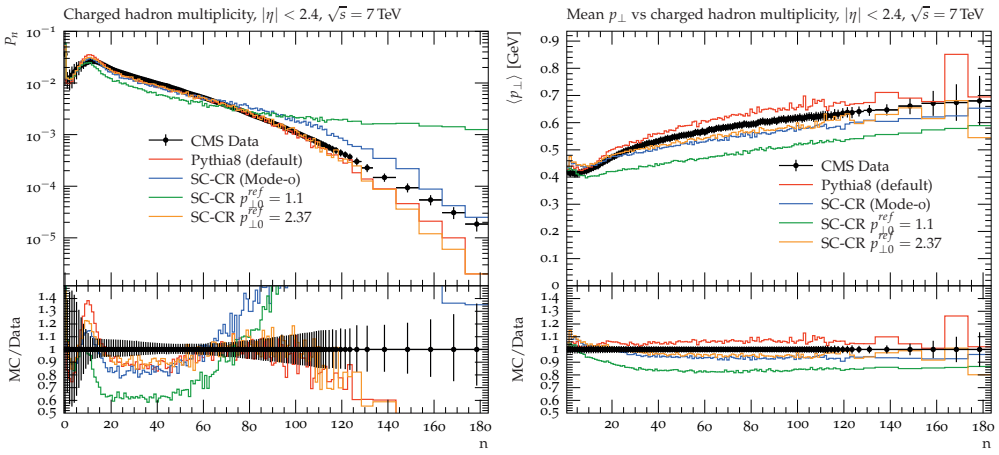


Figure II.4: The effect of varying the $p_{\perp 0}^{\mathrm{ref}}$ parameter in PYTHIA. Events are generated for $\sqrt{s} = 7$ TeV pp non-single-diffractive collisions and compared with CMS data [26]. *Left*: multiplicity in $|\eta| < 2.4$. *Right*: $\langle p_\perp \rangle$ vs N_{ch} for $|\eta| < 2.4$. The red line shows the results for default PYTHIA, the blue line the results for SC-CR (Mode-0), while the green and orange lines show the effects of varying $p_{\perp 0}^{\mathrm{ref}}$ in the latter.

Here α_s is the strong coupling constant, and $p_\perp$ is the transverse momentum of the scattered partons. The energy dependence of the parameter $p_{\perp 0}$ is further given by another parameter $p_{\perp 0}^{\mathrm{ref}}$. The power law dependent relation is given as

$$p_{\perp 0} = p_{\perp 0}(E_{\mathrm{cm}}) = p_{\perp 0}^{\mathrm{ref}} \times \left(\frac{E_{\mathrm{cm}}}{E_{\mathrm{cm}}^{\mathrm{ref}}} \right)^{p_{E_{\mathrm{cm}}}}, \tag{II.4}$$

where $p_{E_{\mathrm{cm}}}$ is a scaling parameter, controlling the growth of $p_{\perp 0}$ with the centre of mass energy of the collision, E_{cm}, with respect to a reference energy, $E_{\mathrm{cm}}^{\mathrm{ref}}$, which by default is set to 7 TeV in PYTHIA.

The effects of varying $p_{\perp 0}^{\mathrm{ref}}$ on the event multiplicity and $\langle p_\perp \rangle (N_{ch})$ distribution is shown in Figure II.4. Here (and also in Figures II.6 and II.7 below) we compare result for default PYTHIA with the results for the SC-CR (with all parameters set as in Mode-0 in Table II.1 and with $\delta b_c = 1$ fm). The main effect of reducing $p_{\perp 0}^{\mathrm{ref}}$ is an increase of multiple scatterings which increases the multiplicity and reduces the average transverse momenta, which is indeed what is shown in the figure.

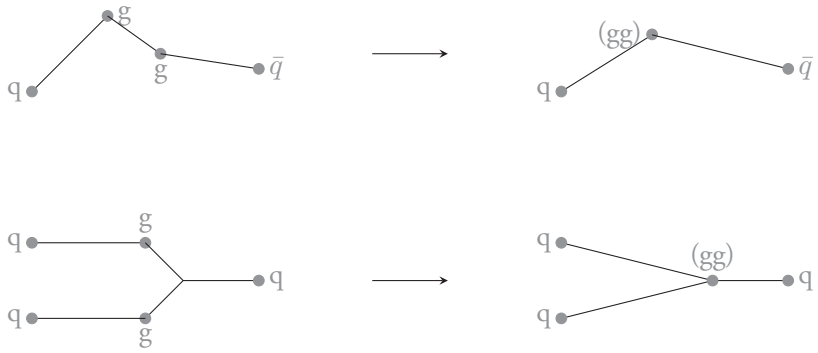


Figure II.5: Top) Pseudo-particle is formed from a dipole that has a smaller invariant mass than m_0 in a string, and Bottom) Pseudo-particles are formed if the dipole is connected to a junction.

3.2 m_0 and C_j parameters

The parameter m_0 is the mass scale in the λ-measure [25] used in the QCD-CR model. C_j is a parameter which modifies this mass scale in string pieces connected to a junction, according to $m_{0j} = C_j m_0$.

In the QCD-CR model, there is a special treatment of small-mass string pieces. Any dipole with invariant mass less than the m_0 scale will not be allowed to reconnect but is instead collapsed into a pseudo-particle.

Figure II.5 shows how one short dipole or two short dipoles are replaced by a *pseudo-particle* in a normal dipole chain (top panel), and in a junction system (bottom panel). The invariant mass of every dipole is compared with m_0, and if it's smaller than m_0 then the dipole is replaced with a *pseudo-particle*, which has the four-momentum of the dipole. Dipoles on both ends of the short dipole are connected to the *pseudo-particle*. This way the value of m_0 will directly control the number of dipoles that undergoes CR in the QCD-CR model. These small dipoles are only removed during the CR, but they do contribute to the hadron production. Therefore the parameter m_0 significantly affects the hadron multiplicity in the event final state. The removal of *small* dipoles is due to technical reasons, and to reduce the complexity of the CR process [17]. It is suggested that in the QCD-CR [17] model the parameter m_0 having its value around Λ_{QCD} has negligible effect, but raising its value beyond 1 GeV significantly reduces the amount of the colour reconnection.

The λ-measure used in [25] is an infrared safe measure of partonic final states, approximately proportional to the the resulting hadronic multiplicity. In the QCD-CR, an approximation is used, which is good for large energies. For string pieces between gluons

and/or massless quarks, it is given by

$$\lambda = ln\left(1 + \frac{\sqrt{2}E_1}{m_0}\right) + ln\left(1 + \frac{\sqrt{2}E_2}{m_0}\right), \tag{II.5}$$

where energies are calculated in the dipole's rest frame. In the parentheses, **1** is added to avoid negative contributions.

In the current QCD-CR implementation in PYTHIA, the value for *pseudo-particle* mass cut-off is also used in the calculation of the λ-measure. But in principle, one could treat them as independent parameters.

Eq. (3) is also used for a string piece connected to a junction, where the energy is measured in the junction rest frame. For a string piece connecting two junctions the λ-measure is given by [28]. The distance between the two junctions is also added to the calculation of the λ-measure of the new system and is given by

$$\lambda = \log\left(\beta_{j1}\beta_{j2} + \sqrt{(\beta_{j1}\beta_{j2})^2 - 1}\right), \tag{II.6}$$

where β_{j1} and β_{j2} are 4-velocities of the two junctions.

The parameter m_{0j} is used in the λ-measure calculation for the junction systems, while the m_0 is used for the λ-measure of the dipoles. Increasing m_{0j} results in a lower value for the λ-measure, which favours junction production. In this way the suppression of junctions in the model compared to proper $SU(3)$ algebra can be compensated by using a value above unity for the parameter C_j.

Figure II.6 shows the effect of varying m_0 on the final state charged multiplicity and $\langle p_\perp \rangle (N_{ch})$ compared with CMS data for $\sqrt{s} = 7$ TeV pp NSD collisions. It is evident that increasing m_0 increases the event multiplicity, by reducing the number of dipoles in CR, and vice versa.

Figure II.7 shows the effect of varying C_j on the final state charged multiplicity and $\langle p_\perp \rangle (N_{ch})$ compared with CMS data for $\sqrt{s} = 7$ TeV pp NSD collisions. The histograms show that reducing C_j below 1 increases the multiplicity because it reduces the number of junctions, which allows the production of many light hadrons. But enhancing its value will not have any significant effect on the observables. It is also evident from the histograms that varying m_0 has relatively strong effects on the observables compared to varying C_j.

3.3 Allowed dipole separation

The parameter δb_c constrains colour reconnection between the colour dipoles by constraining the transverse separation. We now fix all the parameters to QCD-CR (mode-0) and

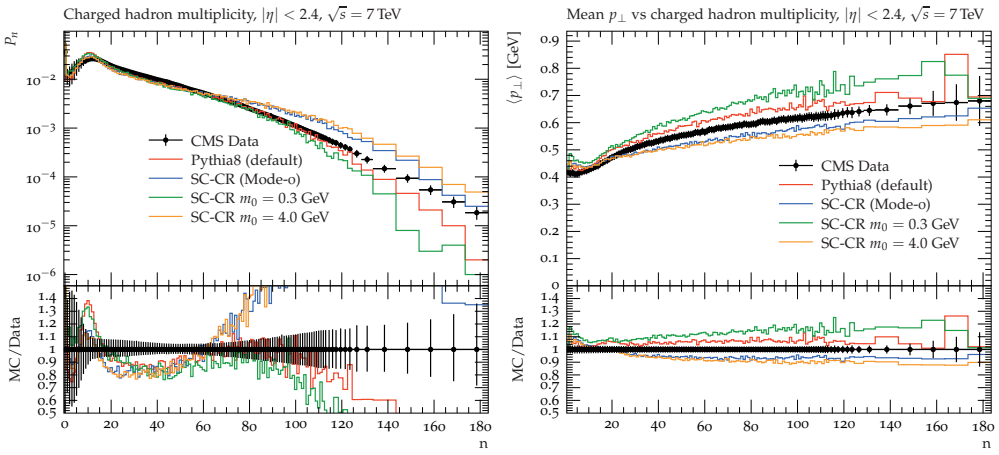


Figure II.6: The same as Figure II.4, but here the green and orange lines show the effect of varying the m_0 parameter in the SC-CR model.

vary the δb_c. The effect on the event multiplicity and $\langle p_\perp \rangle (N_{ch})$ due to varying δb_c between two and three dipoles to be colour reconnected is shown in Figure II.8, and it is compared with CMS data for $\sqrt{s} = 7$ TeV pp NSD collisions. From the figure, we see that reducing δb_c as low as 0.3 fm will reduce the CR significantly, increase the charged multiplicity and make the $\langle p_\perp \rangle (N_{ch})$ distribution flatter.

In a nutshell, each of the parameters discussed above has their contribution to the total charged multiplicity, which is summarised in table II.2. The direction of the arrows in the bracket next to every parameter shows the direction in which the histogram lines will move with respect to the earlier value of the parameter if all the other parameters are fixed.

Table II.2: A list of parameters and the effects on the overall multiplicity in PYTHIA. When the parameter is reduced (↓) the overall multiplicity will increase (↑) or decrease (↓), when keeping all other parameters fixed.

Parameters	Charged Multiplicity
$p_{\perp 0}^{\text{ref}}$ (↓)	↑
m_0 (↓)	↓
C_j (↓)	↑
δb_c (↓)	↑

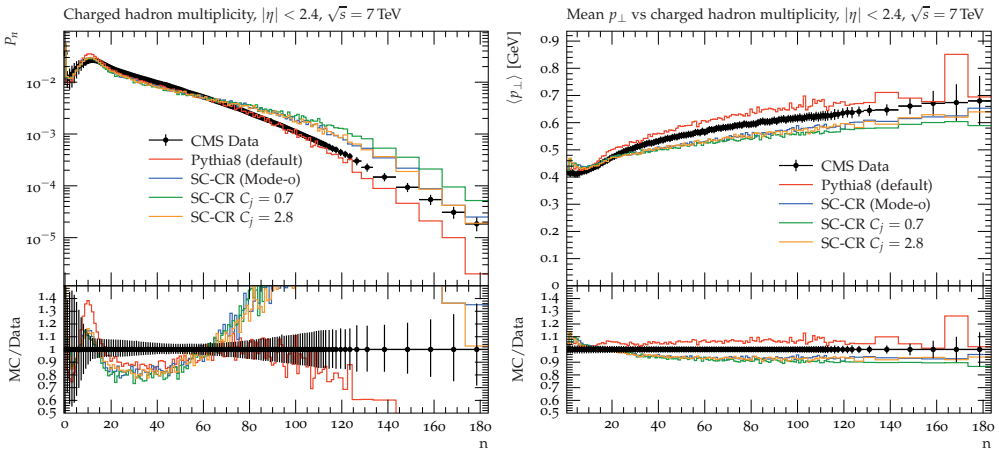


Figure II.7: The same as Figure II.4, but here the green and orange lines show the effect of varying the C_j parameter in the SC-CR model.

3.4 ϵ_{pom} for SND events

The default Angantyr simulates SND events using a proton-like pomeron PDF, a pomeron-proton interaction cross-section similar to the proton-proton non-diffractive interaction cross-section, and the pomeron flux according to Schuler and Sjöstrand [29], which gives a logarithmic distribution in the mass of the diffracted system, dm^2/m^2.

When we now introduce CR also between different sub collisions in a HI event we expect the overall multiplicity to go down, possibly destroying the good reproduction of data reported in [6]. To compensate for this we want to modify the pomeron flux in the SND events, and we use a conventional supercritical description for the pomeron flux, attributed to Berger *et al.* [30] and Streng [31]. The Pomeron Regge trajectory is parameterized as:

$$\alpha(t) = 1 + \epsilon_{pom} + \alpha^{'}(t), \tag{II.7}$$

giving the mass distribution $dm^2/m^{2(1+\alpha(t))}$. We can then vary ϵ_{pom}, which is a parameter to modify the mass distribution in the diffracted system.

The effect of changing the ϵ_{pom} parameter for the low and high multiplicity SND events in Angantyr is shown in Figure II.9. As expected we see an increase in the multiplicity when changing ϵ_{pom} from positive to negative values.

It should be noted that the default value of ϵ_{pom} in PYTHIA is 0.085, which gives a good fit for diffractive events in pp. However, when we here use this to generate SND events, we do not expect them to behave exactly like diffractive events. In [6] we discussed the relationship between the SND in double nucleon scattering and single diffraction (see the

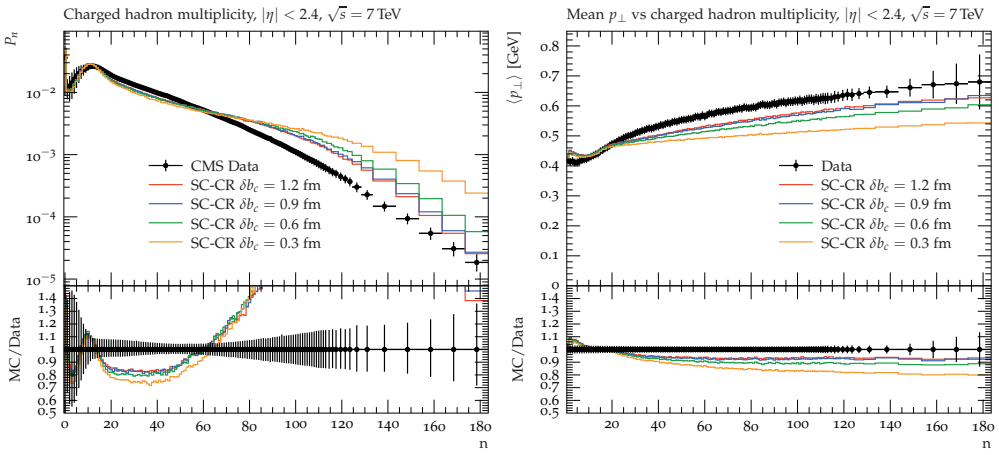
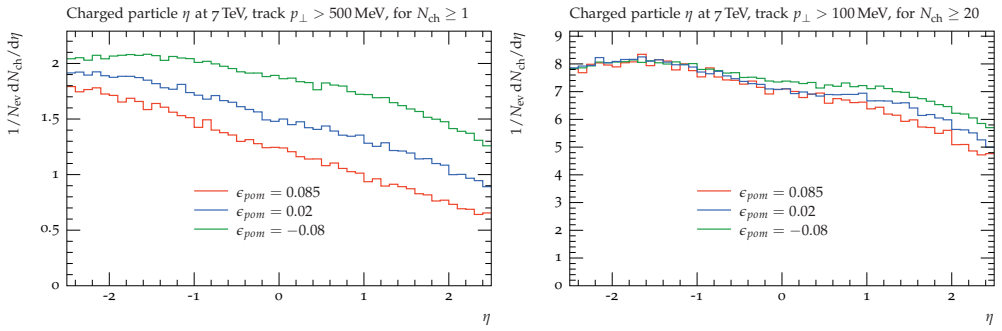


Figure II.8: Comparing variation in the δb_c parameter in PYTHIA. Events are generated for $\sqrt{s} = 7$ TeV pp NSD collisions and compared with CMS data [26]. *Left*: charged hadron multiplicity in $|\eta| < 2.4$. *Right*: $\langle p_\perp \rangle$ vs N_{ch} for $|\eta| < 2.4$.

discussion of Figure 22) and argued that the mass distribution should not be the same in the two cases. The mass distribution in single diffraction is related to the rapidity span of the diffracted system, $dm^2/m^2 \approx d\Delta y_m$, while in the SND we expect it to be proportional to the rapidity gap, $dm^2/m^2 \approx d\Delta y_{gap}$, single diffraction, with $\Delta y_{gap} = \Delta Y - \Delta y_m$, where ΔY is total rapidity span of the pp collision. Therefore, if a positive ϵ_{pom} is needed to describe single diffraction, using a negative value is quite reasonable for the SND, which is what we need in order to compensate for the decrease in multiplicity due to the CR in HI collision.

3.5 CR effects in $\mathrm{p}A$ and AA collisions

This is the first time that the effects of CR have been introduced and studied in a heavy-ion collision event-generator. In section 2, we show the importance of the CR in the context of $\mathrm{e^+e^-}$ and pp collision event simulations in PYTHIA. In this work, we are further extending the Angantyr model of PYTHIA with a global CR, which is constrained by the transverse separation of the colour dipoles. For the sake of completeness, it is interesting to see how large the effects of colour reconnections really are before retuning. We have looked at pPb and PbPb central charged event multiplicities for the default Angantyr setup, where there are colour reconnections only inside individual NN sub-collisions. This we then compared to the case where reconnections are switched off altogether, and to the case where we have global reconnections between (almost) all dipoles in the event. We have used the QCD-CR (mode-0) parameters for the latter, but the δb_c value is set to 7 fm.



(**a**) η distribution for $N_{ch} > 1$ (**b**) η distribution for $N_{ch} > 20$

Figure II.9: Angantyr SND events, which are single diffractive (SD) events generated for $\sqrt{s} = 7$ TeV pp collisions, with `HeavyIon:Mode = 2`, and `Angantyr:SDTests = on` in PYTHIA. Here the SD events are generated in the `Angantyr:SASDmode = 4`. Changing ϵ_{pom} from positive to negative values, the event multiplicity increases, and the overall distribution is a bit flatter, but closer to the diffractive proton side (negative η), its similar to ND as expected.

The results are shown in Figure II.10, where it is clear that the effects of colour reconnections are substantial mainly for the highest multiplicities. The average multiplicities are however only moderately affected, with a 20% increase for PbPb when reconnections are switched off, and a 10% decrease with global reconnections. The corresponding effects in pPb are smaller, with $+10\%$ and -5% respectively.

4 Results

First, we note that we have not done a very sophisticated retuning of the parameters, *e.g.*, using the Professor framework [32]. Our focus has been to get a reasonable multiplicity distribution for pp and also for pA. There is some tension in data, favouring a larger δb_c in pp, but a lower one in pA. In the end, we basically selected the value $\delta b_c = 0.5$ fm.

We show the tuned values under the SC-CR (tuned) column and the default values of those parameters under the QCD-CR (mode-0) column in Table II.3. For pp, all other parameters are the same as for QCD-CR (mode-0), which is shown in Table II.1.

To treat a heavy-ion event as a single event we want to stack all pp-like sub-collisions at the parton level, and then apply CR with spatial constraints on all colour dipoles. To do so, we first turn off colour reconnection in all individual sub-collisions (`ColourReconnection:reconnect = off`) and instead switch it on only for the hadronization stage (`ColourReconnection:forceHadronLevel`

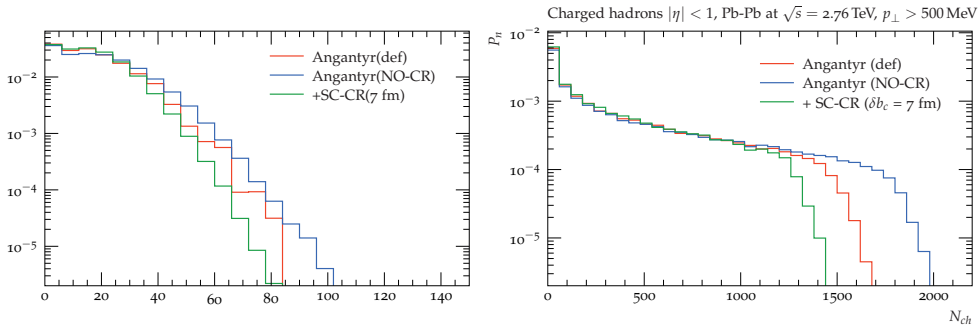


Figure II.10: The distribution in charged hadrons multiplicity in $\mid \eta \mid < 1$ and with $p_\perp > 500$MeV for pPb events at $\sqrt{s_{NN}} = 5.02$ TeV (left) PbPb events at $\sqrt{s_{NN}} = 2.76$ TeV (right). The red lines are for the events generated with the default Angantyr. The blue lines are the events with no CR reconnection between the colour dipoles, and the green lines are the events generated with QCD-CR (mode-0) parameters, but the δb_c value is set to 7 fm, which is almost the radius of a Pb nucleus.

Table II.3: The list of parameters and their new values compared to their values in the QCD-CR (Mode-0) tune. The new values are used in PYTHIA to generate pp collisions with the spatial constraint in the SC-CR model.

Parameters	QCD-CR (Mode-0)	SC-CR (tuned)
`PartonVertex:setVetex`	-	on
`MultiPartonInteractions:pT0Ref`	2.12	2.37
`ColourReconnection:m0`	2.9	1.05
`ColourReconnection:junctionCorrection`	1.43	1.37
`ColourReconnection:dipMaxDist`	-	0.5

`= on`). [2]

Now, with this feature enabled, pA and AA collision events undergo CR only once per HI event, and after CR the entire event proceeds to hadronization. The list of changed and new parameters for HI collisions is shown in Table II.4. The prefix "HI" for some of the parameters means that they are only used for SND sub-collisions.

[2]At the moment this combination works only within the Angantyr framework. Moreover, it does not work with PYTHIA's default MPI-based CR. For pp collisions, this feature is not that useful, as there's only one proton-proton collision which is a single event by default. But if a user wishes, one can use the above combination for pp collisions as well by setting the `Heavyion:mode = 2` flag to generate pp collisions within the Angantyr setup.

Table II.4: The list of parameters and their new values compared to their default values in the QCD-CR (Mode-0) tune. The new values are used in the Angantyr model to generate pA and AA collisions with the spatial constraint in the SC-CR model. The values for the parameters used to generate pp events are not changed, and they are same as in Table II.3, while the other parameters are same as in Table II.1.

Parameters	QCD-CR (Mode-0)	SC-CR (new)
`ColourReconnection:reconnect`	on	off
`ColourReconnection:forceHadronLevelCR`	off	on
`MultiPartonInteractions:pT0Ref`	2.12	2.37
`PartonVertex:setVetex`	-	on
`ColourReconnection:m0`	2.9	1.05
`ColourReconnection:junctionCorrection`	1.43	1.37
`ColourReconnection:dipMaxDist`	-	0.5
`HIMultiPartonInteractions:pT0Ref`	2.12	2.37
`HISigmaDiffractive:mode`	-	0
`HISigmaDiffractive:PomFlux`	-	3
`HISigmaDiffractive:PomFluxEpsilon`	-	-0.04
`HIBeamRemnants:remantMode`	-	1
`HIBeamRemnants:Saturation`	-	5
`BeamRemnants:beamJunction`	off	on
`HIBeamRemnants:beamJunction`	off	on

4.1 pp results

We begin with comparing pp events generated at $\sqrt{s} = 7$ TeV using PYTHIA default and re-tuned spatially constrained QCD-CR in PYTHIA. We show the charged multiplicity and $\langle p_\perp \rangle (N_{ch})$ distributions for these two setups and compare the results with ATLAS [27] and CMS [26, 33] experiments. The pp results compared with ATLAS are minimum bias events, while the events compared with the CMS experiment are generated as non-single diffractive (NSD) events in PYTHIA. We use $c\tau > 10$ mm as the definition for primary particles when comparing our simulated events against ATLAS and CMS data.

In figure II.11 we show that with the new constraint on QCD-CR and the new tuned parameters, we are able to reasonably reproduce the average event multiplicity and $\langle p_\perp \rangle (N_{ch})$ distributions from the ATLAS and CMS for $\sqrt{s} = 7$ TeV pp collisions. This should not be surprising, since these are the distributions we tuned to.

Figure II.12 shows the distribution in pt and N_{ch} compared to ATLAS data, and here we observe that the charged multiplicity distribution for the SC-CR setup drops too quickly

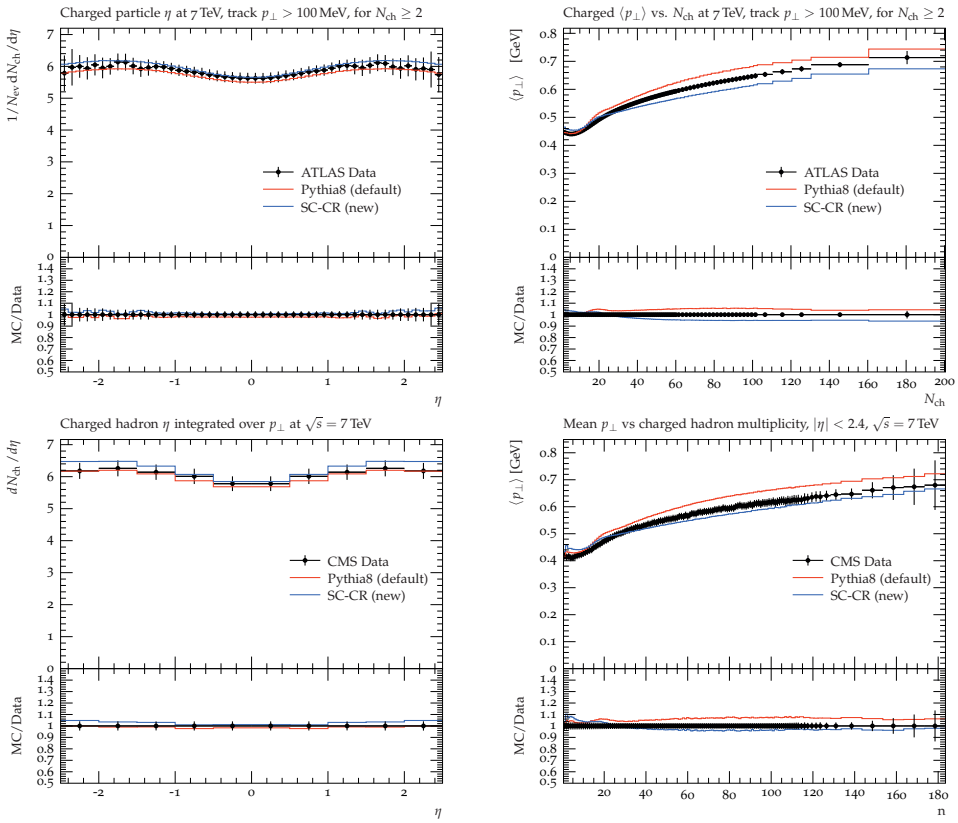


Figure II.11: Events are generated for $\sqrt{s} = 7$ TeV pp collisions minimum bias events are compared with ATLAS [27] results (top row), and non-single diffractive events are compared with CMS [26, 33] results (bottom row). The left plots show the pseudo-rapidity distribution of charged particles, while the right plots show $\langle p_\perp \rangle$ as a function of multiplicity. PYTHIA (default) is the default pp collisions with the PYTHIA Monash tune setup. SC-CR (new) is produced by our spatial constraint CR model, with the parameters given in Table II.3.

for high multiplicities. We also see that the transverse momentum spectrum becomes too hard. These effects are related: when we introduce the spatial constraint, there will be less reconnection and thus higher multiplicity; this is then compensated somewhat by decreasing m_0, but this is not enough (remember that QCD-CR (mode 0) also has a too high multiplicity when our improved junction handling is introduced); so we increase the $p_{\perp 0}^{\text{ref}}$ but that mainly decreases the multiplicity of low transverse momentum particles, giving a too hard spectrum. The conclusion is that our tuning may have been a bit naive, but for now, we are satisfied with that we have the overall multiplicity under control.

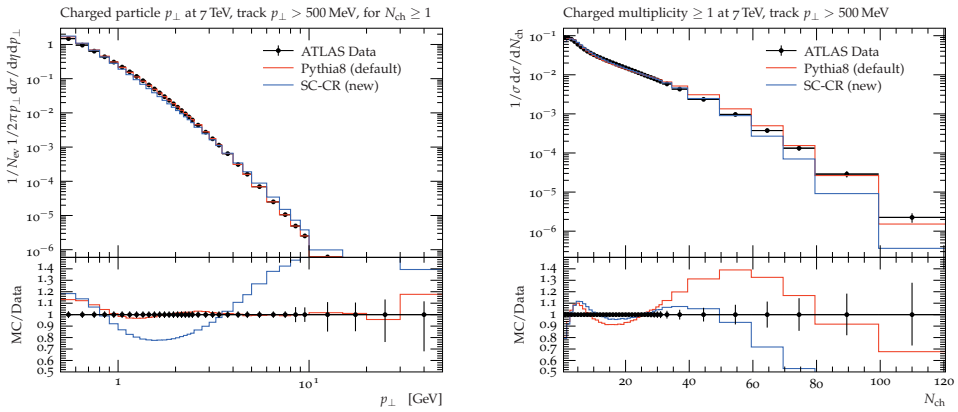


Figure II.12: Events are generated for $\sqrt{s}$ =7 TeV pp collisions minimum bias events and compared with ATLAS [27] results. The left plot shows the distribution of $p_\perp$ for charged particles and the right shows the distribution in charged multiplicity. The lines are the same as in Figure II.11.

Although we in this paper are mainly concerned with general multiplicities and transverse momentum distributions, it is interesting to also look at other effects of the QCD-CR model. In particular the junction reconnections are interesting, as they are known to substantially affect the baryon-to-meson ratios.

Figures II.13 and II.14 show baryon-to-meson ratios for non-strange and strange baryons for the pp collision at $\sqrt{s}$ =7 TeV and compared with ALICE [34] results. The events are generated with default PYTHIA set-up and for the spatially constrained QCD CR model. The ratios of p/π, Λ/π, Λ/K^0_s, Ξ/π, and Ω/π are produced for the combined yield of particles and respective anti-particles. We observe that the baryons production is enhanced, which is an expected outcome of using the QCD CR model [17]. From the p/π ratio plots in figure II.13, we notice that the model produces too many protons, while the model improves the distribution of Λ, Ξ, and Ω baryons. We note that in pp the spatial constraint is not very important for these ratios, and similar results are expected for the standard QCD-CR model without our modifications.

The results of the Λ/K^0_s ratio are showing agreement with the ALICE data similar to the results obtained using the rope-hadronization [35] model in PYTHIA. It is important to note that the CR model acts on the colour dipoles, while the rope-hadronization acts on the Lund strings during the string fragmentation. Hence it will be interesting to investigate if both models together can improve PYTHIA results and to what extent each of the models will influence the hadrons yield in all three collision systems. At the moment from the figures II.12 and II.13 the p_t distribution of the charged particles and the proton yield both are shortcomings of the spatially constrained QCD CR model.

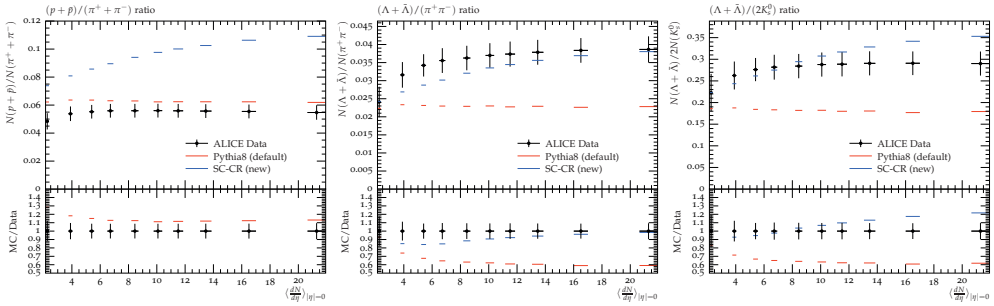


Figure II.13: From left to right the ratios of p/π, Λ/π, and Λ/K^0_s is plotted against the average charged-particle multiplicity in the central pseudo-rapidity. pp collisions minimum bias events are generated at $\sqrt{s}$ =7 TeV and compared with ALICE [34] results. The lines are the same as in Figure II.11.

4.2 pPb **results**

We show here comparisons between the Angantyr-generated average charged particles multiplicity as a function of pseudo-rapidity in different centrality bins for pPb collision events at $\sqrt{s_{NN}} = 5.02$ TeV, and experimental results from ATLAS [36]. For the SC-CR we use the parameter values we tuned to pp with the heavy-ion specific parameters listed in table II.4, while the other parameters retain their values for QCD-CR (mode-0) in table II.1. The centrality determination is made separately for each generated dataset using the standard Rivet [37] routine.[3]

The results are shown in Figure II.15, and it is clear that just allowing for CR between individual sub-collisions (blue lines labelled "+ SC-CR ($\epsilon_{pom} = 0$)") seriously degrades the good reproduction of data for the default Angantyr setup. This is expected as the CR necessarily will reduce the multiplicity in events with many participating nucleons. Trying to counteract this by decreasing ϵ_{pom} and thus increasing the multiplicity of SND subevents, improves the reproduction of data, but as seen in Figure II.15 (green lines), for the most central events it is not quite enough. We also see that for more peripheral events there is a tendency to overestimate the multiplicity, and decreasing ϵ_{pom} further to increase multiplicity in central events would also worsen the description of peripheral events.

It should be noted here that the concept of centrality in pPb, is a bit complicated, and we showed in [6] that any centrality measure will not only be sensitive to the number of participating nucleons, but will also be sensitive to multiplicity fluctuations in individual sub-collisions, especially the most central bins. Looking back on figure II.12, we see that our SC-CR tune have much fewer fluctuations to very large multiplicities than default

[3]The centrality routine is called `ATLAS_pPb_Calib`, and the multiplicity analysis is then made with the option `cent=GEN` in Rivet.

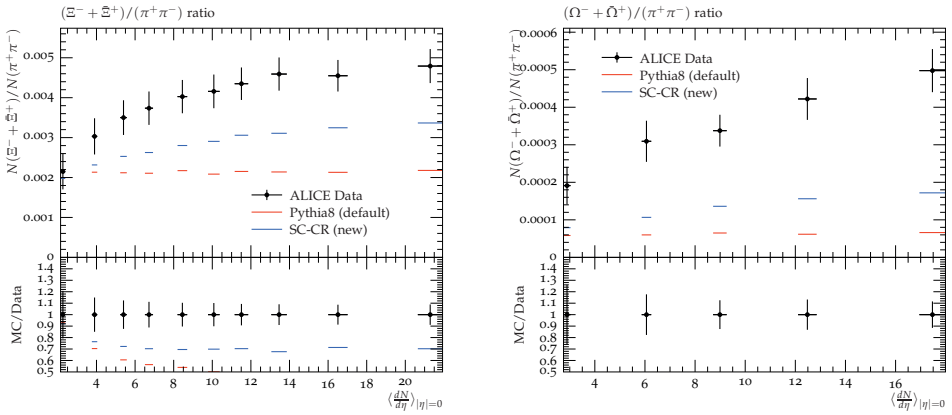


Figure II.14: From left to right the ratios of Ξ/π and Ω/π is plotted against the average charged-particle multiplicity in the central pseudo-rapidity. pp collisions minimum bias events are generated at $\sqrt{s}$ =7 TeV and compared with ALICE [34] results. The lines are the same as in Figure II.11.

Pythia in pp events, and it is possible that improving this could also help the description of very central pPb events.

Figure II.16 shows the baryon-to-meson ratio for protons and Λ baryons in pPb collision events $\sqrt{s_{NN}} = 5.02$ TeV, and the results are compared with ALICE data [38]. We can notice that the results follow the trend of baryon-to-meson ratio in Figure II.13, the SC-CR model produces too many protons irrespective of the event centrality, while the Λ/π distribution is improved compared to Angantyr (default).

4.3 PbPb results

PbPb events at $\sqrt{s_{NN}} = 2.76$ TeV were generated using the Angantyr model with its default setup and the SC-CR setup (table II.4). The simulation results are compared with ALICE [39] in figure II.17 (left) for the central ($\mid \eta \mid < 0.5$) charged multiplicity in different centrality bins[4]. We see here the same tendency as in pPb, that the multiplicity is reduced for central events when including CR between sub-collisions, while for more peripheral events it is less affected. The effect is roughly the same, $\sim 20\%$, for the most central events. Even though the multiplicity in PbPb is much higher, the sub-collisions are more spread out in impact parameter, and the spatial constraint, δb_c, therefore severely restricts the effect of CR.

[4]Again the centrality bins are calculated on the generated data, this time using the `ALICE_2015_PBPBCentrality` analysis in Rivet

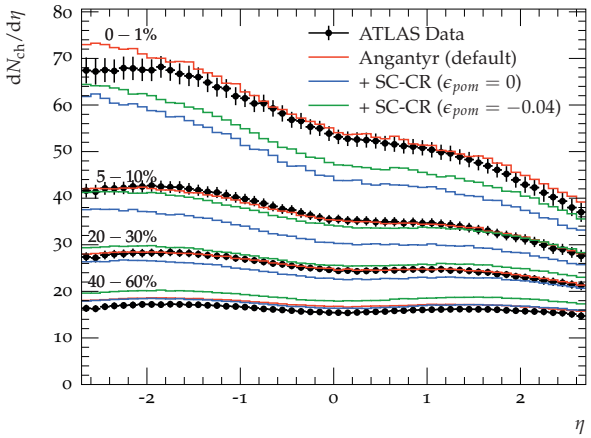


Figure II.15: Average charged hadron multiplicity as a function of η for different centrality bins for pPb events are generated at $\sqrt{s_{NN}} = 5.02$ TeV and compared with ATLAS [36] results. The red lines are Angantyr default, the blue and green ones are also from Angantyr but the blue with the spatially constrained QCD-CR (with the new tune obtained for pp collisions), and green are the same but with the tuned value of $\epsilon_{pom} = -0.04$. The top group of lines show the results from the centrality interval $0-1\%$, followed by the centrality intervals $5-10\%$, $20-30\%$ and, at the bottom, $40-60\%$.

It should be noted that the effects on the centrality binning of fluctuations in sub-collision multiplicity are not so important in PbPb, compared to pPb. This is reflected in the average number of participating nucleons as a function of centrality obtained from a Glauber model, shown in figure II.17 (right). Here we see that the agreement between Angantyr, with and without SC-CR agrees very well with each other and with the number obtained from the data.

We noted in [6] that, while the default Angantyr gives a reasonable description of the multiplicity in PbPb data, other observables were not as well reproduced. In particular this applies to $p_\perp$ spectra, which we show here in figure II.18 compared with ATLAS [40] for central and peripheral events. Considering that we already in figure II.12 showed that SC-CR degrades the description of the $p_\perp$ distribution in pp, it would be unlikely that it would be better in PbPb. Indeed we see in the figure that adding global colour reconnection in PbPb, rather makes the description of data worse in central collisions. For peripheral collisions, one could argue that there is an improvement in the high-$p_\perp$ tail, but the overall performance is still not very good.

So far for pp and pPb, we have shown proton-to-pion and some of the hyperons-to-pion

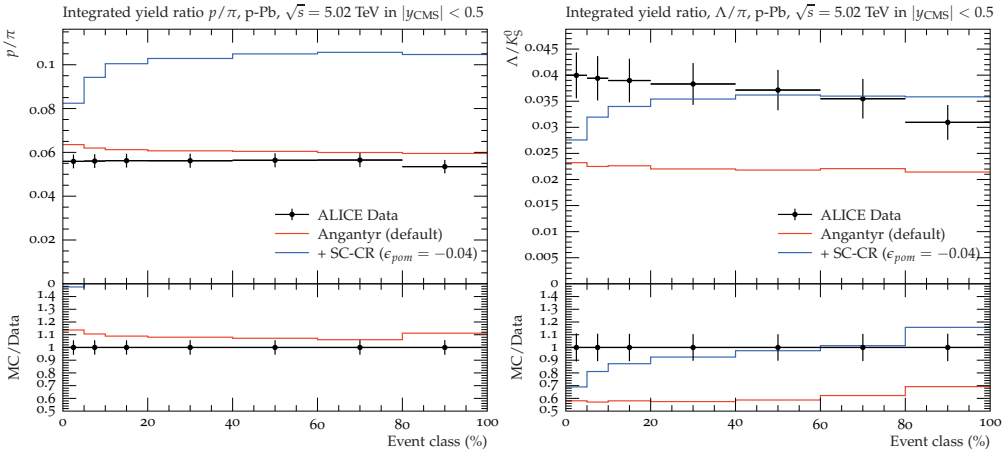


Figure II.16: The ratios of p/π (left) and Λ/π (right) are plotted for different centrality for pPb collision events at $\sqrt{s}$ =5.02 TeV and compared with ALICE [38] results. The red line is generated with the default Angantyr settings, while the blue line is SC-CR with the tune obtained for pA collisions.

ratios. For PbPb we want to refrain from showing any baryon-to-meson results. The SC-CR model fails to reproduce the charged multiplicity for many centrality bins in PbPb collision events. Hence the model's agreement or disagreement with the experimental data of baryon-to-meson ratios will be irrelevant at this stage. We want to improve the model description for the overall charged multiplicity and the p_t distribution of the produced particles before we test the model efficiency against the different identified particle yields in AA collisions.

The SC-CR model increases the baryon production in pp and pPb collision systems. The result of strangeness enhancement in the strange baryons sector without any assumption of thermalised medium opens the possibility for an alternative explanation for the observed strangeness enhancement in heavy-ion collisions. The results from heavy hyperons are much below the experimental observations (figure II.14, but we should note that these results are producing a similar trend of strangeness enhancement to that of rope-hadronization [35]. We hope that in future combining these two models will improve the PYTHIA/Angantyr model efficiency in reproducing the strange hadrons yield.

5 Discussion and outlook

We have here presented a first attempt to extend the concept of colour reconnections to apply also between sub-collisions in heavy ion events. This was done by modifying the

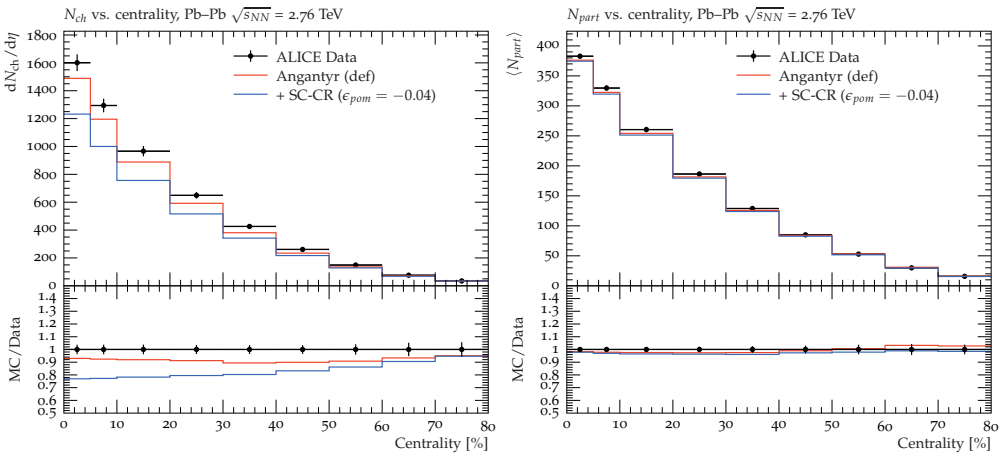


Figure II.17: *Left*: Central ($|\ \eta\ | < 0.5$) charged multiplicity as a function of centrality for PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. The red line is generated with the default Angantyr settings, while the blue line is SC-CR with the tune obtained for pA collisions. The data is from ALICE [39]. *Right*: number of participating nucleons, N_{part}, as a function of centrality for the same event samples, compared with Glauber-model calculations from ALICE [39].

QCD colour reconnection model in PYTHIA to limit the spatial distance between dipoles that are allowed to reconnect. This was done with a simple cutoff in impact parameter distance, δb_c, for which we found a value 0.5 fm, to give reasonable results.

This is clearly a quite naive approach, and the aim is mainly to understand the phenomenology of allowing inter-sub-collision CR in heavy ion collisions. There are many paths to improve our simple model. It is not unnatural, *e.g.*, to let the value of δb_c depend on the local density of dipoles, and/or the transverse momentum of the partons involved in the reconnections. The role of m_0 also needs to be studied more, and this parameter could also be allowed to depend on such local features of the event. There are also other reconnection models that can be considered for use in heavy ion collisions, such as the perturbative swing model in Ref. [12, 23].

Our results are not perfect, but we still feel that they are encouraging. Introducing the cutoff makes the reproduction of pp minimum bias data worse than in default PYTHIA, but not very much worse than what is obtained with the QCD-CR model. For pA collisions we get a reduced multiplicity, as expected, but we find that it can be compensated by modifying the treatment of secondary non-diffractive sub-collisions in the Angantyr model. Finally, we find that when extrapolating to full AA events we again obtain a too low multiplicity for central events, but the reduction is typically below 20%.

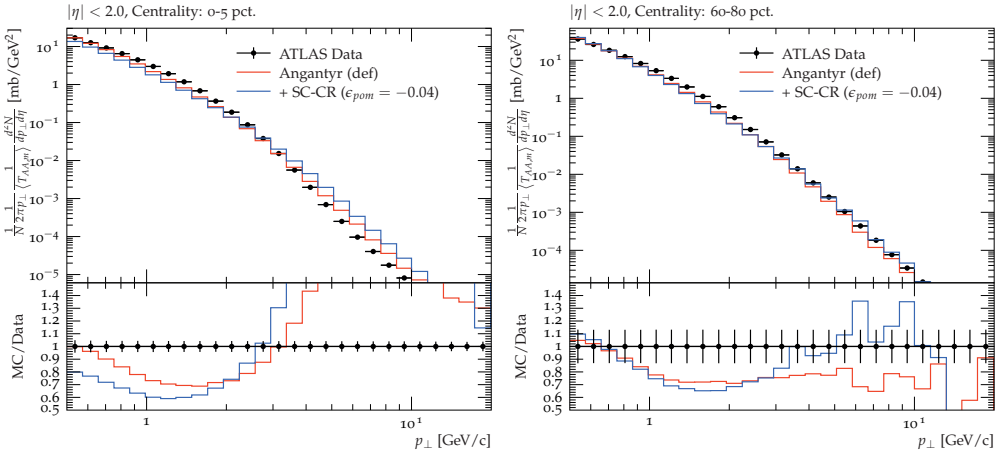


Figure II.18: Transverse momentum distribution of charged particles in different centrality bins for PbPb collisions, generated at $\sqrt{s_{NN}} = 2.76$ TeV and compared with ATLAS [40] results. The lines are the same as in figure II.17.

It should be remembered here that there are many other effects to consider, both in high-multiplicity pp collisions and in heavy ion collisions. In our group, we have considered so-called string shoving [10, 11] and rope hadronization [12], as well as hadronic rescattering [14]. Especially the latter would be interesting study together with colour reconnection in heavy ion collisions, since it has been found that rescattering typically increases the multiplicity for central collisions [15].

The QCD-CR model has shown enhanced production of the baryons compared to the default MPI-based CR model, which is purely a byproduct of junction systems. Hence, there may be some effects similar to strangeness enhancement due to the QCD-CR model in the baryon sector [41]. The QCD-CR model has shown effects similar to collectivity [42]. This work will also enable us to study and investigate the contribution of different event simulation stages like CR, string interactions (shoving and rope), and hadronic rescattering on the various final state observables in pp, pA, and AA collisions. In future work, we will show the results of the new implementation for the flavour production and collectivity-like behaviour in all three collision systems, namely pp, pA, and AA.

A top priority for future work is therefore to develop our implementation to allow the study of the simultaneous effects of all these models. Only then one would be able to do a proper tuning of the involved parameters. In doing so we would again follow the procedure used here, *i.e.*, first tune to minimum bias pp observables of multiplicities and transverse momentum, and then adjust parameters that also depends on inter-sub-event effect by tuning to similar observables in pA in order to be able to have a parameter-free extrapolation to AA events.

Acknowledgments

We thank Gösta Gustafson, Torbjörn Sjöstrand and Christian Bierlich for interesting discussions and important input to this work.

This work was funded in part by the Knut and Alice Wallenberg foundation, contract number 2017.0036, Swedish Research Council, contracts numbers 2016-03291 and 2020-04869, in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No 668679, and in part by the MCnetITN3 H2020 Marie Curie Initial Training Network, contract 722104.

6 Appendix: Junction Fragmentation

The hadronization in Pythia is done through Lund string fragmentation. Pythia has two methods for hadron production; string fragmentation and mini string fragmentation. Both are fundamentally based on the same mechanism, but the latter is an approximation for short strings (mini-strings), which becomes more important with many reconnections in HI events. Depending on the invariant mass of the colour singlet string/junction system, Pythia's algorithm selects one of the two methods to produce primary hadrons.

String fragmentation fragments long strings. For the junction systems, the algorithm fragments two lower energy junction legs first, and it starts from the lowest energy leg. The two low-energy junction legs are fragmented until every leg is left with a parton directly connected to the junction. Later, the two partons from these low-energy legs are combined to form a diquark (or anti-diquark). The diquark is then treated as one end of the remaining highest energy junction leg. At this stage, the junction system no longer exists, and the last junction leg with a diquark at one end undergoes further string fragmentation as a normal string. Figure II.19 shows the progress of the junction break-ups, formation of a diquark, and transformation of the junction system into a string system.

Mini-string fragmentation is used to hadronize short strings. It produces one or two hadrons depending on the energy in the string, and low-energy junction systems are not treated in the mini-string fragmentation. Prior to this work, if low-energy junction systems are found then such an event is aborted and a new event is generated. When we merged multiple pp-like events at the parton level in a HI event and tried to perform CR in the entire event, we observed enhancement in the number of junction systems, and many of them are below the invariant mass cut-off for string fragmentation. All those junction systems have to be fragmented within a mini-string fragmentation module, otherwise, the majority of HI events are aborted due to one or more untreated colour singlet systems in the hadronization stage. Hence, we developed two new fragmentation functions: "Mini-

Figure II.19: Junction fragmentation steps. *Left*: A colour singlet junction system. In this example, the lowest energy leg is the one with a quark directly attached to the junction. The highest energy leg is the one with the longest chain of dipoles. *Centre*: The junction system after both the low-energy legs are fragmented and left with a quark directly connected to the junction. The highest energy leg is intact. *Right*: The final state of the junction system is shown. Here, the quarks of the two low-energy junction legs are combined to form a diquark (represented as (qq)), and it is attached at the end of the highest-energy leg. The remaining junction system is now replaced with a single string, and it will fragment according to the normal string fragmentation.

Junct2Hadrons" and "MiniJunct2Baryons", in the mini-string fragmentation.

Before we provide an overview of the technical details of the two new functions, we like to discuss some additional constraints we introduced inside the QCD-CR model. We observe that most of these junction systems in pA and AA systems, have diquarks in the junction legs. Now, as we mentioned earlier, in the string fragmentation the two lower energy legs of a junction system are fragmented and the remainder is merged to form a string connecting the diquark with the highest energy junction leg. But in the case of the junction being connected to a diquark, it often occurs that either both or one of the low-energy legs is left with a diquark. Pythia can not merge such junction legs, where one or both legs have a diquark.

Sometimes, the highest energy leg also has a diquark at the end. Now, after merging two low-energy legs into a diquark, the string has a diquark at both ends. Although string fragmentation is a probabilistic mechanism and the strings with two diquarks may fragment successfully, sometimes they fail due to the last string piece being left with a diquark at both ends. The last piece can not be hadronized as a tetra-quark system at this moment, although that could in principle be considered. We introduce an additional attempt by forwarding such a failed string fragmentation to mini-string fragmentation, where the string/junction system is forced to produce one or two hadrons.

To avoid recurring failures in the string fragmentation due to junction systems containing diquarks in pA and AA collisions, we do not allow dipoles with a diquark to undergo CR. This constraint is provided with a Boolean flag, so future users can test the options with or without allowing CR for diquarks containing dipoles. In pp collisions, another source of diquarks prior to string fragmentation is beam remnants. Pythia allows users to choose if

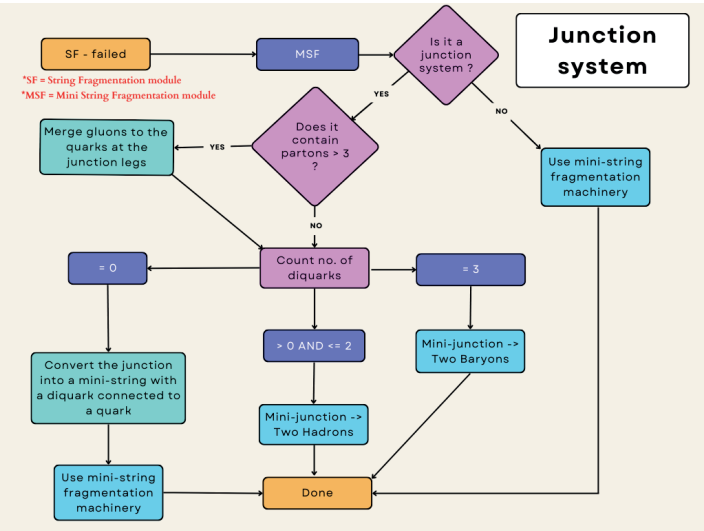


Figure II.20: Overview of the steps taken if the string fragmentation module fails to produce hadrons, and the handling of junction systems in the mini-string fragmentation to produce primary hadrons. The flow chart shows examples of cases with quarks and diquarks, but the junctions can also have anti-quarks and anti-diquarks.

the beam remnants may form a junction or a diquark with the remaining valance quarks of the beam remnant. Here, we choose junction formation to reduce the preexisting number of diquarks in the event during the CR. We continue to use this setup not only for pp collisions but also for pA and AA collisions.

Now, let's go back to the junctions in mini-string fragmentation. The first step is to check if the junction contains any gluons or not. If any of the junction legs have gluons, we reduce the junction system size by absorbing the gluons into the diquark or quark of the respective junction legs. The idea here is to simplify the next steps of producing hadrons via "junction collapse". We reduce the complex junction system to a simple junction system, where every leg contains only a quark or a diquark or respective anti-particle at its end. Now, we calculate the number of diquarks in the junction system. As shown in Figure II.20, the junction can have up to three diquarks, when all the junction legs have a diquark. Depending on the number of diquarks, we have three possible outcomes for the junction to collapse and produce primary hadrons.

Case A (All the legs have quarks):

In this case, we follow the steps similar to the last stage of junction fragmentation in the string fragmentation mechanism. We merge the two lowest energy legs to form a diquark.

Then we reduce the junction into a mini-string containing a diquark and a quark. After that, we use the existing mini string fragmentation functions to produce primary hadrons.

Case B (The junction legs have a maximum of two diquarks):

Here, depending on whether the junction has one or two diquarks we collapse it to produce two mesons, or one baryon and one meson. We observed that the junctions containing one diquark have anti-quarks on the remaining legs and if it's an anti-diquark then quarks. Similarly, in the case of two diquarks, then the remaining leg has an anti-quark, and if two anti-diquarks then a quark. Hence, we treat the junction as a single entity containing a given number of diquarks, anti-diquarks, quarks and anti-quarks, and we produce hadrons accordingly.

Case B.1 (One diquark, and two anti-quarks):

Here, we break the diquark into two quarks. Then we pair each quark with a randomly selected anti-quark from the two anti-quarks. We check if the produced mesons have the correct quark flavours or not. Then we check if the sum of their masses is less than the invariant mass of the junction system or not. If the sum of the masses is below the junction system's invariant mass, then the mesons are assigned momenta as if a mother particle decays into the two daughter particles. If it fails then we repeat the process several times. If all attempts fail, the function returns a message about failed attempts, and the event is aborted.

Case B.2 (Two diquarks and one anti-quark):

Here, a diquark with the lowest invariant mass is broken into two quarks. One of the quarks is randomly selected and assigned to the other diquark, and the other quark is paired with the anti-quark. Again we check the flavours and compare the sum of the masses of the baryon and the meson with the junction system mass. If the sum of the masses is below the junction system's invariant mass then the hadrons are assigned momenta as if a mother particle decays into the two daughter particles. If it fails then we repeat the process several times. If all attempts fail, the function returns a message about failed attempts, and the event is aborted.

Case C (All legs have diquarks):

In this case, we have in total six quarks, and we collapse the junction to produce two baryons. We follow the procedure similar to case B.2, the diquark with the lowest invariant mass is split into two quarks. One of the quarks is randomly assigned to one of the remaining two diquarks. The rest of the steps are the same.

References

[1] Gyulassy, M. & McLeran, L. *Nucl. Phys. A* **30** (2005).

[2] Et al. (ALICE collaboration), J. A. "Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions". *Nature Phys.* **13.** arXiv: 1606.07424 (2017).

[3] Et al. (CMS collaboration), V. K. "Evidence for collectivity in pp collisions at the LHC," *Phys. Lett.B.* **765.** arXiv: 1606.06198 (2017).

[4] Sjöstrand, T. *et al.* "An Introduction to PYTHIA 8.2". *Comput. Phys. Commun.* **191,** 159 (2015).

[5] Bierlich, C. *et al.* "A comprehensive guide to the physics and usage of PYTHIA 8.3". *SciPost Phys. Codebases,* 8. arXiv: 2203.11601 [hep-ph] (2022).

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

[7] Glauber, R. "Cross Sections in Deuterium at High Energies". *Physical Review* **100,** 242–248 (1955).

[8] Miller, M. L., Reygers, K., Sanders, S. J. & Steinberg, P. "Glauber Modeling in High Energy Nuclear Collisions". *Ann.Rev.Nucl.Part.Sci.* **57,** 205–243 (2007).

[9] Good, M. L. & Walker, W. D. "Diffraction dissociation of beam particles". *Phys. Rev.* **120,** 1857–1860 (1960).

[10] Bierlich, C., Gustafson, G. & Lönnblad, L. "Collectivity without plasma in hadronic collisions". *Phys. Lett. B* **779,** 58–63. arXiv: 1710.09725 [hep-ph] (2018).

[11] Bierlich, C., Chakraborty, S., Gustafson, G. & Lönnblad, L. "Setting the string shoving picture in a new frame". *JHEP* **03,** 270. arXiv: 2010.07595 [hep-ph] (2021).

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

[13] Bierlich, C. "Rope Hadronization and Strange Particle Production". *EPJ Web Conf.* **171,** 14003. arXiv: 1710.04464 [nucl-th] (2018).

[14] Sjostrand, T. & Utheim, M. *A Framework for Hadronic Rescattering in pp Collisions* 2020. arXiv: hep-ph/2005.05658v1.

[15] Bierlich, C., Sjöstrand, T. & Utheim, M. "Hadronic rescattering in pA and AA collisions". *Eur. Phys. J. A* **57,** 227. arXiv: 2103.09665 [hep-ph] (2021).

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

[17] Christiansen, J. R. & Skands, P. Z. "String Formation Beyond Leading Colour". *JHEP* **08,** 003. arXiv: 1505.01681 [hep-ph] (2015).

[18] Ames-Bologna-CERN-Dortmund-Heidelberg-Warsaw, Breakstone, A. *et al.* "Multiplicity Dependence of Transverse Momentum Spectra at ISR Energies". *Phys. Lett. B* **132,** 463–466 (1983).

[19] UA1, Albajar, C. *et al.* "A Study of the General Characteristics of $p\bar{p}$ Collisions at $\sqrt{s}$ = 0.2-TeV to 0.9-TeV". *Nucl. Phys. B* **335,** 261–287 (1990).

[20] Gustafson, G., Pettersson, U. & Zerwas, P. M. "Jet Final States in W W Pair Production and Color Screening in the QCD Vacuum". *Phys. Lett. B* **209,** 90–94 (1988).

[21] Sjöstrand, T. & Khoze, V. A. "On Color rearrangement in hadronic W+ W- events". *Z. Phys. C* **62,** 281–310. arXiv: hep-ph/9310242 (1994).

[22] Gustafson, G. & Häkkinen, J. "Color interference and confinement effects in W pair production". *Z. Phys. C* **64,** 659–664 (1994).

[23] Lönnblad, L. "Reconnecting colored dipoles". *Z. Phys. C* **70,** 107–114 (1996).

[24] Friberg, C., Gustafson, G. & Häkkinen, J. "Color connections in e+ e- annihilation". *Nucl. Phys. B* **490,** 289–305. arXiv: hep-ph/9604347 (1997).

[25] Andersson, B., Dahlqvist, P. & Gustafson, G. "ON LOCAL PARTON HADRON DUALITY. 1. MULTIPLICITY". *Z. Phys. C* **44,** 455 (1989).

[26] CMS, Khachatryan, V. *et al.* "Charged Particle Multiplicities in pp Interactions at $\sqrt{s} = 0.9$, 2.36, and 7 TeV". *JHEP* **01,** 079. arXiv: 1011.5531 [hep-ex] (2011).

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

[28] Sjöstrand, T. & Skands, P. Z. "Baryon number violation and string topologies". *Nucl. Phys. B* **659,** 243. arXiv: hep-ph/0212264 (2003).

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

[30] Berger, E. L., Collins, J. C., Soper, D. E. & Sterman, G. F. "Diffractive Hard Scattering". *Nucl. Phys. B* **286,** 704–728 (1987).

[31] Streng, K. H. *HARD QCD SCATTERINGS IN DIFFRACTIVE REACTIONS AT HERA* in *DESY Workshop 1987: Physics at HERA* (Jan. 1988).

[32] Buckley, A., Hoeth, H., Lacker, H., Schulz, H. & von Seggern, J. E. "Systematic event generator tuning for the LHC". *Eur. Phys. J. C* **65,** 331–357. arXiv: 0907.2973 [hep-ph] (2010).

[33] CMS, Khachatryan, V. *et al.* "Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 7$ TeV". *Phys. Rev. Lett.* **105,** 022002. arXiv: 1005.3299 [hep-ex] (2010).

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

[35] Bierlich, C., Chakraborty, S., Gustafson, G. & Lönnblad, L. "Strangeness enhancement across collision systems without a plasma". *Phys. Lett. B* **835,** 137571. arXiv: 2205.11170 [hep-ph] (2022).

[36] ATLAS, Aad, G. *et al.* "Measurement of the centrality dependence of the charged-particle pseudorapidity distribution in proton–lead collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV with the ATLAS detector". *Eur. Phys. J. C* **76,** 199. arXiv: 1508.00848 [hep-ex] (2016).

[37] Bierlich, C. *et al.* "Confronting experimental data with heavy-ion models: RIVET for heavy ions". *Eur. Phys. J. C* **80,** 485. arXiv: 2001.10737 [hep-ph] (2020).

[38] ALICE, Abelev, B. B. *et al.* "Multiplicity Dependence of Pion, Kaon, Proton and Lambda Production in p-Pb Collisions at $\sqrt{s_{NN}}$ = 5.02 TeV". *Phys. Lett. B* **728,** 25–38. arXiv: 1307.6796 [nucl-ex] (2014).

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

[40] ATLAS, Aad, G. *et al.* "Measurement of charged-particle spectra in Pb+Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV with the ATLAS detector at the LHC". *JHEP* **09,** 050. arXiv: 1504.04337 [hep-ex] (2015).

[41] Bierlich, C. & Christiansen, J. R. "Effects of color reconnection on hadron flavor observables". *Phys. Rev. D* **92,** 094010. arXiv: 1507.02091 [hep-ph] (2015).

[42] Ortiz Velasquez, A., Christiansen, P., Cuautle Flores, E., Maldonado Cervantes, I. & Paić, G. "Color Reconnection and Flowlike Patterns in *pp* Collisions". *Phys. Rev. Lett.* **111,** 042001. arXiv: 1303.6326 [hep-ph] (2013).

Paper III

III

Baryon Correlations in Pythia

Leif Lönnblad and Harsh Shah

To be submitted to Eur. Phys. J. C
e-print: arXiv:2309.01557 [hep-ph]
LU-TP-23-12, MCNET-23-08

Abstract: We present the results from our investigation of angular correlations between baryons pairs in the PYTHIA event generator. We show how colour reconnection models and hadronization mechanisms influence such angular correlations and in particular address the effect of gluons on the baryon production mechanism in the Lund string fragmentation model. We conclude by discussing the new theoretical ideas in comparison with the ALICE pp collision results for the baryon angular correlations. We propose a hypothesis for suppressing baryons produced in gluon jets and show how that may influence the angular correlations.

1 Introduction

One of the research interests in particle physics is understanding the production mechanism and the spatial distribution of particles produced in high-energy particle collisions. This can be studied in various ways in colliders, *e.g.*, by measuring single particle distributions as a function of one or more variables or looking at correlations between particles, for different species of particles. Using phenomenological models we can then use these measurements to gain theoretical insights into the underlying particle production mechanisms.

In this work, we address a long-standing open question about the angular correlations of pairs of the produced hadrons. A two-particle correlation function provides information regarding the production of another particle near the first particle. It is often studied as a function of relative pseudorapidity ($\Delta\eta$) and azimuthal angle ($\Delta\phi$) between two particles.

Depending on the chosen range of $\Delta\eta$, the angular distribution can be studied for long-range (large $\Delta\eta$) or short-range ($\Delta\eta \sim 0$). The long-range correlations around $\Delta\phi \sim 0$ are known as the near-side "ridge". They are studied extensively in different collision systems like pp, pA, and AA to understand the collective behaviour of the produced particles (see, *e.g.*, [1] and references therein). The correlation function is defined such that it is unity for completely uncorrelated pairs of particles, and any correlation will show up as a larger value, while lower values indicate that there is an anti-correlation.

The short-range ($|\Delta\eta| < 1.3$) two-particle angular correlations were studied by the ALICE experiment in [2, 3] for low transverse momentum ($p_\perp < 2.5$ GeV) hadrons produced in pp collisions at $\sqrt{s} = 7$ GeV. This angular correlation study shows that the identified hadron pairs have different angular distributions depending on the types of hadrons in the pairs. The meson pairs of the same-sign and opposite-sign particles show correlations peak near $\Delta\phi = 0$ and a wide bump near $\Delta\phi = \pi$ (also known as the jet peak and the away-side ridge). On the other hand, baryons behave differently whether the angular distributions are produced for the same-sign or opposite-sign baryon pairs. For the opposite-sign baryon pairs the angular distribution is similar to that of the meson pairs, with a visible peak near $\Delta\phi = 0$ and almost flat distribution around $\Delta\phi = \pi$. For the same-sign baryon pairs, however, there is a clear anti-correlations near $\Delta\phi = 0$ (except for an indication of a tiny peak for $\Delta\phi = 0$), and a broad peak is observed around $\Delta\phi = \pi$.

When comparing the ALICE experiment results with PYTHIA [4] generated events, the angular correlations for the same- and opposite-sign meson pairs are well reproduced, but PYTHIA is not able to reproduce the angular correlations for any of the baryon pairs types. It is also observed that this peculiar behaviour in the baryon sector is independent of the flavours of the baryons in the pairs, hence ruling out that the Fermi-Dirac correlation effects could play a major role. Some suggestions and hypothesises are proposed in [3]. Following these suggestions, recently PYTHIA's hadronization mechanism was studied by a theory

group [5].

It can be noted that one of the heavy-ion collision experiments, STAR, measures anti-correlations around $\Delta\phi = 0$ for $\mathrm{p\bar{p}}$ pairs produced in Au-Au collisions [6]. These results show that, unlike the observed correlations in pp collisions, anti-correlations are observed for $\mathrm{p\bar{p}}$ pairs in heavy-ion collisions. Furthermore, if we look into $\mathrm{e^+e^-}$ collisions, then PYTHIA is able to reproduce the baryon angular correlations [7] in $\mathrm{e^+e^-}$ collisions. These results from different collision systems reflect the non-triviality of the underlying physics of the angular correlations in the baryon sector. Hence we have decided to investigate the discrepancy in the angular correlations for baryons produced in $\mathrm{e^+e^-}$ collisions and in pp collisions. Moreover, we want to identify if any of the event simulation stages have any significant role in the baryon angular correlations.

Phenomenological models like PYTHIA play important roles in our attempts to quantify the initial and final state effects on the observables. PYTHIA is one of the successful general-purpose event generators, which can reproduce a variety of observables in good agreement with the data for $\mathrm{e^+e^-}$ and pp collisions for a wide range of collision energies. The partons are produced during the hard scattering, multiple partons interactions (MPIs) [8], and the parton showers, in stages in PYTHIA. These produced partons are then treated in terms of chains of colour dipoles between them, forming strings. An important feature in hadron collisions is that colour connections between the partons can be re-arranged by a colour reconnection (CR) [8, 9] model. After the CR, the colour singlet strings are hadronized by the Lund string fragmentation model [10] in PYTHIA. All these steps can influence the production rate of different hadrons, and correlations of the hadrons in the simulation results.

For simplicity, we keep our investigation limited to pp collisions in this paper. We first have to understand which new effects appear in the event simulation when we move from $\mathrm{e^+e^-}$ collisions to pp collisions. Since PYTHIA is able to reproduce the angular correlations for the same- and opposite-sign meson pairs fairly well, we do not discuss the mesons' angular correlations in the rest of the paper. Instead, our investigation is focused on the angular correlations of the same- and opposite-sign baryon pairs. We also keep our results limited to protons while discussing various theoretical aspects.

In the following, we will start in section 2 by outlining the main baryon production mechanism in the PYTHIA implementation of the Lund string fragmentation model. Special attention is given to the role of gluons and how they may affect the production of baryons. This is followed by section 3 where we discuss an alternative way of obtaining baryons in PYTHIA using the QCD-inspired colour reconnection model. In section 4 we then discuss final-state effects and how they could affect baryon correlations, with special attention to the hadronic rescattering model. In section 5 we then look at the phenomenology of these models and try to understand better what could cause the anti-correlation between like-

sign baryons as found in data. Finally, in section 6 we summarise with a discussion and an outlook.

2 Baryons, popcorn and gluons in the Lund Model

Throughout the perturbative phase of the generation of an event in PYTHIA, from multiple scatterings, initial- and final-state showers, the tracing of colour connections between partons is done using a leading-colour ($N_C \to \infty$) approximation. In hadronic collisions there is a possibility to rearrange these connections, as described below in section 3, but the end results is in any case in colour-singlet *strings*, each connecting an anti-quark with a quark via a chain of colour-connected gluons. In the Lund string model, these strings are fragmented into hadrons as the string breaks by quark–anti-quark production in the string-like colour field between the partons.

The production rate of different hadron species depends on their quark content, mass, and spin. The quarks and anti-quarks of different flavours are produced in accordance with various parameters in the Lund String Fragmentation mechanism. The values of these parameters are primarily fixed from the model comparisons with LEP data. In a string breaking, a quark and an anti-quark are produced as virtual particles, which can come on-shell using the energy stored in the string through a tunnelling mechanism. Clearly, the production of heavier quarks would then need more of the string energy than light ones and are therefore suppressed.

The sequence of the further string break-ups will decide if the string piece will form a meson or a baryon as a primary hadron. A series of string breaks of multiple $q\bar{q}$ pairs will produce mesons. The simplest model for baryon production assumes that the string may break by the production of a diquark-antidiquark pair. This we call the 'diquark model' in PYTHIA [11]. The consecutive string breaks of $q\bar{q}$ pairs on either side of the diquark and anti-diquark will form a baryon and an anti-baryon.

In the diquark model, the baryon and anti-baryon are always produced next to each other in rank, and therefore close in rapidity. Experimental results show that this is not the case always [12]. A mechanism was developed to add separation between a baryon and an anti-baryon produced next to each other in the same string. It is called a *popcorn mechanism* [13], and adds a possibility of meson production between the baryon and anti-baryon pair. The idea of the popcorn mechanism for baryon production is favoured by the experimental results [12]. At the moment, the popcorn mechanism is enabled by default although only one meson is allowed to form between the baryon and anti-baryon pair in PYTHIA.

With or without popcorn, it is clear that we expect some correlations between baryons and anti-baryons. In particular, if we consider the case where they are produced next to

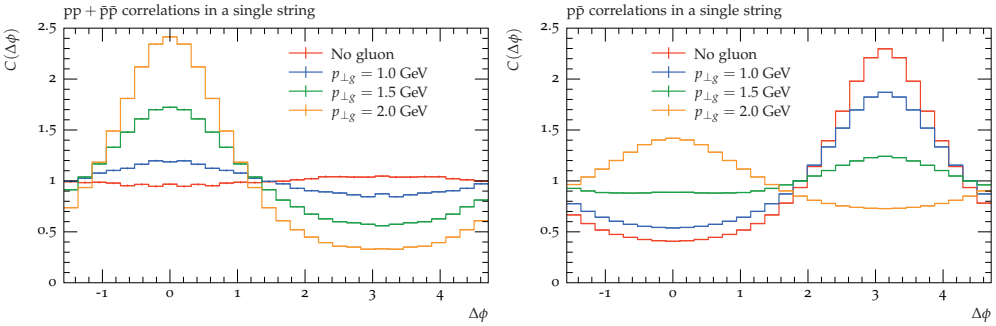


Figure III.1: Azimuthal correlations along a string with or without a soft ($p_\perp = 1.0$, 1.5, and 2.0 GeV) gluon. The left plot shows proton-proton (pp + $\bar{\text{p}}\bar{\text{p}}$) correlations, while the right shows proton–anti-proton correlations. The string is spanned between a quark and an anti-quark with opposite momenta ($p_{q/\bar{q}} = \pm 100$ GeV) along the z-axis and the gluons are placed at $\eta = 0$. Only protons with $|\eta| < 1$ are considered.

each other along the string, their diquark and anti-diquark will have opposite transverse momentum along the string giving an anti-correlation in azimuth angle. However, there is no clear way of obtaining baryon-baryon correlation in the string fragmentation model as such. In a string, there must be at least one anti-baryon between two baryons, and the way transverse momentum is treated in the Lund model, there should be no correlation between them at all.

The MPI machinery for hadronic collisions produces many strings in an event, but they are hadronized independently and would not give rise to correlations between baryons from different strings. It has, however, been shown that the colour reconnection model in PYTHIA gives rise to radial flow [14], which in principle could be responsible for the correlations, and we will discuss that in section 3. Irrespective of colour reconnection it is clear that the strings in hadronic collisions in general are connected to partons from MPI scatterings and are therefore not parallel to the beam axis.

In figure III.1 we show how a jet peak evolves by comparing baryon azimuthal correlations in a single straight string, parallel to the beam axis, with the situation where this string has a (soft) gluon inserted, giving a transverse "kink". For same-sign protons, the straight string has almost no correlations, but already a gluon with $p_\perp = 1$ GeV will give a rather strong correlation. It can be noted that in PYTHIA, around 80% of all hadrons in the central pseudo-rapidity bin come from string pieces connected to a parton with a $p_\perp$ of more than 1 GeV for a 7 TeV pp collision. For $\text{p}\bar{\text{p}}$ we see, as expected, a strong anti-correlation since the di-quark breakup gives opposite transverse momenta for the baryon and anti-baryon. But we see that with a soft gluon, the anti-correlation reduced, and for a 2 GeV gluon it

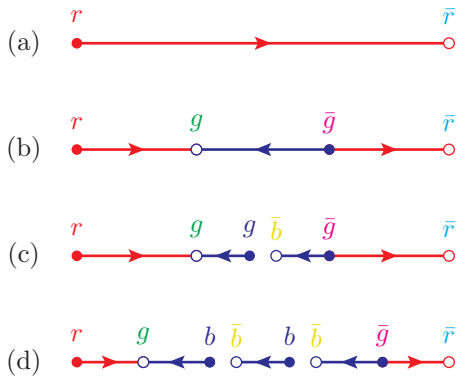


Figure III.2: Illustration of the popcorn mechanism. In (a) no fluctuation has occurred, and a full string is spanned between a red–antired $q\bar{q}$ pair. In (b) a green–antigreen pair has appeared on the string as a quantum fluctuation. If the red and green quarks form an antiblue triplet, this reverses the colour flow in this part of the string, and the net force acting on the green quark is zero. In (c) the string breaks by the production of a blue–antiblue $q\bar{q}$ pair, resulting in two string pieces with diquark ends. In (d) another breakup in the blue triplet field results in an additional meson.

has been turned into a rather strong correlation.

In the MPI machinery, the string with a gluon would be accompanied by another string connected to a gluon going in the opposite azimuth direction. The latter would not be strongly correlated in rapidity, but would give rise to the so-called away-side ridge in a two-particle correlation spectra.

2.1 Gluons vs. popcorn

Since we now have shown that gluon (mini-) jets contribute to the baryon angular correlations, it is relevant to scrutinise the baryon production in a Lund string with gluon "kinks" a bit closer.

The general idea behind the popcorn model used in Pythia is that the creation of a virtual $q\bar{q}$ pair in a string does not necessarily break the string. To do that it has to have the right colours such that the string is divided into two colour singlets. If the colour of the virtual pair does not match the colours of the string ends, the virtual fluctuation can then live for a while before the pair is annihilated again. As an example, in figure III.2 we consider a string stretched between a red quark and an anti-red anti-quark, then imagine a virtual green–anti-green $q\bar{q}$ pair being created where the quark is moving towards the red end, and

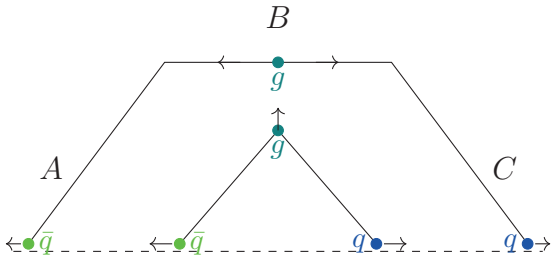


Figure III.3: A schematic diagram shows two different phases of the movement of a $\bar{q}gq$ string, where the initial momentum of the $q(\bar{q})$ is along the (negative) z-axis while the gluon momenta is perpendicular to them. The innermost lines represent the initial phase where a quark and an anti-quark are connected with a gluon kink in between. As the string stretches out and moves, the gluon gradually loses its energy to the string and eventually stops. At this point the string cannot move further upwards, and the gluon kink is basically split into two kinks, and we enter the phase shown with the outermost lines. We thus end up with three pieces of straight string segments, **A**, **B**, and **C**.

vice versa. The field between the virtual quarks will then effectively become antiblue–blue, and if another virtual pair occurs in this region the string can break. With the two quarks moving towards the quark end and two anti-quarks moving towards the anti-quark end. We created two string pieces, each carrying a non-zero baryon number.

For the $q\bar{q}$-fluctuation to live long enough for the string to break in between, the momenta of the q must be longitudinal towards the quark end of the string and vice versa for the $\bar{q}$. Any transverse momenta ($k_\perp$) would be suppressed with a factor $\propto \exp(-\pi k_\perp^2/\kappa)$, where κ is the string tension. Also, if the q and $\bar{q}$ had opposite momenta, the field in between would effectively be between two octet charges, which have more than twice the string tension[1] giving rise to an extra attractive force between the virtual $q\bar{q}$ pair, making long-lived fluctuations heavily suppressed.

This picture works nicely for a straight string. But if there is a gluon along a string, the picture changes. In figure III.3 we show two snapshots of a string stretched between a quark and an anti-quark via a gluon. At first, there are two straight-string pieces (A and C) with the gluon as a kink. But the gluon is here retarded by two string pieces and will eventually stop, resulting in a new straight string piece (B) being formed, and the gluon kink is split into two.

In the current implementation of string fragmentation in PYTHIA, there is no special treatment of baryon production close to such gluon kinks. From the description of the popcorn

[1] see, *e.g.*, [15].

model above, however, it is clear that for a non-breaking virtual $q\bar{q}$ fluctuation, it would be very difficult for the quark or the anti-quark to propagate across such a kink. The pair should have only longitudinal momenta along the string piece where they are created, but the propagation across the kink corresponds to a non-zero transverse momenta in the string piece on the other side of the kink, such fluctuations would be suppressed.

Since we have here shown that gluons are important for the azimuthal correlations between baryons we will in section 5.3 use a toy model to investigate the possible effects of the suppression of baryon production close to gluon kinks.

3 Junctions and colour reconnections

The popcorn and di-quark models are not the only way of obtaining baryons in PYTHIA. In some cases non-trivial colour topologies may arise prior to the string fragmentation stage, *e.g.*, from the treatment of remnants in hadronic collisions, or when looking at baryon number violating BSM processes. In the MPI machinery, it is not uncommon that two (valence) quarks are taken from a proton, leaving a remnant in a colour-triplet state. Similarly, baryon number violating processes may decay a colour-triplet particle into two anti-triplet particles. In both cases, we may obtain colour-singlet string systems connecting three quarks (or three anti-quarks) in a so-called string junction topology [16]. PYTHIA is able to hadronise such systems, in a process that always will produce a net baryon number.

We will not be concerned with BSM here, and the junctions formed in the MPI remnant treatment mainly affects baryons in the far forward or backward regions of rapidity. There is, however, another way of creating junctions available in PYTHIA, using the so-called QCD colour reconnection model [9].

CR models re-arrange the colour connections of the colour dipoles produced after MPIs and parton showers. The primary objective of CR is to reduce the net string length so that the model can reproduce the charged particles' multiplicity and the observed enhancement in $\langle p_\perp \rangle$ (N_{ch}) distribution. PYTHIA has a default CR model, which is based on MPIs [8], where the different MPIs are colour reconnected in the $N_c \to \infty$ limit, and the only criteria to satisfy is to reduce the net string length.

PYTHIA has an alternative model, the QCDCR model [9], which follows QCD colour rules while performing CR. The QCDCR model allows the formation of junction systems, where two or three string pieces can be colour connected to a junction and an anti-junction system, each of which will produce at least one (anti-)baryon. The different colour reconnections possible in this model are summarised in figure III.4. For case (a) two string pieces where the colour end of one is in a colour-singlet state w.r.t. the anti-colour end of the other, can reconnect in a so-called *swing*. In (b) we instead have the situation where the two

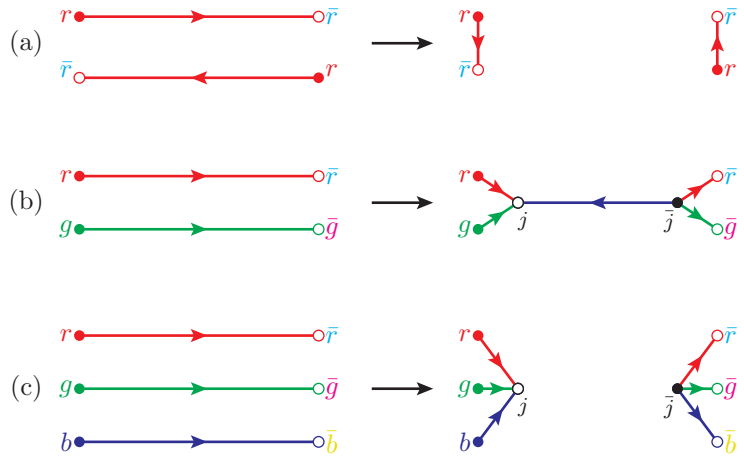


Figure III.4: Illustration of the possible reconnections in the QCDCR model. (a) A "swing" between two dipoles in the same colour state. (b) Two dipoles in different colour states can form a connected junction–anti-junction system. (c) Three dipoles in different colour states form separate junction and anti-junction systems. In all cases, the total string length must be reduced in the process. Note that the dipole ends may be gluons that connect to other dipoles in a string system.

(anti-) colour ends together are in an anti-triplet (triplet) state and can reconnect to two junction systems connected by a dipole. Finally, in (c) the (anti-) colour ends of three dipoles together form a colour singlet and can reconnect into a separate (anti-) junction system. In each case only reconnections that reduce the overall string lengths[2] are allowed. This means that dipoles that are approximately anti-parallel in momentum space are more likely to reconnect like in case (a), while the opposite is true for cases (b) and (c).

The number of junctions in e^+e^- collisions is very low, and it is pointed out in [9] that the effect of the QCDCR model is not clearly visible there. But in pp collisions there are sometimes many MPIs, which enhance the possibility of junction formation during the CR. This means the QCDCR will produce additional baryons, on top of what is produced in the subsequent string fragmentation. It should also be noted that since the dipoles that are reconnected can have a large rapidity span, the correlation between the resulting baryon and anti-baryon is much weaker than for the baryon–anti-baryon pairs produced in the string breaking. We can therefore expect that the correlations will be diluted by the additional baryons from the QCDCR model.

[2]The *length* of a string is approximately given by the sum of the logarithm of the invariant masses of the dipoles forming the string (see ref. [9] for details).

4 Final-state effects on correlations

There are many potential final-state effects that may affect correlations between hadrons produced in the string fragmentation. The Lund group has studied several such models, e.g., a model for Fermi-Dirac correlations [17], the so-called rope hadronisation model [15] and a model for repulsion between strings [18]. Of these, the rope model mainly affects the flavour composition, and is not expected to give significant effects on correlations. Also, the string repulsion will give a flow effect in high multiplicity pp events, but the effect is overall quite small in pp, and it will increase correlations both at $\Delta\phi = 0$ and $\Delta\phi = \pi$ and would therefore not improve the description of baryon-baryon correlations in PYTHIA at small angles. Fermi-Dirac effects would decrease the correlation at small angles for identical baryons, but again the effect is expected to be small[3]. Also, as already pointed out in [5], the effect found in [2, 3] is the same for pp and pΛ, this can also not improve the situation.

Instead, we will focus on the model for hadronic rescattering [19, 20]. By following the production vertices of all partons in the event, it is possible to calculate the production points of all hadrons in the string fragmentation [21]. Then one can study the possible scatterings between these hadrons in a way similar to the UrQMD [22] and SMASH [23] models. Clearly, the rescatterings will mainly affect hadrons that are propagating in the same general direction, and one may expect that it will reduce correlations at $\Delta\phi = 0$, and we will therefor investigate this model in the following section.

5 Comparison with data

So far we have presented a set of ideas that may affect baryon correlations in PYTHIA, and in this section, we will confront these ideas with data. It can be noted that we have also tested varying standard string fragmentation parameters, such as flavour ratios, di-quark production rate, spin ratios of the di-quarks, and $p_\perp$ assignment to the produced hadrons. We found, however, that none of these changes significantly affects the angular correlations of the same-sign baryon pairs in PYTHIA.

In Ref.[5], the main conclusion was that only by forcing PYTHIA to produce at most one di-quark breakup per string, it was possible to understand the correlations found in data. Such an artificial change in the behaviour of the string fragmentation is of course not a satisfactory solution, but it gives us hints as to what is needed. We will therefore concentrate on *reducing* the number of di-quark breakups in strings with (semi-) hard gluons, but also to introduce alternative baryon production mechanisms that do not exhibit the correlations

[3]We have confirmed that the effect is small by making a rudimentary implementation in PYTHIA of the Fermi-Dirac model described in [17]

found in string fragmentation. In addition, we will also consider final state effects from hadronic rescattering.

In all simulations, we have used the PYTHIA version 8.306 to generate pp events at $\sqrt{s} = 7$ TeV. The analysis of the generated events was done using the Rivet [24] routine `ALICE_2016_I1507157` which mimics the analysis in [3].[4]

5.1 The QCD colour reconnection model

We begin with the QCDCR model, which introduces a completely new way of producing baryons. We have used the so-called "mode-0" tune presented in [9], with no further changes. The results are presented in figure III.5 and show a remarkable improvement in the baryon–anti-baryon correlations as compared to the default CR model in PYTHIA.

The choice of the CR model does not, however, improve the angular (anti-) correlations for the same-sign baryon pairs. The effect is rather a general reduction in correlations, which is expected since the model will produce additional baryons with fewer correlations.

The reduction is more clearly seen for the opposite-sign baryon pairs, where QCDCR model reduces the amplitude of the baryon-antibaryon pair correlations near $\Delta\phi = 0$, and also reduces the corresponding away side anti-correlation, bringing the simulation results in agreement with the data. It is clear that the separation between the junction and anti-junction systems created by the QCDCR model plays a significant role in improving the angular correlations between the opposite-sign baryon pairs.

It should be noted that we have also studied the effects in meson correlations (which are not shown here) but found no significant effect of the choice of CR model there.

Since the QCDCR shows significant improvement in the angular correlations of the opposite-sign baryon pairs, we will in the following use the QCDCR as our base-line set-up when adding other modifications.

5.2 Hadronic rescattering

In a high multiplicity event, the produced hadrons can interact with nearby hadrons via elastic or inelastic scattering. A model for hadronic rescattering [19] was recently added in PYTHIA, implementing $2 \to 2$ and $2 \to 3$ type inelastic and elastic hadronic rescatterings.[5]

[4]Note that we have made slight corrections to the kinematical cuts used for different particle species in the Rivet routine, to better reflect the cuts made in the experiment. These corrections will be included in a future release of Rivet.

[5]It should be noted that the absence of $3 \to 2$ scatterings mean that the overall multiplicity of an event is slightly increased. In [19] this was compensated by a slight change in the p_{T0}^{Ref} parameter regulating the

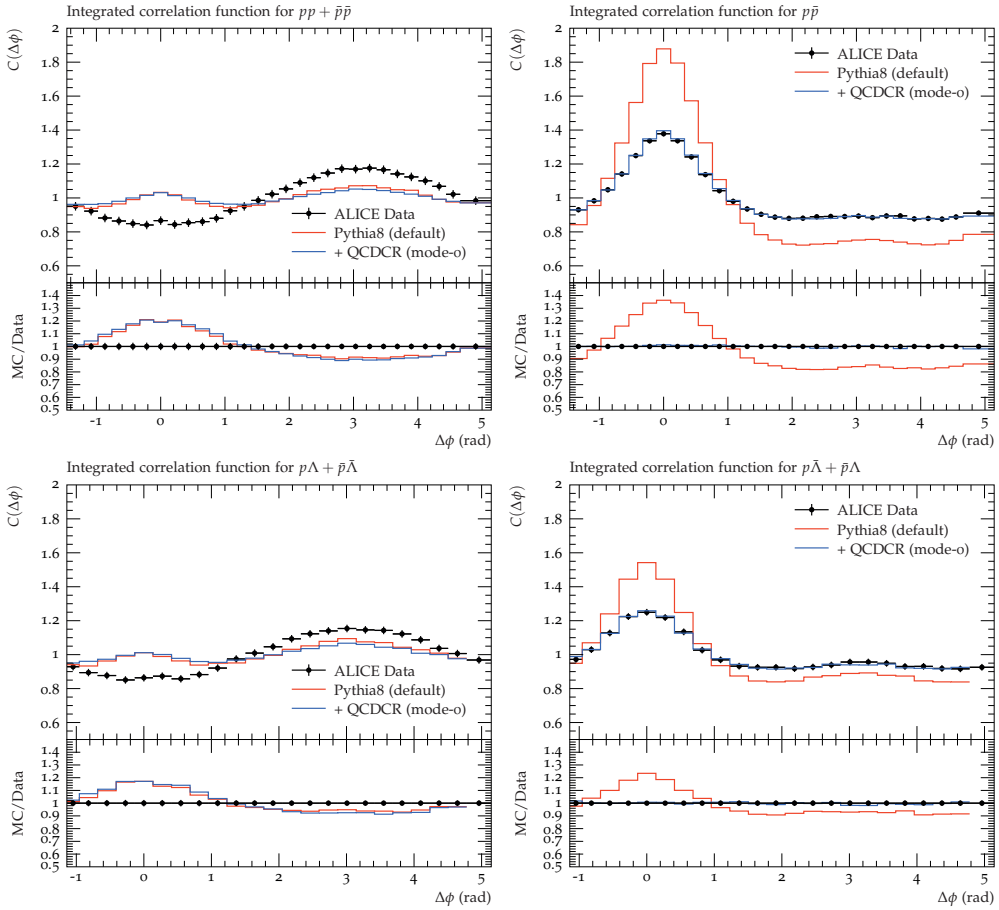


Figure III.5: Baryon azimuthal correlations. *Top*: pp + $\bar{\text{p}}\bar{\text{p}}$ pairs on the left, and p$\bar{\text{p}}$ pairs on the right. *Bottom*: pΛ + $\bar{\text{p}}\bar{\Lambda}$ pairs on the left, and p$\bar{\Lambda}$ + $\bar{\text{p}}\Lambda$ pairs on the right. Events are generated for pp collisions at $\sqrt{s} = 7$ TeV and are compared with ALICE data [3]. The red and blue lines represent results using the PYTHIA Monash tune with MPIs-based colour reconnection, and using QCDCR (mode-0) colour reconnection respectively.

The naive expectation is that rescattering will blur preexistent correlations between particles going in the same direction, and that is indeed what is seen for the p$\bar{\text{p}}$ correlations in figure III.6. The effect is not very large, but we know that rescattering effects in general are quite modest in pp collisions. We note, however, that for the like-sign proton correlations in figure III.6 the effect is much more visible. In fact, the correlation around $\Delta\phi = 0$ is all

divergences in the MPI scattering cross sections. We have checked that a similar increase does not significantly influence the correlation results presented here.

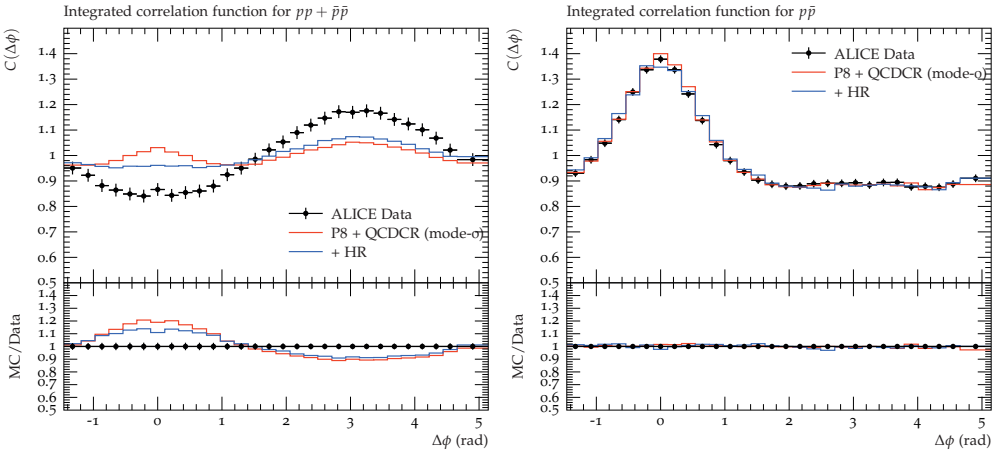


Figure III.6: Proton azimuthal correlations for pp + $\bar{\text{p}}\bar{\text{p}}$ pairs on the left, and p$\bar{\text{p}}$ pairs on the right. Events are generated for pp collisions at $\sqrt{s} = 7$ TeV and are compared with ALICE data [3]. The red lines show results from PYTHIA with QCDCR (mode 0), while blue lines show the same but with hadronic rescattering.

but wiped out.

The reason for this is somewhat non-trivial, and is related to the annihilation of baryons–anti-baryon pairs in the rescattering. As explained in section 2, the peak at $\Delta\phi = \Delta\eta = 0$ mainly comes from jets, where the two particles are typically produced in the same string. For p$\bar{\text{p}}$ pairs the main contribution is pairs produced in a single diquark breakup, and since the diquark and anti-diquark will have opposite transverse momenta along the string, it is very unlikely that the baryons formed would rescatter with each other. For pp and $\bar{\text{p}}\bar{\text{p}}$ pairs, however, we would need two baryon–antibaryon pairs produced close together along the string and a baryon in one pair could then more easily annihilate with the anti-baryon in the other. This effect turns out to be rather large. Adding rescattering to the $p_{\perp g} = 2$ GeV runs in figure III.1 does not affect the shape of the correlations very much, but the number of like-sign pairs will be reduced by around 40%. We, therefore, conclude that the reason for the relatively large effect for pp and $\bar{\text{p}}\bar{\text{p}}$ in figure III.6 is that the number of pairs stemming from the same string is reduced.

5.3 Suppressing baryon production close to gluon kinks

There is currently no proper implementation for the possible suppression of baryon production in string fragmentation close to gluon kinks in the popcorn mechanism discussed in section 2.1. Instead, we will study a simplified toy model to understand what the effects

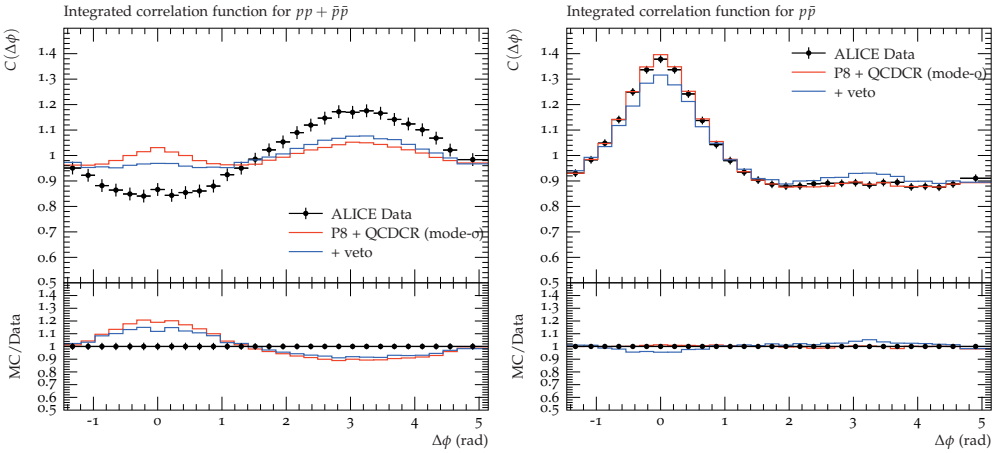


Figure III.7: Proton azimuthal correlations for pp + $\bar{\text{p}}\bar{\text{p}}$ pairs on the left, and p$\bar{\text{p}}$ pairs on the right. Events are generated for pp collisions at $\sqrt{s} = 7$ TeV and are compared with ALICE data [3]. The red lines show results from PYTHIA with QCDCR (mode-0), while blue lines show the same but with a veto on primary baryons spanning a gluon kink as explained in the text.

may be.

We have decided to constrain the baryon production using the `UserHooks` facility in PYTHIA [4], which allows a user to intervene at different stages of the event generation. In particular, there are options to intervene in the string fragmentation procedure and one possibility is to simply veto the production of a single hadron, based on additional criteria implemented by the user.

In our crude implementation, we veto any baryon produced in a diquark breakup if the previous breakup was in a different string region. As an example consider the case in figure III.3 where the gluon has lost all its energy. If there has been a normal $q\bar{q}$ breakup in string region **C** and the next breakup is a diquark–anti-diquark breakup in region **B**, we veto the baryon to be produced, and tell PYTHIA to try another breakup instead. It should be noted that the Lund string fragmentation model is left–right symmetric, and if we go from the other ($\bar{q}$) end and the same diquark breakup occurred in region **B** the following $q\bar{q}$ breakup in **C** producing the same baryon from a *kinky* string piece, is not vetoed. The reason for implementing it in this way is technical, but effectively it will result in a suppression of baryons produced around a string corner with a factor of 0.5.

In figure III.7 see the effect of applying this toy model to PYTHIA including QCDCR. As expected the jet peak for baryon pairs is reduced, and for pp + $\bar{\text{p}}\bar{\text{p}}$ the lines are moved closer to the data. Unfortunately, the reduction of the jet peak is also present for p$\bar{\text{p}}$, which

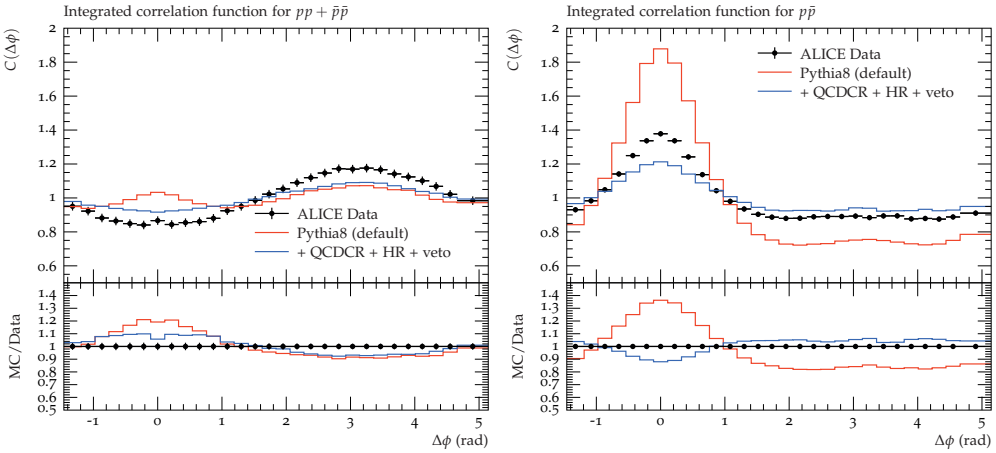


Figure III.8: Proton azimuthal correlations for pp + $\bar{\text{p}}\bar{\text{p}}$ pairs on the left, and p$\bar{\text{p}}$ pairs on the right. Events are generated for pp collisions at $\sqrt{s} = 7$ TeV and are compared with ALICE data [3]. The red lines show results from the default PYTHIA, while blue lines show the result with QCDCR (mode-0) colour reconnection and hadronic rescattering (HR) switched on and with a veto on primary baryons spanning a gluon kink as explained in the text.

worsens somewhat the excellent agreement with data obtained from the QCDCR model.

It should be noted that our toy model will reduce the overall number of baryons produced in general, even in e^+e^- collisions, since also there we have gluon kinks. To be completely fair we should therefore have retuned the parameters affecting baryon production to obtain the same reproduction of LEP data. We would, however, expect the reduction of the jet peak to stay more or less the same.

As a final result, we show in figure III.8 a comparison between the default PYTHIA and the accumulated changes of all three models investigated her: QCDCR, hadronic rescattering and the vetoing of baryons close to gluon kinks. We see that the reproduction of the ALICE date is far from perfect if the models are added, but there is a clear improvement over the default PYTHIA. The jet peak is reduced both for pp +$\bar{\text{p}}\bar{\text{p}}$ and p$\bar{\text{p}}$, and we see that there is even an anti-correlation for like-sign proton pairs around $\Delta\phi = \Delta\eta = 0$.

6 Discussion and summary

We have shown that the observed anti-correlation for the same-sign baryon pairs in the ALICE experiment for pp collisions is a non-trivial outcome of many-fold effects. We also

presented the first steps towards understanding the failure of PYTHIA in reproducing these experimental results. We found that two already existing models in PYTHIA, the QCDCR model and the hadronic rescattering model, have a significant effect on the correlations, and adding these to the default PYTHIA improves the description of data significantly.

The QCDCR model produces additional baryons due to junction systems forming as a part of the reconnections of the colour dipoles. Such junction baryons are much less correlated than those produced in string fragmentation. We show that it visibly reduces the correlations between the opposite-sign baryon pairs in the jet peak near $\Delta\phi = 0$. As a result, PYTHIA is able to reproduce the angular correlation distribution for the opposite-sign baryon pairs.

The anti-correlations in the same-sign baryon pairs are rather complex results. Although the QCDCR model improves the PYTHIA results, it is not sufficient. Adding the hadronic scattering model, we found that the effect of annihilation of baryon–anti-baryon in jets with more than one baryon–anti-baryon pair is quite significant, while if there is only one pair, there is typically no annihilation. This gives a further reduction of the jet peak for same-sign baryons, while the effects on unlike-sign correlations are small. Still, the jet peak for same-sign baryons in PYTHIA needs to be further reduced in order to reproduce data.

The authors in [5] managed to make PYTHIA reproduce data, by forcibly forbidding more than one baryon–anti-baryon pair to be produced in a string. This effectively removed the jet peak in the same-sign baryon correlations, leaving only the anti-correlation in the away-side ridge. Here we instead propose a more physical mechanism, where baryon production close to gluon kinks in a string is suppressed. The motivation for this comes from the popcorn model of baryon production in a string. Here an extra non-breaking virtual $q\bar{q}$ is required to exist before a $q\bar{q}$ pair breaking occurs to produce an effective di-quark breakup, and we argue that it is less likely to have such an extra pair close to a gluon kink.

Since the jet peak around $\Delta\phi = \Delta\eta = 0$ in the angular correlations in pp collisions mainly consist of particle pairs from the same (mini-) jet (which at the LHC is likely to be a gluon jet), one would then expect a reduction of the peak for baryon pairs in general. For same-sign baryon pairs, we expect the reduction to be even bigger since we require two such popcorn breakups in the same string. We have here qualitatively confirmed that this is the case using a toy model, where we simply disallow some such breakups close to gluon kinks, which has motivated us to attempt a more realistic modelling of the effect in the future. Such a model would have to take into account the size of the transverse momentum of the gluon kink, as well as the distance between the breakup and the kink. Since the overall number of baryons would be reduced, such a model would also require a proper retuning of the baryon parameters in PYTHIA, but it is still likely that the jet peak for same-sign baryon correlations would be reduced. Whether it will be reduced enough to reproduce data remains to be seen.

Finally, we note that there are other independent measurements that could verify our hypothesis of suppressed baryon production close to gluons. One obvious example is to compare the baryon-to-meson ratio inside a gluon jet to that of a quark jet, which could be done by comparing inclusive jets to jets produced together with hard photons. We are not aware of any study where the jet substructure has been studied for identified hadrons, but we would certainly like encourage our experimental colleagues to pursue such measurements at the LHC.

Acknowledgements

We would like to thank Gösta Gustafson for coming up with the idea of suppressing popcorn production close to gluons. We would also like to thank Christian Bierlich and Torbjörn Sjöstrand for useful discussions.

This work was funded in part by the Knut and Alice Wallenberg Foundation, contract number 2017.0036, Swedish Research Council, contracts numbers 2016-03291 and 2020-04869, in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 668679, and in part by the MCnetITN3 H2020 Marie Curie Initial Training Network, contract 722104.

References

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

[2] ALICE, Graczykowski, Ł. K. & Janik, M. A. "Angular correlations measured in pp collisions by ALICE at the LHC". *Nucl. Phys. A* **926** (eds Armesto, N., McLerran, L., Pajares, C. & Salgado, C.) 205–212. arXiv: 1401.4306 [hep-ex] (2014).

[3] ALICE, Adam, J. *et al.* "Insight into particle production mechanisms via angular correlations of identified particles in pp collisions at $\sqrt{s} = 7$ TeV". *Eur. Phys. J. C* **77.** [Erratum: Eur.Phys.J.C 79, 998 (2019)], 569. arXiv: 1612.08975 [nucl-ex] (2017).

[4] Bierlich, C. *et al.* "A comprehensive guide to the physics and usage of PYTHIA 8.3". *SciPost Physics CodeBases.* arXiv: 2203.11601 [hep-ph] (Mar. 2022).

[5] Demazure, N., Sebastián, V. G. & Llanes-Estrada, F. J. "Baryon Anticorrelations in PYTHIA". *Few Body Syst.* **64,** 57. arXiv: 2210.02358 [hep-ph] (2023).

[6] STAR, Adam, J. *et al.* "Beam-energy dependence of identified two-particle angular correlations in $\sqrt{s_{NN}}$ = 7.7–200 GeV Au+Au collisions". *Phys. Rev. C* **101,** 014916. arXiv: 1906.09204 [nucl-ex] (2020).

[7] TPC/Two Gamma, Aihara, H. *et al.* "Study of baryon correlations in e^+e^- annihilation at 29-GeV". *Phys. Rev. Lett.* **57,** 3140 (1986).

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

[9] Christiansen, J. R. & Skands, P. Z. "String Formation Beyond Leading Colour". *JHEP* **08,** 003. arXiv: 1505.01681 [hep-ph] (2015).

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

[11] Andersson, B., Gustafson, G., Ingelman, G. & Sjöstrand, T. "Baryon Production in Lepton - Nucleon Scattering and Diquark Fragmentation". *Z. Phys. C* **13,** 361 (1982).

[12] TPC/Two Gamma, Aihara, H. *et al.* "Baryon production in e^+e^- annihilation at $\sqrt{s} = 29$ GeV: clusters, diquarks, popcorn?" *Phys. Rev. Lett.* **55,** 1047 (1985).

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

[14] Ortiz Velasquez, A., Christiansen, P., Cuautle Flores, E., Maldonado Cervantes, I. & Paić, G. "Color Reconnection and Flowlike Patterns in *pp* Collisions". *Phys. Rev. Lett.* **111,** 042001. arXiv: 1303.6326 [hep-ph] (2013).

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

[16] Sjöstrand, T. & Skands, P. Z. "Baryon number violation and string topologies". *Nucl. Phys. B* **659,** 243. arXiv: hep-ph/0212264 (2003).

[17] Duran Delgado, R. M., Gustafson, G. & Lönnblad, L. "String Effects on Fermi-Dirac Correlation Measurements". *Eur. Phys. J. C* **52,** 113–119. arXiv: hep-ph/0702241 (2007).

[18] Bierlich, C., Gustafson, G. & Lönnblad, L. "Collectivity without plasma in hadronic collisions". *Phys. Lett. B* **779,** 58–63. arXiv: 1710.09725 [hep-ph] (2018).

[19] Sjostrand, T. & Utheim, M. *A Framework for Hadronic Rescattering in pp Collisions* 2020. arXiv: hep-ph/2005.05658v1.

[20] Bierlich, C., Sjöstrand, T. & Utheim, M. "Hadronic rescattering in pA and AA collisions". *Eur. Phys. J. A* **57,** 227. arXiv: 2103.09665 [hep-ph] (2021).

[21] Ferreres-Solé, S. & Sjöstrand, T. "The space–time structure of hadronization in the Lund model". *Eur. Phys. J. C* **78,** 983. arXiv: 1808.04619 [hep-ph] (2018).

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

[23] SMASH, Weil, J. *et al.* "Particle production and equilibrium properties within a new hadron transport approach for heavy-ion collisions". *Phys. Rev. C* **94**, 054905. arXiv: 1606.06642 [nucl-th] (2016).

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

Paper IV

IV

The Dynamic Hadronization of Charm Quarks in Heavy-ion Collisions

Christian Bierlich, Gösta Gustafson, Leif Lönnblad, and Harsh Shah

To be submitted to Eur. Phys. J. C
e-print: arXiv:2309.12452 [hep-ph]
LU-TP-23-13, MCNET-23-10

Abstract: The PYTHIA/Angantyr model for heavy ion collisions was recently updated with a mechanism for *global colour reconnection*. The colour reconnection model used is QCD colour algebra inspired, and enhances baryon production due to the formation of string junctions. In this paper we present updates to the junction formation and string fragmentation mechanisms, connected to heavy quark fragmentation. This allows for the simulation of heavy quark fragmentation, using junction formation, in heavy ion collisions. The framework is validated for proton collisions, and we show results for charm baryon production in proton-lead collisions.

1 Introduction

In high-energy particle collisions, hadrons with heavy quark content, are a uniquely versatile probe of fragmentation dynamics. Their defining feature, a charm (c) or bottom (b) flavoured quark, cannot originate from the hadronization process but must be created either in the hard process or in the parton shower, both calculable with perturbative techniques.

As opposed to the even heavier top (t) quark, hadrons containing c- and b–type quark content, are still understood to fragment through the same mechanisms as their light counterparts, the u, d, and s quarks. When comparing experimental data to theory, two quite different (and thus complementary) techniques are used: The factorisation approach and the route taken by Monte Carlo event generators. In the factorisation approach [1, 2], the cross-section is separated into a convolution of three factors: 1) a Parton Distribution Function (PDF) of the incoming hadron, 2) the parton level hard scattering cross-section, where state-of-the-art calculations today are next-to-leading-order (NLO) in the strong coupling (α_s) (see *e.g.*[3–5]) often with next-to-leading-log (NLL) resummation techniques applied as well, such as *e.g.*GM–VFNS [6] or FONLL [7, 8], and finally 3) fragmentation functions, analytical expressions fitted to e^+e^- and epdata [9, 10] giving differential probabilities for the charm quark to fragment to various hadron species. It has been known at least since SPS [11] that the underlying assumption of independent fragmentation does not hold, but it has generally been assumed that universal fragmentation functions can be applied across systems, when studying inclusive quantities, such as total charm hadron yields per event. Recent work by the ALICE collaboration [12–15] has, however, clearly shown that fragmentation functions tuned to e^+e^- and ep, cannot describe the fragmentation of charm into baryons in pp.

In the Monte Carlo event generator approach, as used in *e.g.* PYTHIA [16], PDFs are still used to extract the participating partons from the colliding nucleons. But where the focus in the factorisation approach tends to be more directed towards formal precision in the calculation of the hard scattering, the focus in the Monte Carlo generators is more towards coherent modelling of both perturbative and non-perturbative aspects, such as hadronization. Once the total amount of charm quarks present in the event is determined by means of a leading order calculation, plus parton shower [17, 18], the amount of hadron species, is determined by the dynamical fragmentation model, the Lund string model [19], and its extensions. This makes charm hadrons very well suited for studies of dynamical hadronization models. For charm baryon production in particular, the so-called QCD colour reconnection (CR) model [20] in PYTHIA has gained a lot of attention, due to its ability to correctly reproduce the Λ_c^+ yield and Λ_c^+/D^0 ratio as a function of $p_\perp$ in pp collisions at various collision energies at LHC [12, 14, 15]. However, the predicted production rates of Ξ_c and Ω_c baryons are still undershooting data, even with the QCDCR model [21–23]. Furthermore, the model has, until recently not been usable for heavy ion collisions.

One of the key aspects of the QCDCR model is the formation of junction-like configurations between two or three colour dipoles. These junction systems contribute to baryon production in addition to the baryons produced during the string fragmentation in PYTHIA. We have recently improved the junction fragmentation for the low-energy junction systems and extended the QCDCR model with a spatial constraint [24]. As a result, the QCDCR model can be used as a global CR model for heavy-ion collision simulations in the Angantyr model [24, 25].

In this paper, we further improve the junction formation and fragmentation for the colour dipoles containing heavy quarks. We use pp collisions to validate the framework and show for the first time how pPb collisions generated with Angantyr + QCDCR, give a satisfactory description of Λ_c^+ production. We show results primarily for charm baryons, but a similar outcome can be expected for the bottom quark containing baryons as well. We present the results using the upgraded QCDCR model from [24] with the new changes we have made in this paper.

We first provide an overview of the Angantyr model for heavy-ion event simulation in PYTHIA in the next section. In section 3 we discuss the perturbative production of the charm quarks, and in section 3.2, we show the non-perturbative aspects of the charm hadrons production. We also discuss the changes we have made in junction formation and fragmentation. Finally, results for charm hadron production in pp and pPb are shown in section 4.

2 Heavy ion collisions with the Angantyr model

The Angantyr model [25, 26] is an extension of PYTHIA to simulate heavy-ion collision events without assuming the creation of a Quark–Gluon plasma. It uses a modified Glauber model [27, 28] to obtain the number and types (*e.g.*elastic or inelastic (diffractive or non-diffractive) interactions) of sub-collisions in a heavy-ion collision event. Based on the number and type of sub-collisions, multiple pp-like collisions are generated and stacked together to produce the heavy-ion event.

The arrangement of the nucleons inside a nucleus is obtained using the Woods-Saxon distribution in the GLISSANDO parametrization [29]. When nuclei collide with each other at relativistic energies, they are Lorentz contracted. The wave functions of the nucleons inside the nuclei can be treated as frozen at the time of the collision. This is realized in the so-called Glauber-Gribov [30–32] formalism for nucleon wave-function fluctuations and extended it to include cross-section fluctuations in projectile and target nucleons for pA and AA collisions.

Once the types of nucleon-nucleon (NN) sub-collisions are decided, the Angantyr model

uses the PYTHIA model for multiparton interactions to generate respectively non-diffractive, diffractive, and elastic pp events. Often it occurs that a nucleon is participating in more than one NN non-diffractive sub-collision. A key feature of the model is the special treatment for nucleons participating in multiple non-diffractive interactions. Given a single projectile nucleon interacting with several target nucleons, the NN pair with the smallest impact parameter is denoted the "primary" non-diffractive sub-collision. The others are denoted "secondary". The primary sub-collision is generated as a normal non-diffractive pp collision, whereas the secondaries are generated as a modified single diffractive collision (see section 5 in reference [25] for further explanation). A secondary non-diffractive interaction will be discarded once sufficient energy is no longer available.

There is no interaction between the partons produced in different sub-collisions in the default Angantyr. All multiple sub-collisions are stacked together at the parton level as colour singlet Lund strings. Later, the Lund strings are hadronised and produce a heavy-ion collision event.

Recently we have added a global colour reconnection (CR) in Angantyr [24]. We have extended the QCDCR model [20], by adding a spatial constraint on the colour dipoles to be colour reconnected. We stack the colour dipoles from different sub-collisions and use the spatially constrained QCDCR model such that colour dipoles from nearby sub-collisions can undergo CR. In this work, we continue to use this upgraded Angantyr set-up to simulate pPb collision events.

3 Charm hadron production in PYTHIA

Since the masses of charm (≈1.5 GeV) and bottom (≈4.8 GeV) are large compared to the light quarks, they will never be produced through the tunnelling mechanism by which the string breaks, but only in the hard process and the parton shower. In sub-section 3.1 we will briefly review the PYTHIA formalism for heavy quark production, and in sub-section 3.2 we give an overview of the impact of CR. In the following sub-sections, the modifications relevant to charm production in pPb will be introduced.

3.1 Charm quark in hard process and parton shower

Several different QCD processes in pp collisions in PYTHIA can produce heavy quarks. The leading order (LO) processes like $qq \to Q\bar{Q}$ and $gg \to Q\bar{Q}$ hard scatterings are the primary processes for heavy quarks production in PYTHIA. Another source of heavy quark production is weak decays (Z and $W^{\pm}$ bosons decays), Higgs decay, and top and bottom quark decay, though of those, only the latter contributes in any significant amount when

considering total charm production down to low $p_\perp$. Furthermore, parton showers, where initial or final state partons (mostly gluons) produce the heavy quarks by pair creation, flavour excitation or gluon splitting. This is a significant source of charm production, in addition to that produced in the hard scattering. Furthermore, the "hidden charm" from the PDF of one of the colliding beams, may come on a mass shell due to the scattering. The interaction is like $Qq \to Qq$ or $Qg \to Qg$, but since the Q is not a valence quark it has to be produced in pairs by a gluon splitting.

The LO processes have the matrix elements containing the heavy quark mass. Since quark masses are included, full phase space down to $p_\perp \to 0$ can be populated. For low $p_\perp$ production, however, using the PYTHIA multiparton interaction framework [33], which introduces a general parameter $p_{\perp 0}$, is more suitable, in particular when extending to heavy ion collisions. The heavy quark masses are an important parameter in the perturbative description of their production. In PYTHIA, the default values for the charm and bottom quark masses are set to 1.5 GeV and 4.8 GeV respectively. The masses affect the matrix elements, splitting kernels, and the phase space of the heavy quarks production cross-section. These values are fitted to D-meson production rates. To better fit production rates at LHC, we have in this paper reduced the charm and bottom quark masses to 1.3 GeV and 4.2 GeV respectively.

3.2 Colour reconnection and hadronization

After the multiple parton scatterings and parton showers, outgoing quarks and gluons are connected by strings. We speak of a string connecting a colour and an anti-colour – either quark and anti-quark or through one or more gluon "kinks" – as a chain of colour dipoles. These colour connections are reassigned through colour reconnection (CR) [20, 33] models. The conventional argument, and indeed the logic behind the default CR model in PYTHIA, is that while the parton shower generates a colour configuration in the $N_c \to \infty$ limit, nature has $N_c = 3$. The choice of specific colour connections for a single event is ambiguous and should therefore be corrected. The calculation itself, however, cannot provide any guidance as to how to do the reconnection, and one must resort to models. A common feature is a reduction of the so-called λ-measure, which is an indirect representation of the rapidity span of the colour dipoles, which is again a logarithmic sum of the potential energy of the dipoles, and hence a measure of the number of hadrons produced by the dipoles. Further details in sub-section 3.3.1. The CR in PYTHIA helps to reproduce the charged particle multiplicity and the increase in $\langle p_\perp \rangle$ as a function of (N_{ch}) distribution as observed in the experiments.

The QCDCR model [20] is developed with the idea of applying **SU(3)** colour algebra on non-correlated colour dipoles before calculating the λ-measure for the new configurations of the dipoles. Colour algebra allows the formation of a colour singlet by three colour

string pieces being connected to a "junction" point (see figure IV.1). The "string system contains a junction" (junction system) formed by two or three dipoles is not possible in the earlier case of $N_c \to \infty$ limit. Hence in the QCDCR model, the two and three dipoles can have three string pieces that are colour-connected to a single "junction" point after the CR. A junction system produces at least one baryon per junction during the hadronization. In the QCDCR model, junctions are always produced as junction and anti-junction pairs and conserve the baryon number. These baryons (and anti-baryons) are additional baryons due to QCDCR.

After the CR, the colour singlet Lund strings hadronised by sequential fragmentation. The different flavours of quarks and anti-quarks are produced according to the Lund string model [34]. Parameter values are fixed from the model tuning with LEP data [20, 35, 36].

The sequence of the string breaks decides if a string piece will form a meson or a baryon as a primary hadron. The Lund string fragmenting into $q\bar{q}$ pairs will produce mesons. For baryons production, the string has to break into a diquark–anti-diquark pair, where the consecutive string breaks of a $q\bar{q}$ pair on either side of the diquarks will produce a baryon and an anti-baryon. PYTHIA uses the "popcorn mechanism" [37], which includes a probability for a meson production between the baryon and the anti-baryon, and the results of the "popcorn mechanism" are supported by the experiments [38].

A $q\bar{q}$ pair production rate,

$$\mathrm{d}P \simeq \mathrm{d}^2 p_\perp \exp\left(-\pi m_\perp^2/\kappa\right), \quad m_\perp^2 = p_\perp^2 + m_q^2, \tag{IV.1}$$

where κ is string tension, and $m_\perp$ is transverse mass of the quark with mass m_q and opposite transverse momenta $p_\perp$. The pair production rate is mass-dependent, and it gives an extremely low probability for the production of heavy quarks pair (*e.g.*"charm" pairs) during the string fragmentation. Therefore all of the heavy quarks are produced either in hard scatterings or in parton showers in PYTHIA as mentioned in section 3.1. In this paper, we show results for charm baryons only, so we refer to charm quarks as the heavy quarks for the rest of the paper.

The heavy quarks form mesons or baryons depending on which quarks/diquarks are produced next to them during string fragmentation. The only way the hadronization of the heavy quarks can be influenced is either by modifying fragmentation parameters or by colour reconnection. LEP data constrains the fragmentation parameters, while in the CR we have some freedom in rearranging the colour connections among the partons. Moreover, the QCDCR model allows junction configurations, which contribute to baryon production. In such a case the type of partons attached to the junction legs and the choice of junction fragmentation sequence can influence the baryon production. Moreover, the probability for junction formation increases in a more dense environment like high multiplicity pp or heavy nuclei collisions.

Until now, no special attention was given if a heavy quark is involved during the junction formation or fragmentation. In this paper, we improve the junction treatments if one or more heavy quarks are involved. We discuss these improvements below in detail.

3.3 The role of junctions

We have made several improvements in the QCDCR model in [24]. A crucial addition is the impact parameter-dependent constraint on the colour dipoles to be colour reconnected, which allowed us to have a global CR among the partons produced in different sub-collisions in heavy-ion event simulation in Angantyr. In addition, we implemented a few technical improvements in the hadronisation of junctions.

In PYTHIA, there may occur situations when a string system in an event cannot be hadronised properly. There are several reasons for such failures, and often they involve junctions. If such a failure arises, PYTHIA will throw away the whole event and generate a new one. In pp collisions such failures are typically rare, but in heavy-ion collisions there can be very many strings and the failure rate per event increases. And since the rate is higher for high multiplicity events (many strings), there is a risk that the overall multiplicity distributions may be skewed. With the QCDCR model, the number of junctions increases, which also increases the failure rate, and we found in [24] that the effect on multiplicity distributions was substantial in the heavy-ion collision, and even visible in pp collisions.

A majority of the discarded events are found to have at least one junction system with a very low invariant mass ($\lesssim$1 GeV). We added a "junction collapse" mechanism to hadronize the low-mass junction systems, which were not treated in PYTHIA prior to [24]. This "junction collapse" mechanism produces two hadrons from the junction system. These hadrons can be two baryons or a meson and a baryon depending on the types of partons attached to the end of every leg in a junction system. We have also introduced an additional trial if the string fragmentation fails to hadronise a junction system. In such a scenario, the junction system is fragmented by the special version of the junction collapse procedure.

As we mentioned at the beginning of the paper, the conflicting results from e^+e^- and pp collisions raise questions about the universality of charm hadron production. In PYTHIA, this applies in particular to charm baryons. We have seen that adding QCDCR improves the description for Λ_c, but for heavier charmed baryons there is still a problem. Since the additional junctions from the QCDCR model are responsible for the increased charm baryon rates, we want to look in more detail into how junctions involving charm quarks are handled there, and also how they are treated in the subsequent string fragmentation.

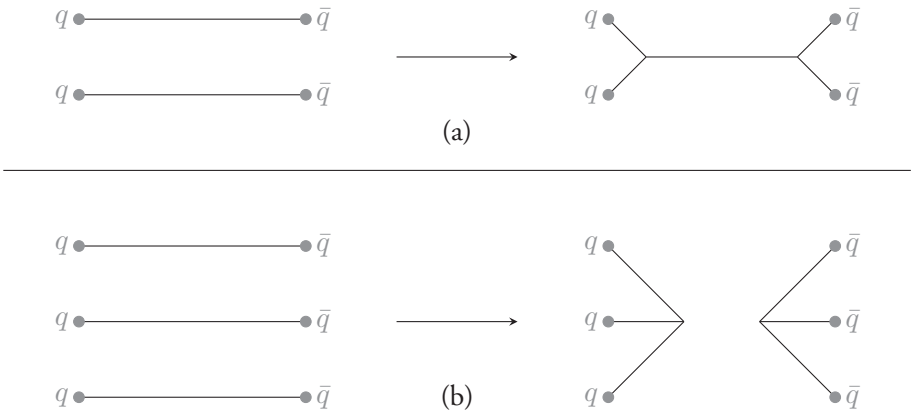


Figure IV.1: Illustration of colour reconnections forming junctions. Two dipoles can form a colour connected junction–anti-junction system (a), and three can form two separate (anti-) junction systems (b).

3.3.1 Junction formation

In pp collisions, junctions are normally only formed in the treatment of the proton remnants, when more than one valence quark undergo scattering in the multiple parton interactions machinery, but such junctions mainly influence baryon production in the forward rapidity region. The junction formation due to colour reconnection is unique to the QCDCR model. The QCD colour algebra-based reconnection treatment includes the colour connections beyond the leading colour approximations. This means that besides the case where two uncorrelated dipoles having the exact same colour state can "swing" so that the coloured parton in one dipole becomes colour connected with the anti-colour of the other, and vice versa, there can also be reconnections between dipoles that have different colour states. In this way, the partons in two or three colour dipoles can become colour-connected to junction points (as shown in figure IV.1) with a certain probability that they are carrying the right colour charges. Each of the junction legs has to have a different colour charge so that the junction system becomes a colour singlet.

After all possible reconnections are tabulated, the QCDCR model will order them so that the reconnection reduces the string lengths the most, as defined by the λ-measure are performed first. For dipoles between two partons the λ-measure in the model is given by[1]

$$\lambda = \ln\left(1 + \frac{\sqrt{2}E_1}{m_0}\right) + \ln\left(1 + \frac{\sqrt{2}E_2}{m_0}\right), \tag{IV.2}$$

[1]The infrared λ-measure for a string in Ref. [39], is in the QCDCR model approximated by the sum of these contributions from the individual dipoles.

where the energies E_i are given in the dipole's rest frame, and m_0 is a tunable parameter. For a dipole connecting a parton to a junction, the model similarly defines the λ-measure as

$$\lambda_j = \ln\left(1 + \frac{\sqrt{2}E}{m_j}\right), \tag{IV.3}$$

where E is the energy of the parton given in the junction rest frame[2] and m_j is a tunable parameter not necessarily the same as m_0.

As discussed in the introduction the λ-measure is an estimate of the rapidity range for the hadrons in the string breakup. The definitions in equations (IV.2 and IV.3) are well motivated for light quarks and massless gluons. However, for a string piece connected to a heavy quark, these expressions are not good estimates of the rapidity range. In this case, we instead use the rapidity of the heavy quark in the rest frame of the junction:

$$\lambda_{HQ} = \frac{1}{2}\log\left(\frac{E+p}{E-p}\right). \tag{IV.4}$$

Here E and p are the energy and momentum of the heavy quark in the junction rest frame. The λ_{HQ} will give a lower value than λ_j for heavy quarks, especially for small p. Hence with this new change, we enhance the possibility for a heavy quark to be part of a junction system during CR in PYTHIA. We note that there is no need for the parameter m_j to set the scale in λ_{HQ} since the quark mass does that for us. Also, the "1+" in the logarithm, which protects the λ from becoming negative is also not needed.

3.3.2 Junction fragmentation

After the Colour Reconnections, the colour strings will undergo string fragmentation. In the QCDCR model in PYTHIA, the three junction legs are treated separately according to the following steps.

- A few attempts are made to move the junction system to the junction rest frame.
- If the algorithm fails to obtain the junction rest frame, then the junction system fragments in the centre of the mass frame of the junction system.
- Once the frame is found, the summed energy of the partons on each junction leg is calculated in that frame and the junction legs are tagged as low-, middle-, and high-energy legs.

[2]In the junction rest frame, the angles between all momenta of the connected partons are 120°.

- The low-energy leg is fragmented first. A fictitious particle is assumed on the opposite side of the junction point for the given junction leg, and the string fragments from the endpoint towards the junction point until a parton closest to the junction point is left on the junction leg.
- Similarly the middle-energy leg is fragmented.
- A diquark is formed by combining the flavour and momenta of the two partons left on the low and middle legs closest to the junction point.
- The diquark is connected at one end of the high-energy leg, the junction no longer exists and the string is fragmented via the usual string fragmentation mechanism.

Finding the junction frame for three *massless partons* is trivial, but as soon as one or more of them are massive, the process does not always converge to a stable solution, because such a solution does not exist. (Also in the case where a junction leg has a long chain of gluons, the proper frame can be difficult to find.)

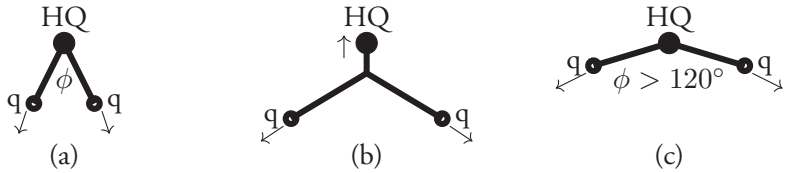


Figure IV.2: (a) A representation of a junction system with a heavy quark (denoted with HQ) and two light partons denoted with (q), (the arrows show their momentum vectors) with an angle ϕ between them, (b) a junction rest frame system for the same configuration, and (c) a scenario where the heavy quark coincides with the junction point and the angle between two light partons is greater than $120°$.

Figure IV.2 (a) illustrates a system with a heavy quark and two light colour charges (quarks or gluons) in the initial rest frame of the heavy quark. If the angle ϕ between the light charges is smaller than $120°$, there is always a frame, in which a junction is at rest as in figure IV.2 (b). The massive quark moves more slowly, and the corresponding string piece is shorter. For $\phi = 120°$ this length goes to zero, and for $\phi > 120°$ the junction coincides with the heavy quark, see figure IV.2 (c). In this case, we find that it is most natural to hadronize the system in the rest frame of the heavy quark. We have implemented this in PYTHIA, and one consequence of this new procedure is that the heavy quark is more likely to be the lowest energy leg, and will in addition not be able to fragment into a heavy meson before being joined into a diquark and then ending up in a baryon. (We note that in this situation the PYTHIA fragmentation system instead hadronizes three strings connected at

the centre in the rest frame of the whole junction system. This reduces the probability of producing a heavy baryon and overestimates the number of produced hadrons.)

We have made one more change in PYTHIA to enhance the chance of a heavy quark ending up in a baryon, and that is to change the ordering of the junction legs. Instead of taking the leg with the lowest summed energy of the connected partons, we use the sum of absolute spatial momentum instead. In analogy with the change in the λ-measure in the QCDCR, this will more closely correspond to how long the actual string is, and will more often put the leg with a heavy quark among the two legs that are fragmented first. Again this improves the chances that the heavy quark ends up in a baryon.

Besides changing the actual algorithms in PYTHIA and in the QCDCR model, we have also investigated some of the parameters that can affect the production of charmed baryons. In the diquark formation by combining the two quarks from the low- and middle-energy legs, the spin assignment is done by a set of parameters[3] suppressing the expected ratio of 3 spin-1 vs. spin-0 states. The default values in PYTHIA are 0.5, 0.7, 0.9, and 1, for the cases where the heaviest quark is u/d, s, c, and b, respectively, but these are not well constrained by experimental data. In the so-called mode-0 tune for the QCDCR model, these were instead all set at 0.0275, which is close to the more well constrained value used for the diquark–anti-diquark breakups in a normal string. There is, however, no reason to expect that these parameters should be the same, since the formation of diquarks in the joining of junction legs is very different from the breakup in strings. And since we know that the QCDCR model has difficulties in describing the production of heavier baryon states, we have checked the effect of raising the values to the default ones in PYTHIA also when using QCDCR.

Since we will here mostly be concerned with charmed baryons that also includes strange quarks, there are also other effects that can influence the production. It is well-known that strangeness enhancement is present not only in heavy-ion collisions but also in high multiplicity pp collisions (see, *e.g.*, [40]). In Lund we have studied the so-called rope hadronisation model [41–43], where overlapping strings gives an increase in the string tension, κ. This results in an increased probability of strange quarks in the string breakups (*c.f.*, eq. IV.1) and the results are promising. Our current implementation does not, however, handle junctions very well, which is why we here have decided to emulate the effect by increasing the overall relative probability of having strange quarks in sting breakups[4] from the default value of 0.217 to 0.4. The number may seem to be high but since most of the charmed baryons are produced at high multiplicities, where there are many dipoles that can reconnect, and hence also many strings can overlap, we do not consider it to be unreasonably high.

[3]The PYTHIA parameter used to set these values is `StringFlav:probQQ1toQQ0join`.

[4]The PYTHIA parameter for this is called `StringFlav:probStoUD`.

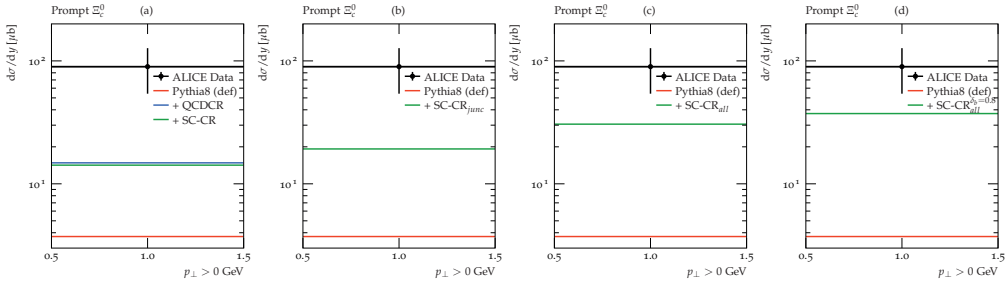


Figure IV.3: Integrated prompt Ξ_c^0 cross-section for pp at $\sqrt{s} = 5.02$ TeV, for $|y| < 0.5$ using different options for PYTHIA compared to ALICE data [22]. In all cases, we show PYTHIA default and, from the left, we show (a) QCDCR (mode-0), and spatially constrained QCDCR (as SC-CR); (b) SC-CR with corrections in the heavy quarks junction formation and fragmentation; (c) SC-CR will all the new changes from this work; and (d) SC-CR (with all changes) with the δ_b parameter set to 0.8 fm.

The overall charm content in an event is mainly governed by perturbative effects, and can be gauged by the rate of the most common charmed D-mesons, which are reasonably well described by the default PYTHIA. With the modifications we have described here, however, a larger fraction of charm quarks will end up in baryons, reducing the rate of D-mesons, and we have therefore decided to compensate for this by increasing the overall charm production by reducing the charm (and bottom) quark mass i PYTHIA from the default value of 1.5 (4.8) GeV to 1.3 (4.2) GeV.

To get an indication of the overall effects of the changes we have proposed here, we show in figure IV.3 the rate of direct production of Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 5.02$ TeV. The model results are compared with the ALICE data [22] using the Rivet [44] routine called `ALICE_2021_I1863039`. In the left-most histogram, we show the results of the default PYTHIA (red line), QCDCR (mode-0) (blue line), and spatially constrained QCDCR (green line). Here we see clearly the effect of introducing the junction reconnections in QCDCR. Our spatially constrained version of the QCDCR gives a slightly reduced rate, mainly because of the constraint, but also because of differently tuned parameters (see [24] for details). In the second to the left histogram, we show the effect of the changes in junction formation and fragmentation for the SC-CR case, and find an increase of around 35%. In the third histogram, we have also added the parameter changes described above and found an additional increase of almost 60%, giving an almost doubled rate compared to the default SC-CR, and and a factor 8 more than the default PYTHIA. We are, however, still far away from the lower bound of the experimental error bar, and a factor almost three below the central value.

Finally, in the right-most histogram of figure IV.3, we show that if we allow reconnection

of dipoles farther separated in the transverse plane by increasing the spatial constraint (δ_b) value in the SC-CR model from 0.5 fm to 0.8 fm on top of the other changes we have made, then we can further enhance the Ξ_c^0 baryon's production in pp collision events. However, since one of the aims of this paper is to compare to pPb data using the Angantyr model we will in the following keep the tuned value of 0.5 fm, which we have shown in [24] gives a more reasonable description of multiplicities in pPb.

4 Results

In this section, we want to look more in detail at the effects of the changes we made. We will concentrate on the charmed baryons, but will also look at non-charmed hyperons. We first look at pp collisions to check that we get reasonable results there before we extrapolate the models to pPb collisions using Angantyr.

4.1 Hyperon production in pp collisions

Since we have forcibly increased the overall strangeness rate in the PYTHIA string fragmentation, it is important to check that what we have done is not unreasonable. In figure IV.4 we therefore show $p_\perp$ distribution for Σ^+, $\bar{\Sigma}^-$, and $(\Xi^0 + \bar{\Xi}^0)/2$ baryons respectively. The ALICE experiment [45] results for pp collisions at $\sqrt{s} = 7$ TeV are used here[5]. The measurements are reported for inelastic collisions and for the particles in the mid-rapidity region ($|y| < 0.5$).

Comparing the default PYTHIA with and without QCDCR, it is clear that the junction reconnections do not contribute much to strange baryon. Instead, the main production mechanism is diquark breakups in the string fragmentation. We can therefore conclude that the main effect when looking at the changes we have done here is the enhancement of strange (di-)quarks in the string breakups. It can be argued that our enhancement is a bit high, but it is clearly not completely unreasonable.

4.2 Charmed baryon production in pp collisions

We now turn to the charmed baryons and will start with Λ_c, where we know that the QCDCR model does a reasonable job. Looking back at figure IV.3, we see that our changes increase the Ξ_c^0 rate substantially, and one can fear that this is compensated by a decrease of Λ_c.

[5]The plots were generated using the `ALICE_2014_I1300380` routine in Rivet.

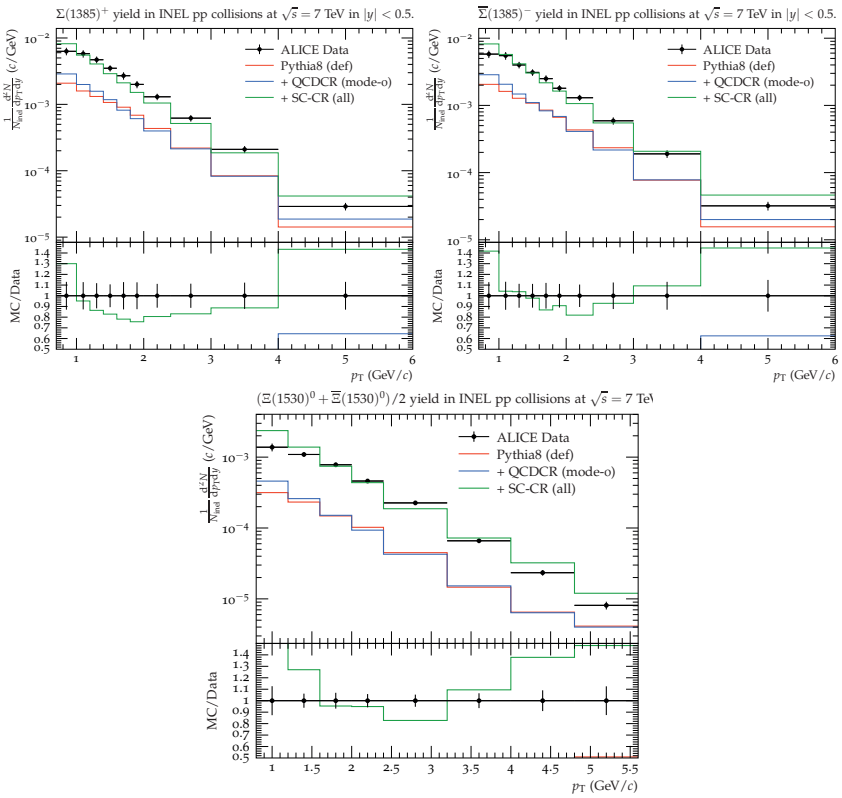


Figure IV.4: *Top*: $p_\perp$ differential yield of Σ^+ (left) and $\bar{\Sigma}^-$ (right). *Bottom*: $(\Xi^0 + \bar{\Xi}^0)/2$ yield as a function of $p_\perp$. The results are from the ALICE experiment [45] for pp collisions at 7 TeV and for mid-rapidity ($|y| < 0.5$). The experimental results are compared with PYTHIA default, QCDCR (mode-0), and spatially constrained QCDCR with all the new changes from this work, which are shown as red, blue, and green lines respectively.

To check this, we show in figure IV.5 the $p_\perp$ distribution of the prompt Λ_c^+ baryons and Λ_c^+/D^0 ratio is compared with the ALICE data [12] (using the same rivet routine as in figure IV.3). Clearly, we maintain a good description of the Λ_c^+ cross section and Λ_c^+/D^0 ratio, even after all the new changes we have introduced in this work.

We note that the QCDCR, both with and without our changes, gives more enhancement for Λ_c^+ for low $p_\perp$, as seen both for the yield and for the ratio to the D^0 yield. The reason for this is that most strings in an event are fairly parallel to the beam, connecting low-$p_\perp$ partons produced by MPI. So the largest chance to get baryons from junction reconnections is from two or three dipoles from such strings along the beam direction, which then results in low-$p_\perp$ baryons.

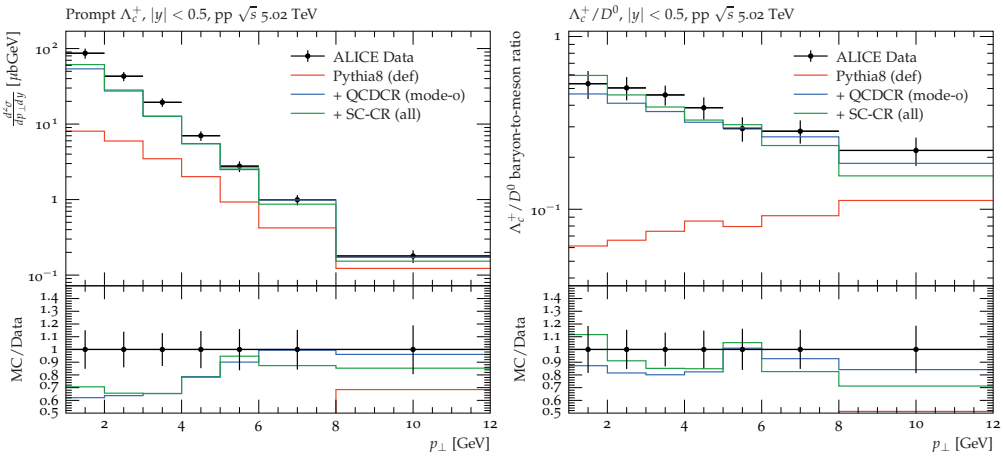


Figure IV.5: Prompt Λ_c^+ distribution as function of $p_\perp$ on the left. The baryon-to-meson ratio for Λ_c^+/D^0 as a function of $p_\perp$ on the right. The pp collision at 5.02 TeV results are from the ALICE experiment [22]. The coloured lines represent the same setups as in figure IV.4.

The effect is less visible for the strange baryons in figure IV.4 since the relative contribution from junction reconnection is smaller but it is still reflected in a small increase of small $p_\perp$ for QCDCR.

In figure IV.6, we then show the corresponding comparison for $p_\perp$ distribution of the prompt Ξ_c^0 baryons and the Ξ_c^0/D^0 ratio results obtained at the ALICE experiment [22] for pp collisions at $\sqrt{s}$=5.02 TeV. The cross section distribution basically shows the same thing that we previously showed in figure IV.3, where the overall yield for PYTHIA is far below the data while adding QCDCR brings it closer, and with our changes even more so.

The Ξ_c^0/D^0 ratio is arguably more relevant for assessing our changes, since the overall (perturbatively modelled) charm rate is factored out, and only the change in the non-perturbative modelling is important. Both for the $p_\perp$ distribution and the integrated ratio our changes actually come quite close to the data (note that there is a linear scale for the ratios here). We note that for the $p_\perp$ shape, the data has a tendency to decrease a bit for the lowest $p_\perp$ bin, while the QCDCR model, with and without our changes, seems to continue to rise, mirroring the behaviour in the figure IV.5.

Last year, the ALICE collaboration presented results [14] [6] also for Σ_c baryons in pp collision, this time using data from the LHC run 2 at $\sqrt{s}$= 13 TeV. Figure IV.7 shows a differential production cross-section for Λ_c^+ on the left and $\Sigma_c^{0,+,++}$ on the right as a function of $p_\perp$, and in figure IV.8 the same is shown as a ratio to the D^0 cross section. We

[6] The analysis is implemented in the Rivet routine `ALICE_2022_I1868463`.

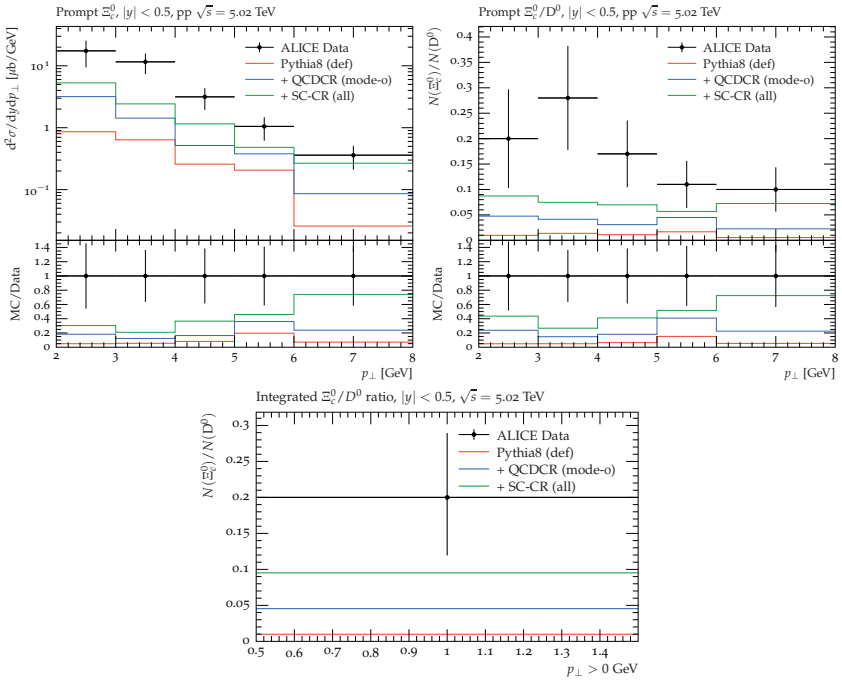


Figure IV.6: *Top Left*: The prompt Ξ_c^0 cross section as a function of $p_\perp$. *Top Right*: The Ξ_c^0/D^0 ratio as a function of $p_\perp$. *Bottom*: The integrated Ξ_c^0/D^0 ratio for all $p_\perp > 0$. The data is from ALICE experiment [22] pp collisions at $\sqrt{s} = 5.02$ TeV. The coloured lines represent the same setups as in figure IV.4.

can clearly see that the modification of the QCDCR model done in this paper not only maintains the Λ_c^+ description but also controls the $\Sigma_c^{0,+,++}$ production rate in PYTHIA. Finally in figure IV.9, we show that due to the reduced $\Sigma_c^{0,+,++}$ production cross-section, the fraction of Λ_c^+ coming from $\Sigma_c^{0,+,++}$ decays, and the ratio to the inclusive Λ_c^+ both are improved by our modifications to the QCDCR model.

From our changes to the QCDCR, the one mainly influencing the Σ_c rate is the change in the parameter controlling the diquark formation in the joining of the smallest junction legs in the fragmentation (see section 3.3.2). Increasing the probability for a charmed diquark to be in a spin-1 rather than a spin-0 state, means that $\Sigma_c^\star$ states are favoured over the Σ_c ones in the subsequent fragmentation of the largest leg. As mentioned in section 3.3.2 these parameters were previously completely unconstrained by data and in [16], the authors described the chosen default values as guesswork. In QCDCR (mode-0) the values were set to the same, rather low, value for all quark types, but in our change, we decided to keep the default ones which are higher and dependent on the heavy quark mass. That the probability should be mass dependent is reasonable since the mass splitting between

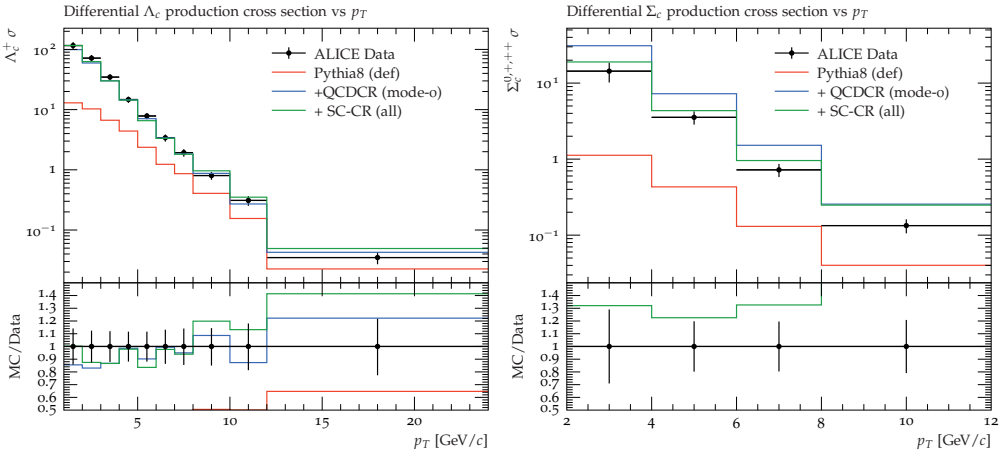


Figure IV.7: $p_\perp$ differential production cross section of Λ_c^+ on the left and of $\Sigma_c^{0,+,++}$ on the right. The data is from the ALICE experiment for pp collisions at 13 TeV [14]. The coloured lines represent the same setups as in figure IV.4.

the spin-1 and spin-0 state should be smaller when heavier quarks are involved. (See, *e.g.*, [46] for a discussion on this). Thanks to ALICE we now have data [14] that can actually constrain this parameter. Here also we notice that the Λ_c^+/D^0 ratio for low $p_\perp$ is increased.

4.3 pPb collisions

With a reasonable description of charmed baryon production in pp collisions, we can now use the Angantyr model to extrapolate our results in heavy-ion collisions. We have in a previous publication shown that, by introducing a cut in the transverse distance between dipole below which we allow them to reconnect in QCDCR. We can allow for a global colour reconnection between sub-collisions in heavy-ion collisions, and still retain a reasonable description of hadron multiplicities. Since the charmed baryon production has been shown to be a sensitive probe into how the junction reconnections in the model behave, we can now see in more detail, if our extrapolation to heavy-ion collisions is reasonable.

In figure IV.10 we show pPb results from the ALICE experiment at $\sqrt{s_{NN}}$= 5.02 TeV for the Λ_c^+ cross section, and the ratio of this w.r.t. D^0 cross section, in comparison with Angantyr model. As expected, the default Angantyr, with colour reconnections only within each sub-collision separately, severely underestimated the rate of Λ_c^+. Adding our spatially constrained version of the QCDCR model improves the description of data significantly, although the Λ_c cross section is still somewhat underestimated. Adding the changes introduced in this paper, however, does not influence the result much. This was not to be expected, since also in pp the effect on Λ_c^+ was minor.

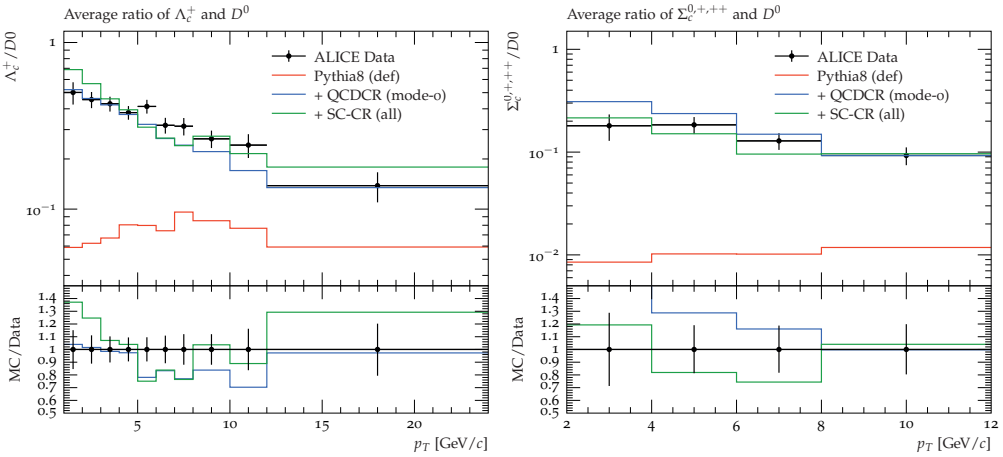


Figure IV.8: Baryon-to-meson ration for Λ_c^+/D^0 on the left and $\Sigma_c^{0,+,++}/D^0$ on the right. The data is from the ALICE experiment for pp collisions at 13 TeV [14]. The red, blue, and green lines are the same as in the figure IV.7. The coloured lines represent the same setups as in figure IV.4.

We can see that for low $p_\perp$ the model fails to reproduce the behaviour of the data. This is best seen in the ratio to D^0, where our model completely shows no sign of reduction of the ratio at small $p_\perp$. Also this could be expected, as we had also seen indications of this in pp collisions above.

5 Conclusion

This paper has shown, for the first time, the effect of applying a modern colour reconnection model to a heavy ion collision, in order to better describe baryon yields. We have shown that the production rates of Λ_c^+ are dramatically improved in pPb collisions using the QCDCR model, which has previously worked well in pp collisions. We also show that the diquark formation in the joining of the junction legs influences the spin-dependent baryon production, and we require experimental data similar to ALICE [14] to constrain the parameter in PYTHIA.

Heavy quarks can only be produced in hard scattering or in a parton shower mechanism in PYTHIA. We show that the application of colour algebra in the QCDCR model allows junction formation by connecting three colour dipoles in a junction point. These junctions contribute significantly to baryon production.

We show that for a heavy quark connected to a junction the λ-measure used in the QCDCR

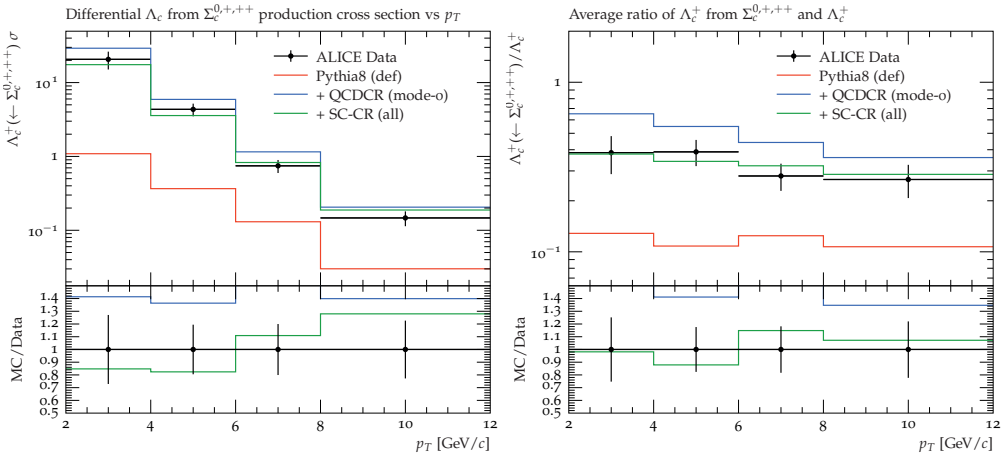


Figure IV.9: *Left*: $p_\perp$ differential production cross-section of Λ_c^+ from $\Sigma_c^{0,+,++}$ decays. *Right*: Ration of Λ_c^+ from $\Sigma_c^{0,+,++}$ decays to the total Λ_c^+ as a function of $p_\perp$. The data is from the ALICE experiment for pp collisions at 13 TeV [14].

model should be improved. Usually, the λ-measure calculates the logarithm of the energy of the dipole in the junction rest frame. But if the dipole contains a heavy quark then often the invariant mass of the quark has a non-negligible contribution to the energy of the dipole. Therefore the rapidity span of the heavy quark from the junction point in the junction rest frame should be used as the λ-measure for such a dipole.

Moreover, when a heavy quark dipole is directly connected to the junction point, the system often fails to obtain a junction rest frame. If the momentum of the heavy quark is low, it is possible that the string piece between the junction and the heavy quark collapses to zero. Thus the heavy quark is directly connected by two strings to the lighter quarks. We show that under such a scenario fragmenting the junction system in the rest frame of the heaviest quark is a good choice.

During the fragmentation of the junction system, the convention is to calculate the energy in every leg and start fragmenting the junction system from the lowest energy leg. Here again, we show that the choice of the new λ-measure should be the scalar value of the momentum instead of the energy because we should avoid counting the invariant mass of the quarks as the potential energy available in the junction leg.

We notice that apart from the two modifications in the junction formation during CR and junction fragmentation during hadronization, we need strange quarks as many of the heavy baryons contain strange quarks. We have a rope hadronization model, which contributes to the strangeness enhancement in PYTHIA. The junction topologies are complex and the string-string interactions in rope hadronization haven't been implemented for junc-

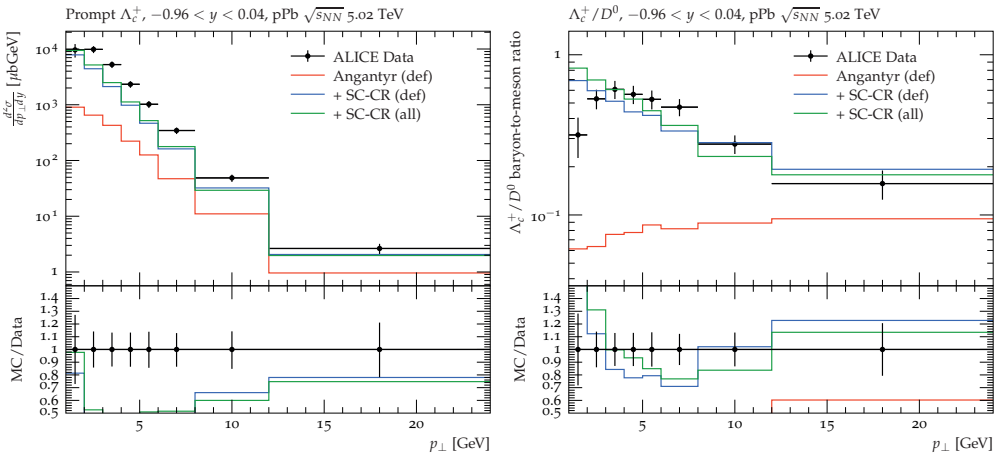


Figure IV.10: Prompt Λ_c^+ distribution as a function of $p_\perp$ on the left. The baryon-to-meson ratio, Λ_c^+/D^0, as a function of $p_\perp$ on the right. The red, blue, and green lines are Angantyr default, Angantyr new tune with global CR as SC-CR (def), and the changes we have made in this work in the global CR as SC-CR (all) respectively. The pPb collision at 5.02 TeV results are from the ALICE experiment [12]. (Note that the rapidity region, $-0.96 < y < 0.04$ is given in the collision rest frame, and corresponds to the central, $|\eta| < 0.5$ region in the laboratory frame.)

tion configurations. Hence we have compensated it by increasing the string fragmentation probability for the strange quarks.

The charm and bottom quark masses are the other parameters we changed in this paper. To enhance the charm and bottom quark production in the first place we decided to use slightly lower mass values within the proposed mass ranges for the respective quarks.

All these changes together helped us to improve the PYTHIA description for Ξ, Σ, Ξ_c, and Σ_c baryon production rates. We also managed to keep a good description of the Λ_c^+ and Λ_c^+/D^0 for different collision energies in pp collisions. For the first time, we show the Λ_c^+ and Λ_c^+/D^0 results in pPb collisions. The results are generated with the global CR in Angantyr and with the changes we introduced in this paper, and they show a visible improvement over the default Angantyr setup.

At this stage, it is also important to note that the increased strange quark fragmentation should be replaced with the appropriate treatment of the rope hadronization model. We may be required to retune some of the parameters because the production of light baryon and other hyperons has not been tracked against the new changes we have made in this work. Moreover, so far we have applied the λ-measure correction only to the junction

formation, but a similar correction should also be applied to the "swing" CR between two dipoles. We hope that the complete treatment of λ-measure correction will affect the Quarkonium production. Hence exploring the possibility of reproducing Quarkonia suppression in heavy-ion collisions is one of the tasks for future work.

Acknowledgements

We would like to thank Torbjörn Sjöstrand for useful discussions.

This work was funded in part by the Knut and Alice Wallenberg Foundation, contract number 2017.0036, Swedish Research Council, contracts numbers 2016-03291 and 2020-04869, in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 668679, and in part by the MCnetITN3 H2020 Marie Curie Initial Training Network, contract 722104.

References

[1] Collins, J. C., Soper, D. E. & Sterman, G. F. "Heavy Particle Production in High-Energy Hadron Collisions". *Nucl. Phys. B* **263,** 37 (1986).

[2] Collins, J. C., Soper, D. E. & Sterman, G. F. "Factorization of Hard Processes in QCD". *Adv. Ser. Direct. High Energy Phys.* **5,** 1–91. arXiv: hep-ph/0409313 (1989).

[3] Nason, P., Dawson, S. & Ellis, R. K. "The Total Cross-Section for the Production of Heavy Quarks in Hadronic Collisions". *Nucl. Phys. B* **303,** 607–633 (1988).

[4] Nason, P., Dawson, S. & Ellis, R. K. "The One Particle Inclusive Differential Cross-Section for Heavy Quark Production in Hadronic Collisions". *Nucl. Phys. B* **327.** [Erratum: Nucl.Phys.B 335, 260–260 (1990)], 49–92 (1989).

[5] Mangano, M. L., Nason, P. & Ridolfi, G. "Heavy quark correlations in hadron collisions at next-to-leading order". *Nucl. Phys. B* **373,** 295–345 (1992).

[6] Helenius, I. & Paukkunen, H. "Revisiting the D-meson hadroproduction in general-mass variable flavour number scheme". *JHEP* **05,** 196. arXiv: 1804.03557 [hep-ph] (2018).

[7] Cacciari, M., Greco, M. & Nason, P. "The P(T) spectrum in heavy flavor hadroproduction". *JHEP* **05,** 007. arXiv: hep-ph/9803400 (1998).

[8] Cacciari, M. *et al.* "Theoretical predictions for charm and bottom production at the LHC". *JHEP* **10,** 137. arXiv: 1205.6344 [hep-ph] (2012).

[9] Belle, Seuster, R. *et al.* "Charm hadrons from fragmentation and B decays in e+ e- annihilation at s**(1/2) = 10.6-GeV". *Phys. Rev. D* **73,** 032002. arXiv: hep-ex/0506068 (2006).

[10] Kneesch, T., Kniehl, B. A., Kramer, G. & Schienbein, I. "Charmed-meson fragmentation functions with finite-mass corrections". *Nucl. Phys. B* **799,** 34–59. arXiv: 0712.0481 [hep-ph] (2008).

[11] WA82, Adamovich, M. *et al.* "Study of D+ and D- Feynman's x distributions in pi-nucleus interactions at the SPS". *Phys. Lett. B* **305,** 402–406 (1993).

[12] ALICE, Acharya, S. *et al.* "Λ_c^+ Production and Baryon-to-Meson Ratios in pp and p-Pb Collisions at $\sqrt{s_{NN}}$=5.02 TeV at the LHC". *Phys. Rev. Lett.* **127,** 202301. arXiv: 2011.06078 [nucl-ex] (2021).

[13] ALICE, Acharya, S. *et al.* "Charm-quark fragmentation fractions and production cross section at midrapidity in pp collisions at the LHC". *Phys. Rev. D* **105,** L011103. arXiv: 2105.06335 [nucl-ex] (2022).

[14] ALICE, Acharya, S. *et al.* "Measurement of Prompt D^0, Λ_c^+, and $\Sigma_c^{0,++}(2455)$ Production in Proton–Proton Collisions at $\sqrt{s}$ = 13 TeV". *Phys. Rev. Lett.* **128,** 012001. arXiv: 2106.08278 [hep-ex] (2022).

[15] "Exploring the non-universality of charm hadronisation through the measurement of the fraction of jet longitudinal momentum carried by Λ_c^+ baryons in pp collisions". arXiv: 2301.13798 [nucl-ex] (Jan. 2023).

[16] Bierlich, C. *et al.* "A comprehensive guide to the physics and usage of PYTHIA 8.3". *SciPost Physics CodeBases.* arXiv: 2203.11601 [hep-ph] (Mar. 2022).

[17] Norrbin, E. & Sjostrand, T. "Production mechanisms of charm hadrons in the string model". *Phys. Lett. B* **442,** 407–416. arXiv: hep-ph/9809266 (1998).

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

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

[20] Christiansen, J. R. & Skands, P. Z. "String Formation Beyond Leading Colour". *JHEP* **08,** 003. arXiv: 1505.01681 [hep-ph] (2015).

[21] ALICE, Acharya, S. *et al.* "Measurement of the Cross Sections of Ξ_c^0 and Ξ_c^+ Baryons and of the Branching-Fraction Ratio BR($\Xi_c^0 \to \Xi^- e^+ \nu_e$)/BR($\Xi_c^0 \to \Xi^- \pi^+$) in pp collisions at 13 TeV". *Phys. Rev. Lett.* **127,** 272001. arXiv: 2105.05187 [nucl-ex] (2021).

[22] ALICE, Acharya, S. *et al.* "Measurement of the production cross section of prompt Ξ_c^0 baryons at midrapidity in pp collisions at $\sqrt{s}$ = 5.02 TeV". *JHEP* **10,** 159. arXiv: 2105.05616 [nucl-ex] (2021).

[23] "First measurement of Ω_c^0 production in pp collisions at $\sqrt{s} = 13$ TeV". arXiv: 2205.13993 [nucl-ex] (May 2022).

[24] Lönnblad, L. & Shah, H. "A spatially constrained QCD colour reconnection in pp, pA, and AA collisions in the Pythia8/Angantyr model". *Eur. Phys. J. C* **83.** [Erratum: Eur.Phys.J.C 83, 639 (2023)], 575. arXiv: 2303.11747 [hep-ph] (2023).

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

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

[27] Glauber, R. J. "Cross-sections in deuterium at high-energies". *Phys. Rev.* **100,** 242–248 (1955).

[28] Miller, M. L., Reygers, K., Sanders, S. J. & Steinberg, P. "Glauber modeling in high energy nuclear collisions". *Ann. Rev. Nucl. Part. Sci.* **57,** 205–243. arXiv: nucl-ex/0701025 [nucl-ex] (2007).

[29] Bożek, P., Broniowski, W., Rybczynski, M. & Stefanek, G. "GLISSANDO 3: GLauber Initial-State Simulation AND mOre..., ver. 3". *Comput. Phys. Commun.* **245,** 106850. arXiv: 1901.04484 [nucl-th] (2019).

[30] Heiselberg, H., Baym, G., Blaettel, B., Frankfurt, L. L. & Strikman, M. "Color transparency, color opacity, and fluctuations in nuclear collisions". *Phys. Rev. Lett.* **67,** 2946–2949 (1991).

[31] Blaettel, B., Baym, G., Frankfurt, L. L., Heiselberg, H. & Strikman, M. "Hadronic cross-section fluctuations". *Phys. Rev.* **D47,** 2761–2772 (1993).

[32] Alvioli, M. & Strikman, M. "Color fluctuation effects in proton-nucleus collisions". *Phys. Lett.* **B722,** 347–354. arXiv: 1301.0728 [hep-ph] (2013).

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

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

[35] Skands, P., Carrazza, S. & Rojo, J. "Tuning PYTHIA 8.1: the Monash 2013 Tune". *Eur. Phys. J. C* **74,** 3024. arXiv: 1404.5630 [hep-ph] (2014).

[36] Bierlich, C. & Christiansen, J. R. "Effects of color reconnection on hadron flavor observables". *Phys. Rev.* **D92,** 094010. arXiv: 1507.02091 [hep-ph] (2015).

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

[38] TPC/Two Gamma, Aihara, H. *et al.* "Baryon production in e^+e^- annihilation at $\sqrt{s} = 29$ GeV: clusters, diquarks, popcorn?" *Phys. Rev. Lett.* **55,** 1047 (1985).

[39] Andersson, B., Dahlkvist, P. & Gustafson, G. "AN INFRARED STABLE MULTIPLICITY MEASURE ON QCD PARTON STATES". *Phys. Lett. B* **214,** 604–608 (1988).

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

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

[42] Bierlich, C., Chakraborty, S., Gustafson, G. & Lönnblad, L. "Jet modifications from colour rope formation in dense systems of non-parallel strings". *SciPost Phys.* **13,** 023. arXiv: 2202.12783 [hep-ph] (2022).

[43] Bierlich, C., Chakraborty, S., Gustafson, G. & Lönnblad, L. "Strangeness enhancement across collision systems without a plasma". *Phys. Lett. B* **835,** 137571. arXiv: 2205.11170 [hep-ph] (2022).

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

[45] ALICE, Abelev, B. B. *et al.* "Production of $\Sigma(1385)^{\pm}$ and $\Xi(1530)^0$ in proton-proton collisions at $\sqrt{s} = 7$ TeV". *Eur. Phys. J. C* **75,** 1. arXiv: 1406.3206 [nucl-ex] (2015).

[46] Bierlich, C., Chakraborty, S., Gustafson, G. & Lönnblad, L. "Hyperfine splitting effects in string hadronization". *Eur. Phys. J. C* **82,** 228. arXiv: 2201.06316 [hep-ph] (2022).

Printed by Tryckeriet i E-huset, Lund 2023

LUND
UNIVERSITY

Department of Physics
Faculty of Science
ISBN 978-91-8039-838-1 (print)
ISBN 978-91-8039-837-4 (pdf)