NLO BFKL and perturbative QCD predictions for Mueller-Tang jets at the LHC

Master Thesis

Pablo Enrique González Durán

Double Master Degree
University of Münster and University of Seville

First advisor: Prof. Dr. Michael Klasen
Second advisor: Dr. Jens Salomon
Second examiner: Prof. Dr. Joaquín José Gómez Camacho

Contents

Chapter 1

Introduction

Although dijet events are some of the most common occurrences in particle colliders, there are still phenomena that have not been thoroughly explained by quantum chromodynamics (QCD), even though we have known about them for more than half a century. One such example is jet-gap-jet events, where the two jets produced have a region between them that is devoid of particles.
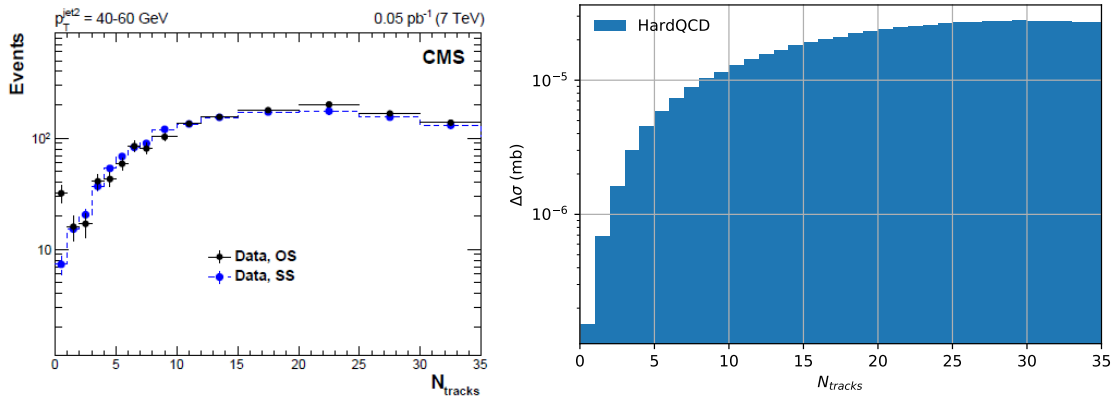

Figure 1.1: On the left the number of charged particles produced between the two leading jets is plotted for events with two jets on the same side (SS) and on opposite sides (OS). Image taken from [Col17]. On the right is a Monte Carlo simulation of the same observable for OS events using Born-level QCD processes.

As seen in Fig. 1.1, if both jets from the event are in opposite hemispheres there is an excess of events with no charged particles between them, when compared with events in which the jets exit in the same hemisphere.

Using Born level quantum chromodynamic predictions it is possible to reproduce the observed behaviour when jets end in the same hemisphere, but it is not possible to explain the gap that forms when jets are back to back. For that we need to go beyond regular perturbative QCD.

This master thesis will give an introduction to the physics involved in such events and will perform a phenomenological analysis of results using an up-to-date Monte Carlo event generator.

In Ch. 2 the framework which serves as a foundation for subsequent work, quantum chromodynamics will be explored, paying close attention to the most important properties of the theory. Then the DGLAP formalism will be briefly laid out as an introduction for the next sections as well as explore how QCD deals with soft physics in hadrons. The last two sections focus on the special cases considered in this work, BFKL dynamics and the Mueller-Tang analysis.

In Ch. 3 the means by which QCD is explored in hadron-hadron collisions is discussed, Monte Carlo (MC) event generators. Each part of the MC used in this work will be explained, starting from the partonic processes, parton showers, secondary soft processes and finally hadronisation. An overview for the physics behind each one will be given so that later sections can be easily understood.

In Ch. 4 the Mueller-Tang analysis is performed for different settings and always comparing the results with experimental data. Extensive analysis of various effects observed is also done in order to understand the exact mechanisms that lead to our results.

Finally in Ch. 5 the main points and conclusions obtained in the previous chapter, as well as an overview of possible next steps, will be laid out.

Chapter 2

Physics Background

Quantum Chromodynamics is the non-Abelian quantum field theory used to model the strong interactions in the Standard Model. Although it has had many successes, there are still challenges when it comes to making predictions. In order to get around those problems two main approaches are taken: Lattice QCD and Perturbative QCD. The former makes numerical computations on a lattice of points, given the potentials and interactions that arise from the theory, and the latter makes predictions making use of perturbation theory, with the limitations and constraints that it carries. This chapter will focus on this last approach and explain the strengths and weaknesses of it.

2.1   Quantum Chromodynamics

First of all, let us consider the main characteristics present in Quantum Chromodynamics (QCD) in this section. We do not attempt to give an exhaustive explanation of the theory, instead we try to give a sufficient introduction to it in order to be able to make sense of subsequent chapters.

2.1.1   QCD Basics

The basic components of QCD are quarks and gluons. Quarks are massive fermions that carry electric charge either $+2/3$e or $-1/3$e and one colour charge, either red, green or blue. The quark antiparticles, antiquarks, carry opposite electric charge and one anticolour charge. Gluons on the other hand are massless bosons, force carriers, with no electric charge but they carry one colour and one anticolour charge.

The simplest Lagrangian for QCD is enough to illustrate the properties and kinks of the theory that are required to be understood for this work. Denoting a quark field of flavour f and carrying colour i by $q_i^f(x)$, it's anti-quark field by $\bar{q}_i^f(\mathrm{x})$ and the gluon fields carrying colour index a in the adjoint representation by $A_\mu^a(x)$, the Lagrangian density is

$$\mathcal{L}_{\mathrm{QCD}} = \sum_f \bar{q}_i^f(x)[i\gamma^\mu D_\mu - m_f]_{ij} q_j^f(x) - \tfrac{1}{4} F_{\mu\nu}^a F^{a\mu\nu}, \qquad\qquad [2.1]$$

where the gauge covariant derivative is

$$D_\mu = \partial_\mu - igt^a A_\mu^a$$

and the gluon field strength tensor

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc} A_\mu^b A_\nu^c.$$

The generators of SU(3) in the fundamental representation are t^a, equivalent to the Gell-Mann matrices multiplied by a factor of $1/2$, $\lambda^a/2$, and f^{abc} are the structure constants of colour SU(3).

2.1.2 Perturbation Theory

When introducing interactions between free particles it is no longer possible to solve the field equations exactly, which means that some kind of approximation needs to be taken. One such method is called perturbation theory, in which one expects the interaction to be small and expands any observable A as a power series,

$$A = A_n \alpha^n + A_{n+1} \alpha^{n+1} + A_{n+2} \alpha^{n+2} + A_{n+3} \alpha^{n+3} + \dots, \qquad [2.2]$$

where n denotes the order of the leading contribution and α is a parameter of the theory which has to be sufficiently small in order for the expansion to converge.
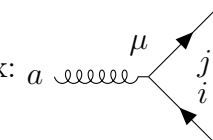
The most common way to study different orders of perturbation theory is by use of the graphical method of Feynman diagrams, in which each element of a graph represents a mathematical element in the amplitude calculation.
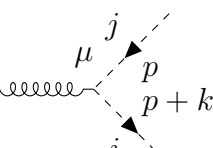
One common occurrence is that many, or even all, orders of perturbation theory will have terms that can be treated as a convergent series, e.g. a geometric series. When that happens it is possible to perform a *resummation* in said factors. On the other hand, using perturbation theory to calculate an observable up to an order n is referred to as *fixed-order perturbation theory*.
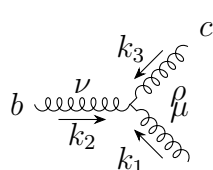
2.1.3 Feynman Rules
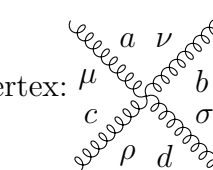
In order to calculate the amplitude of a process at a given order of perturbation theory, all Feynman diagrams that give a contribution to said order must be considered and the resulting amplitudes added together. Translation between Feynman diagrams and mathematical expressions is done with the use of Feynman rules, which for QCD are:

$$\text{quark propagator: } i \xrightarrow{p} j \quad = i\frac{\slashed{p} + m}{p^2 - m^2 + i\epsilon}\delta^{ab},$$

$$\text{gluon propagator: } {}^a_\mu \underset{}{\overset{p}{\sim\!\sim\!\sim\!\sim}} {}^b_\nu \quad = i\frac{-D_{\mu\nu}(p)}{p^2 + i\epsilon}\delta^{ab},$$

$$\text{ghost propagator: } a \dashrightarrow{p} b \quad = i\frac{1}{p^2 + i\epsilon}\delta^{ab},$$

quark-gluon vertex: $= ig_s\gamma^\mu(t^a)_{ij},$

ghost-gluon vertex: $= g_s(p+k)^\mu f^{abc}$

three-gluon vertex: $= -g_s f^{abc}[(k_1-k_3)^\nu g^{\mu\rho} + (k_2-k_1)^\rho g^{\mu\nu} + (k_3-k_2)^\mu g^{\nu\rho}]$,

four-gluon vertex: $= ig_s^2[f^{abe}f^{cde}(g^{\mu\rho}g^{\mu\sigma} - g^{\mu\sigma}g^{\nu\rho}) + f^{ace}f^{bde}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\sigma}g^{\nu\rho}) + f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma})]$.

In addition to those rules one needs to add the ones that apply to all field theories, such as conservation of energy and momentum or adding a -1 factor for each loop in the diagram.
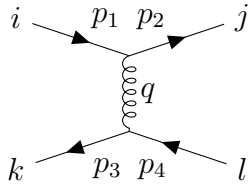
A tree level gluon exchange between a $q\bar{q}$ pair is then represented by the following Feynman diagram and has the amplitude,

$$= (t^a)_{ji}(t^a)_{kl}(ig_s)^2 \bar{u}_j(p_2)\gamma^\mu u_i(p_1)\frac{-i\left[g_{\mu\nu} - (1-\xi)\frac{q_\mu q_\nu}{k^2}\right]}{k^2}\bar{v}_k(p_3)\gamma^\nu v_l(p_4).$$

This diagram also serves to illustrate how colour charges interact. The colour behaviour is determined by the factor $(t^a)_{ji}(t^a)_{kl}$, which translates to potentials [Sch14, § 26.2]

$$V(r) = -\frac{4}{3}\frac{g_s}{4\pi r} \qquad \text{if charges are in colour singlet state,} \qquad [2.3a]$$

$$V(r) = \frac{1}{6}\frac{g_s}{4\pi r} \qquad \text{if charges are in colour octet state.} \qquad [2.3b]$$

Therefore only colour singlet states are stable. This potential does not hold for partons inside hadrons as perturbation theory breaks down for low energy interactions, as discussed in the next section.

2.1.4 Running Coupling

When performing precise calculations, loop diagrams such as the one in Fig. 2.1 need to be taken into account. Unfortunately they lead to infinities when integrating momenta over the loops, which requires some treatment in order to make the theory predict finite observables.
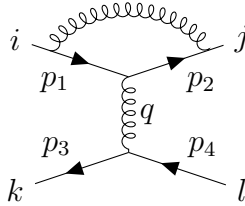

Figure 2.1: One loop diagram correction.

There are several ways to deal with these divergences, known as renormalisation schemes. One well known scheme is the *Modified Minimal Subtraction* or MS-bar scheme, $\overline{\text{MS}}$, [Sch14, § 26.6] which introduces a renormalisation scale, μ^2, that absorbs the divergences. Since this scale is purely theoretical, observables must not depend on it. That is, for an observable O,

$$\mu^2 \frac{d}{d\mu^2} O(\mu, \alpha_s) = \mu^2 \left(\frac{\partial}{\partial \mu^2} + \frac{\partial \alpha_s}{\partial \mu^2} \frac{\partial}{\partial \alpha_s} \right) O(\mu, \alpha_s) = 0. \tag{2.4}$$

This equation is known as the renormalisation group equation. The right hand side tells us that the coupling must change in order for the observable to remain independent of μ^2. Solving it for QCD at leading order yields,

$$\alpha_s(\mu^2) = \frac{1}{\beta_0 \ln\left(\frac{\mu^2}{\Lambda_{\text{QCD}}^2} \right)} = \frac{12\pi}{(33 - 2n_f) \ln(\mu^2/\Lambda_{\text{QCD}}^2)}, \tag{2.5}$$

where n_f is the number of flavours considered and Λ_{QCD} is a parameter of the theory, commonly called *Landau pole of QCD*, which can't be predicted and depends on the specifics of the renormalisation scheme used. However, its value is commonly around $\Lambda_{\text{QCD}} \sim 300\text{MeV}$, i.e similar to the constituent mass of valence quarks in protons.

The coupling constant depending on the energy scale involved is called *Running Coupling*. The running of the strong coupling in Eq. 2.5 gives rise to the two main properties of QCD, what is known as *asymptotic freedom* and *colour confinement*. Asymptotic freedom refers to the fact that, at high energy scales, $\mu^2 \gg \Lambda_{\text{QCD}}^2$, the coupling becomes very small and interacting particles can be approximated as free particles. Meanwhile, the phenomenon of never encountering free colour octet charges in nature is called colour confinement and it is a bit harder to explain from a theory point of view but will be expanded upon in Sec. 3.4. Basically, for $\mu^2 \sim \Lambda_{\text{QCD}}^2$, $\alpha_s(\mu^2)$ grows and ultimately diverges. Since it is possible to equate high energy with smaller distances and low energies with longer distances, the aforementioned behaviour results in the confinement of particles within hadrons, and given the nature of the interactions given by Eq. 2.3, in colour singlet states.

Lastly, perturbative QCD uses α_s as the expansion parameter. The running of the coupling provides the main limit to the use of *perturbative QCD*; requiring α_s to be sufficiently small translates to μ^2 being sufficiently large compared with Λ^2_{QCD}. Therefore it is not possible to apply perturbation theory to situations with low energy interactions such as the interior of hadrons.

2.2 DGLAP

When studying scattering events involving hadrons one encounters two different scales: a soft, non-perturbative scale that dominates the interactions between partons belonging to the same hadron and a hard, perturbative scale that is given by the transferred momentum between the interacting particles. The hard scale interaction can be readily calculated using the perturbative Feynman rules given in Sec. 2.1.3. However, the soft scale interaction is not so easy to deal with, and, in fact, there is no perturbative or analytical way to predict how partons behave inside hadrons.

There is a way around this limitation, however, since it is possible to "lump" all non-perturbative physics dominated by the soft scale into *parton distribution functions* (PDF), which give the probability density of finding any given parton carrying longitudinal momentum fraction x when being probed at energy scale Q^2. This is known as the *factorisation theorem* [Sch14, § 32.4] which, roughly speaking, states that cross sections can be calculated in two parts as

$$\sigma = f * H, \tag{2.6}$$

where f is a PDFs containing all the soft physics, H is the perturbatively calculated cross section and $*$ denotes convolution. This approach works as long as $Q^2 \gg \Lambda^2_{\text{QCD}}$, which equates to saying the *interaction time* (also known as *Ioffe time*, cf. [KL12, § 2.3]) and *spatial resolution* of the hard interactions are small when compared with the soft interactions.

Fortunately, even though it is not possible to predict PDFs, it is also not required to measure them at all energy scales; just one is enough. After measuring the PDF at energy scale Q_0^2, applying Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [Sch14, § 32.3] equations it is possible to "evolve" it to another value of the scale Q^2.

DGLAP equations account for diagrams in which a splitting occurs, such as the ones seen in Fig. 2.2, where the final, interacting parton comes from a succession of splittings from other partons.

Taking into account these corrections amounts to performing a resummation of all diagrams in the parameter

$$\alpha_s \ln\left(\frac{Q^2}{\Lambda^2_{\text{QCD}}}\right),$$

with the energy scale $Q \gg \Lambda_{\text{QCD}}$. The reason to resum all diagrams with this factor instead of just α_s is that α_s is considered small while $\ln Q^2/\Lambda^2_{\text{QCD}}$ becomes big enough so that $\alpha_s \ln Q^2/\Lambda^2_{\text{QCD}} \sim 1$. For this reason it is often referred to as the *leading log approximation* (LLA) or *leading order DGLAP*.

Working from that assumption, the DGLAP equations can be found, entirely within the realm of perturbative QCD. This gives a set of coupled differential equations that tell how PDFs change as the interaction scale changes, based on the values of all other PDFs and
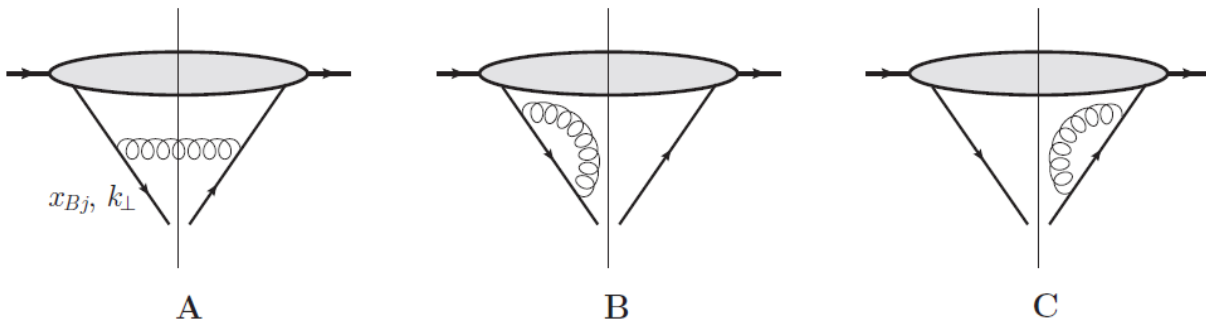
Figure 2.2: Diagrammatic representation of the lowest order corrections to parton distribution functions, represented by the blob, where the interacting parton takes part in one gluon loop diagrams in addition to the hard interaction (lower vertex, not represented). The right side of the line represents the complex conjugate. Figure taken from [KL12, p. 45].

the possible vertices in QCD seen in Sec. 2.1.3. There are different ways to represent the DGLAP equations, but one of the most intuitive is, at leading order,

$$Q^2 \frac{d}{dQ^2} \begin{pmatrix} f_i(x, Q^2) \\ f_g(x, Q^2) \end{pmatrix} = \sum_j \frac{\alpha_s}{\pi} \int_x^1 \frac{d\xi}{\xi} \begin{pmatrix} P_{q_i q_j}(\frac{x}{\xi}) & P_{q_i g}(\frac{x}{\xi}) \\ P_{g q_j}(\frac{x}{\xi}) & P_{gg}(\frac{x}{\xi}) \end{pmatrix} \begin{pmatrix} f_j(x, Q^2) \\ f_g(x, Q^2) \end{pmatrix}, \quad [2.7]$$

with $f_i(x, Q^2)$ being the ith flavour parton distribution function and $f_g(x, Q^2)$ being the gluon distribution function. The functions $P_{ji}(z)$ are called splitting functions and give the probability density of a parton i taking part in a split that produces parton j. The exact form of the splitting functions at leading order is

$$P_{qq}(z) = C_F \left[\frac{1 + z^2}{[1 - z]_+} + \frac{3}{2} \delta(1 - z) \right] \quad [2.8a]$$

$$P_{qg}(z) = T_F \left[z^2 + (1 - z)^2 \right] \quad [2.8b]$$

$$P_{gq}(z) = C_F \left[\frac{1 + (1 - z)^2}{z} \right] \quad [2.8c]$$

$$P_{gg}(z) = 2C_A \left[\frac{z}{[1 - z]_+} + \frac{1 - z}{z} + z(1 - z) \right] + \frac{\beta_0}{2} \delta(1 - z), \quad [2.8d]$$

with the plus notation meaning,

$$\int_x^1 dz \frac{1}{[1 - z]_+} f(z) = \int_x^1 dz \frac{1}{1 - z} [f(z) - f(1)] + f(1) \ln(1 - x). \quad [2.9]$$

A problem with the DGLAP equations arises when the centre of mass energy of the interaction, $\sqrt{s}$, is much larger than the interaction scale Q. For a $2 \to 2$ process, kinematics dictate that

$$x_1 x_2 = \frac{4p_T^2}{s} \cosh^2(\hat{\eta}). \quad [2.10]$$

Increasing s will make the product $x_1 x_2$ become smaller, leading the conclusion

$$s \sim 1/x, \implies \text{ high s} \equiv \text{small x}, \qquad [2.11]$$

which is an important heuristics when studying high energy events. As a consequence terms with large $\ln(s) = \ln(1/x)$ become large and need to be taken into account.

2.3 BFKL

The Balinsky-Fadin-Kuraev-Lipatov (BFKL) [KL12, § 3] equation sets out to do what the DGLAP equation does, but instead of evolving the distribution in the Q^2 direction, it evolves it in the $1/x$ direction, as illustrated in Fig. 2.3. This means that now the resummation is performed in the parameter

$$\alpha_s \ln \left(\tfrac{1}{x}\right).$$
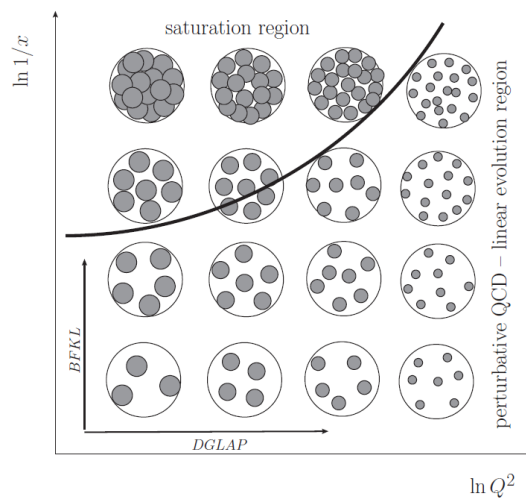
Figure 2.3: Parton distribution function representation in the transverse plane as a function of $\ln Q^2$ and $\ln 1/x$. Figure taken from [KL12, p. 115].

As a result, BFKL takes into account large logarithms in s that were ignored in the DGLAP equations presented in the previous sections. The derivation of the BFKL equation is long and arduous [KL12; FR97], so instead let us present the equation and discuss its properties and consequences. In light cone perturbation theory, the BFKL equation takes the form

$$\frac{\partial G(\vec{l_\perp}, \vec{l_\perp'}, Y)}{\partial Y} = \frac{\alpha_s N_C}{\pi^2} \int \frac{d^2 \vec{q_\perp}}{(\vec{l_\perp} - \vec{q_\perp})^2} \left[G(\vec{q_\perp}, \vec{l_\perp'}, Y) - \frac{\vec{l_\perp}^2}{2\vec{q_\perp}^2} G(\vec{l_\perp}, \vec{l_\perp'}, Y) \right], \qquad [2.12]$$

where $G(\vec{l_\perp}, \vec{l_\perp'}, Y)$ is called the Green function of BKFL.

The derivation of the BFKL equation relies on two assumptions: perturbative QCD is valid and the Regge limit is applicable. As discussed previously, perturbation theory is

reliable only when the running coupling is sufficiently small (cf. Sec. 2.1.4) which means that energy scales are far away from the QCD Landau pole, that is, $Q^2 \gg \Lambda^2_{\text{QCD}}$.

The second assumption makes use of the *Regge limit* [FR97, § 1], named after Italian physicist Tulio Regge, who studied high energy collisions using S-matrix formalism. What would latter be called Regge theory predicted that as centre of mass energy becomes much greater than the exchanged momentum for a $2 \to 2$ process, $s \gg Q^2$, cross sections asymptotically approach

$$\sigma \propto s^{\alpha(Q^2)}, \quad [2.13]$$

with the function $\alpha(Q^2)$ being called *Regge trajectory*. When particle exchanges follow this behaviour they are said to *Reggeize*.

Thus, with these two assumptions the BFKL equation imposes a strict hierarchy of energy scales; where DGLAP demanded only that $Q^2 \gg \Lambda^2_{\text{QCD}}$, BFKL instead demands $s \gg Q^2 \gg \Lambda^2_{\text{QCD}}$.

When talking about these evolution equations there is a certain level of ambiguity. They always *do* the same, but the separation between the diagrams in Fig. 2.2 belonging to the collective wave function of the hadron or being part of the hard interaction diagram is arbitrary and a matter of the factorisation scheme used. This is possible only because those diagrams are calculated using perturbative theory and are not part of the soft physics enclosed by PDFs, otherwise it would not be possible to extract them. Therefore, the BFKL equation can also be interpreted as an exchange diagram. In this case the corresponding diagram would be a "ladder" diagram, like the one in Fig. 2.4.
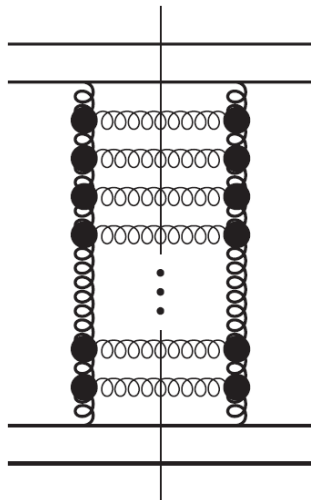
Figure 2.4: BFKL ladder diagram. The sides of the ladder are reggeized gluons (bold corkscrews) and the vertices are effective Lipatov vertices (Fig. 2.5). Figure taken from [KL12, p. 94].

The sides of the ladder are *Reggeized gluons*, whereas the rungs of the ladder are regular gluons. The black dots represent *Lipatov effective vertices*, represented in Fig. 2.5, which take into account all diagrams with a real emitted gluon.

The result of all of this is an exchange of momentum, with no transfer of colour or electric charges, or flavour across the colliding particles. For this reason it is often called the *hard*

Figure 2.5: Lipatov effective vertex. It takes into account all diagrams with a real emission gluon. Figure taken from [KL12, p. 85].

Pomeron, as it matches with the qualities of the Pomeron, a Regge trajectory that carries the quantum numbers of the vacuum and is the only trajectory that has a positive intercept, but that is entirely derived using perturbative QCD.

2.4 Mueller-Tang Analysis

The focus of this work is the analysis of a type of event commonly referred to as *Mueller-Tang jets* [MT92] (MT). The MT analysis is based on two foundations: forward jets and a gap.

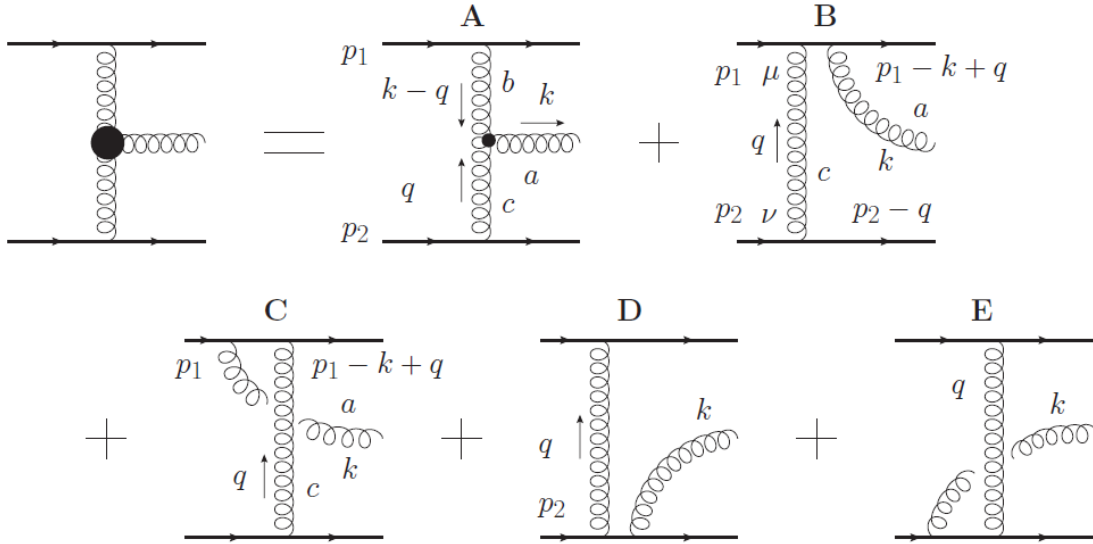Imposing forward jets enhances the Pomeron exchange. This is easiest seen in terms of kinematical relations. For a $2 \to 2$ process (cf. Appendix A), the relation

$$\cos \theta = 1 - \frac{2Q^2}{s},$$

[2.14]

where θ is the scattering angle, makes it so that having a small scattering angle implies $Q^2 \ll s$, placing the event inside the Regge limit. Since the particles exchanged can be considered as reggeized, thus following Eq. 2.13, and the Pomeron is the only Regge trajectory with a positive intercept, there is an enhancement of Pomeron exchange with respect to basic QCD reggeized components. Unfortunately two problems arise: detectors cannot cover extremely forward scattering angles and even if they could,

$$\Delta \eta \gg 1 \implies Q^2 \sim p_\mathrm{T}^2$$

[2.15]

means that as $p_\mathrm{T}^2 \to 0$, predictability is lost as all scales become soft and perturbative QCD becomes unreliable. This would also break the second condition of the BFKL hierarchy $Q^2 \sim p_\mathrm{T}^2 \gg \Lambda_\mathrm{QCD}^2$.

If p_{T}^2 is made to be not too small the BFKL Pomeron enhancement that motivated this setup in the first place is partially lost. Here enters the second foundation: a gap with no particles produced between the two leading jets. Since the Pomeron carries no electric charge, no QED emissions take place, and since no colour charge is transferred there is no QCD production of particles in the space between the two resulting jets. However, because in QCD all leading order particle exchanges carry colour charge particle production is heavily favoured. Therefore a lack of particles found between the leading jets can be used as a signature of BFKL events. We shall discuss the mechanisms of particle production through colour interactions in more detail in Sec. 3.4.

The importance of this analysis to study BFKL events cannot be overstated. It gives a kinematical regime that is extremely easy to enforce in particle accelerators, greatly enhancing BFKL cross sections and additionally provides a way to discriminate BFKL from other types of events.

Using the MT prescription [MT92] at NLL-BFKL [KMR10] the resulting amplitude for MT jets is

$$A(\Delta\eta, p_{\mathrm{T}}^2) = \frac{16 N_C \pi \alpha_s^2(p_{\mathrm{T}}^2)}{C_F p_{\mathrm{T}}^2} \sum_{p=-\infty}^{\infty} \int \frac{d\gamma}{2i\pi} \frac{[p^2 - (\gamma - 1/2)^2] \exp\{\bar{\alpha}(p_{\mathrm{T}}^2)\chi_{eff}\left(2p, \gamma, \alpha(\bar{p}_{\mathrm{T}}^2)\right)\Delta\eta\}}{[(\gamma - 1/2)^2 - (p - 1/2)^2][(\gamma - 1/2)^2 - (p + 1/2)^2]},$$

[2.16]

with the integral running from $1/2 - i\infty$ to $1/2 + i\infty$ and with only even *conformal spins*, p, contributing. In the amplitude $\chi_{eff}(p, \gamma, \bar{\alpha})$ is the effective kernel obtained from the NLL kernel using the formula

$$\chi_{eff} = \chi_{NLL}(p, \gamma, \bar{\alpha}\chi_{eff}).$$

[2.17]

Once again we see the enhancement in BFKL cross section in the $\exp(\Delta\eta)$ factor in the amplitude. For details and discussion on the amplitude calculation we refer to [MR09].

Chapter 3

PYTHIA 8

Monte Carlo (MC) event generators have proven an invaluable tool in high energy physics. The use of random numbers to sample from probability distributions is a natural way of probing theories that deal with complex and hard to disentangle effects that would otherwise be unfeasible to compute analytically. Here we present a brief overview of the main components of Monte Carlo generators.

The mechanism by which MC event generators work is by sampling from distributions in order to make decisions about the event evolution. The result is a single, random "trajectory" of the event out of all possible ones for the process considered. When enough events are sampled, the overarching distribution can be revealed through statistic analysis.

There are many different Monte Carlo event generators for high energy physics on the market such as HERWIG or PYTHIA. This work uses PYTHIA 8 (v8.244) since it is still regularly updated to this day and its robustness has been extensively tested. Therefore, while the different components and the physics behind them apply to all MC event generators, any implementation details refer only to how PYTHIA 8 deals with them.

The sections presented here are not meant to be an extensive explanation of the algorithms and theory behind each component. Instead they are meant to give a good picture of the inner workings of the event generators, so that latter explanations may be more easily understood. For example, multi-parton interactions, final and initial state radiations do not take part independently, as might be suggested by their presentation in this chapter, but are instead computed in an inter-leaved algorithm that generates all of them using the same evolution scale. This is, however, an implementation detail that is not strictly required to understand the basis of each section.

3.1   Partonic Processes

We are mainly interested in $2 \rightarrow 2$ processes, a kind of process that PYTHIA is optimised to handle along with $2 \rightarrow 1$ processes. One such process, in generic form, would be $ab \rightarrow cd$, with the corresponding diagram in Fig. 3.2. The cross section for regular, Born-level partonic QCD interactions can be easily calculated using perturbation theory. In colliders, however, the basic objects are not partons, but hadrons, so the cross sections needed are
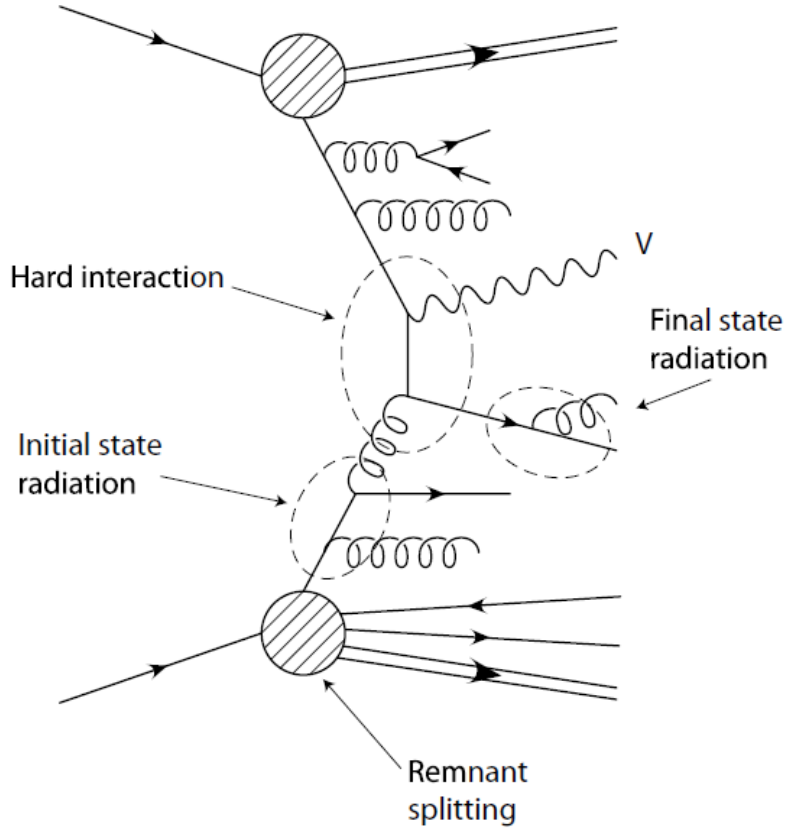
Figure 3.1: Representation of different parts of a MC event generator. Figure taken from [Enb03, p. 45].

those convoluted with PDFs as

$$\sigma_{ab\to cd} = \int dx_1 \int dx_2\, f_a^1(x_1, Q^2)\, \hat{\sigma}_{ab\to cd}\, f_b^2(x_2, Q^2), \qquad [3.1]$$

where $f_i^j(x, Q^2)$ is the parton distribution function corresponding to the parton flavour i, inside hadron j and probed at Q^2.
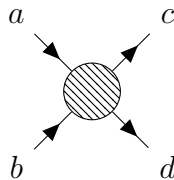
Figure 3.2: Generic diagram for $ab \to cd$ process.

PYTHIA offers all possible $2 \to 2$ possible events present in vanilla QCD, both elastic and inelastic. One thing to note is that PYTHIA uses an unregularised QCD cross section, which means that a cutoff, $p_{T,0}$, needs to be used in order to avoid divergences as $p_T \to 0$. In effect the cutoff acts as a generation limit below which no calculations are considered.

Besides the $2 \rightarrow 2$ QCD processes, BFKL events are made use of in the Mueller-Tang analysis. As BFKL events are always elastic there is no need to consider different end particles. It is worth mentioning that BFKL couples *equally* to quarks and gluons, only differing in a colour factor.

BFKL ladder exchange is added into PYTHIA by introducing the differential cross sections of the process and making PYTHIA sample from it. The cross section is at next-to-leading log (cf. Eq. 2.16) using the S4 resummation scheme [MR09] and is not calculated on the spot, since that would require performing a calculation that would be too computationally expensive to be useful. Instead, the cross section parametrisation from [KMR10] has been fed to PYTHIA machinery,

$$\frac{d\sigma}{dp_\mathrm{T}^2} = \frac{\alpha_S^4(p_\mathrm{T}^2)}{4\pi p_\mathrm{T}^4}[a+bp_\mathrm{T}+c\sqrt{p_\mathrm{T}}+(d+ep_\mathrm{T}+f\sqrt{p_\mathrm{T}})\cdot z+(g+hp_\mathrm{T})\cdot z^2+(i+j\sqrt{p_\mathrm{T}})\cdot z^3+\exp(k+lz)],$$

[3.2]

where $z(p_\mathrm{T}^2) = \bar{\alpha}(p_\mathrm{T}^2)\Delta\eta/2$ and $\bar{\alpha}(p_\mathrm{T}^2) = \alpha_S(p_\mathrm{T}^2)N_c/\pi$. The use of this fit to the cross section is why our results will be often referred to as "approx".

In addition to this, another element needs to be taken into consideration, *impact factors*. These impact factors account for the coupling between partons and the BFKL ladder, such that the parton distribution function can be evolved using regular DGLAP while maintaining consistency with BFKL calculations. The impact factors are at leading-order, and together with the BFKL fit have been implemented into PYTHIA by Jens Salomon and Federico Deganutti.

It is worthwhile to note that in the event generator flow, it is the partonic process at the very beginning what sets the scale for all subsequent routines.

3.2 Parton Showers

Showers take into account diagrams where one or more emissions happen in addition to the hard interaction, as seen in Fig. 3.3. This happens in almost every event, since the partons in hadrons are constantly splitting inside hadrons, Sec. 2.2, and the splitting functions have a collinear singularity, giving raise to jets. In the context of Monte Carlo event generators, the former is called initial state radiation while the latter is called final state radiation.

Showering algorithms are based on Sudakov form factors and splitting functions, like the ones in Eq. 2.8, in addition to the electrodynamic splitting functions, as quarks and many of the final state particles (hadrons) have electric charges and can, consequently, engage in such interactions.

Figure 3.3: Initial and final state radiation diagrams.

Sudakov form factors give the probability of a split happening in a given interval. If the

differential probability of a single particle branching out is given by

$$dP = \sum_{ij} \frac{\alpha_{ij}}{2\pi} P_{ij}(z) \, dt \, dz, \qquad [3.3]$$

where the sum over ij goes over all possible branchings, including QED ones, and α_{ij} is the coupling constant involved in the splitting in question. In general, t can be any variable singular in the collinear limit [Sch14, § 32.3], but in PYTHIA [SMS06, § 10.1.1], dt is particularly defined as

$$dt = d\ln(Q^2) = \frac{dQ^2}{Q^2}, \qquad [3.4]$$

and z is the proportion of energy that one of the daughter particles in the branching get from the mother particle, while the other one gets $1 - z$.

Integrating over all allowed values of z one obtains

$$\mathcal{I}(t) = \sum_{ij} \int_{z_-(t)}^{z_+(t)} \frac{\alpha_{ij}}{2\pi} P_{ij}(z) \, dz, \qquad [3.5]$$

which is the differential probability that any branching occurs for the particle observed with respect to the variable t. The probability that no branching has occurred in an interval δt is $1 - \mathcal{I}(t)\delta t$.

For a parton that begins at t_0, the cumulative probability that it has not branched by t is equal to the product of the probabilities that no branching occurs in all previous t differentials. That is, the probability that no branching has happened between t_0 and t exponentiates and is given by

$$\mathcal{P}_{\text{no-branch}}(t_0, t) = \exp\left\{-\int_{t_0}^{t} dt' \, \mathcal{I}(t')\right\} = S(t, t_0). \qquad [3.6]$$

$S(t)$ is also known as the *Sudakov form factor*. The probability of an emission happening in an interval δt is therefore the differential probability that it happens within said interval times the probability that no branching has happened before,

$$dP(t) = \mathcal{I}(t) \, S(t, t_0). \qquad [3.7]$$

Given this cumulative probability it is possible to iteratively generate random numbers and proceed through the branching any given parton undergoes. Notice that the variable t fulfils a role similar to time, in the sense that as the shower progresses in time, t always decreases until it reaches a cutoff, at which point the shower generation is stopped.

This sets the basis for showering algorithms in most event generators. While the algorithm above is efficient for final state radiation, which would account for showers between the hard collision, at $t_{\max}$, and a cutoff for infrared safety, at t_0, initial state radiation needs a slight modification to the algorithm in order to be efficiently computed using PDFs. Since the hard interaction scale is fixed at the very beginning, instead of trying different "time"-forwards until a matching branching history that ends in the correct $t_{\max}$ is found, a "time"-backwards algorithm is used that evolves from $t_{\max}$ to the initial scale t_0.

3.3   Soft Interactions

This section revolves around non-perturbative interactions that may happen during the scattering event. It is roughly divided into two parts: *multiparton interactions* and *colour reconnections*. One thing to note is that since these are non-perturbative, very little is known about the exact details of how they happen in the context of QCD. Therefore, models may vary widely, with their success being measured mainly by their ability to reproduce data.

Multiparton interactions (MPI) try to correct the naïveness of assuming there is only one single interaction between partons when hadrons collide. In PYTHIA these secondary interactions happen at a lower interacting scale than the main hard event, but are, nonetheless, considered to have the same perturbative cross sections as if they were hard interactions. The basic probability parameter is the ratio between partonic hard interaction cross section, $\sigma_{\mathrm{hard}}(p_{\mathrm{T,min}})$ and non-diffractive inelastic hadronic cross sections, $\sigma_{\mathrm{nd}}(s)$, such that the average number of secondary interactions is

$$\bar{n} = \frac{\sigma_{\mathrm{hard}}(p_{\mathrm{T,min}})}{\sigma_{\mathrm{nd}}(s)}. \tag{3.8}$$

There is an additional scale that has been introduced in the partonic cross section, $p_{\mathrm{T,min}}$, in order to regularise the infrared divergence these cross sections have. With this, a Sudakov-like factor can be constructed, which allows the use of an algorithm similar to that of the showers.

Additionally, an impact parameter, b, is introduced. The physical intuition of this parameter is that it is equal to the closest approach between the two hadrons, and is used in conjunction with some model for the mass distribution of the hadrons as a prefactor to cross sections so that "direct" impacts, with more overlap, produce more secondary interactions than a grazing one, with barely any overlap. A bit more nuance is required, as experimentally the dependence on this factor seems to saturate and some mass distribution models, such as the double Gaussian, never return a null value, which demands the imposition that no soft interactions can occur if not preceded by a hard one.

Finally, each MPI needs to fulfil some further conditions, such as the beam remnant, i.e. what remains of colliding hadrons, taking the recoil from any emissions while preserving energy and momentum, as well as assigning colour charges in a way that is consistent with the initial hadrons being in singlet states. Each MPI system can also have their own showers, by the way.

One important concept to understand is that of the *colour connection*. When a system of partons is in a colour singlet state, the colour charges of the partons are not static since the attractive interaction they experience is mediated by gluons, which cause rotations in colour space. However, even though the colour charges can change, defining one colour charge immediately defines the other so the colour singlet is preserved, e.g. a colour singlet dipoles can alternate between the states $|r\bar{r}\rangle$, $|g\bar{g}\rangle$ and $|b\bar{b}\rangle$. Instead of tracking colour charges that can randomly change it is much more efficient to use colour connections, which just indicate which partons have their colour charges linked so they form a colour singlet state.

After MPI, all colour connections are determined. Partons that were scattered may be colour connected if they exchanged a triplet or octet charge and all other emissions from a colour singlet partonic system will be assigned a colour in order to preserve total

colour charge. This, however, completely negates the possibility of exchanging super soft gluons that basically carry zero momentum but still carry colour charge, which results in a rearrangement of colour connections, which will have a great impact on the hadronisation procedure (cf. Sec. 3.4). *Colour reconnection* (CR) deals with such events.
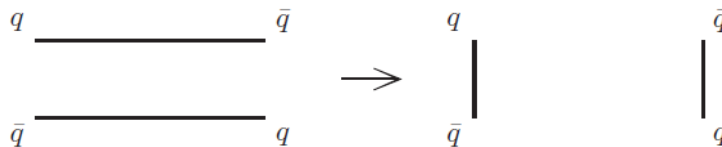

Figure 3.4: Colour reconnection example. The two dipoles on the left get their colour connections rearranged into the right side configuration. Figure taken from [CS15, Fig. 7].

There are many different models for CR but the default one in PYTHIA decides which colour singlet systems to reconnect based on a probability defined as

$$\mathcal{P} = \frac{p_{\text{T,Rec}}^2}{p_{\text{T,Rec}}^2 + p_{\text{T}}^2},$$

[3.9]

where $p_{\text{T,Rec}}$ is $p_{\text{T,min}}$ times *range*. The latter is a free parameter of the model while the former is an energy dependent dampening parameter set in the MPI routine. This definition of probability favours reconnections between low p_{T} systems and in effect seeks to minimise total string length (cf. Sec. 3.4).

PYTHIA works in *next-to-leading colour* [CS15], which means it considers also the creation of colour connection junction topologies. The simplest example of a junction colour connection would be a baryon, which can not be neatly connected with simple connections that begin in triplet and end in anti-triplet colours.

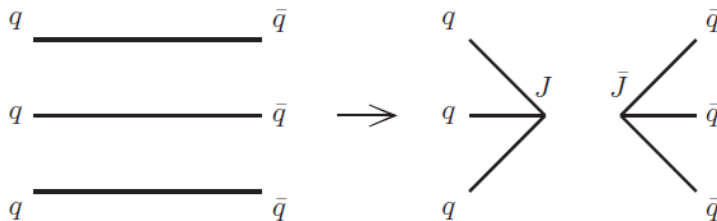Figure 3.5: Colour reconnection example. The three dipoles on the left side get their colour connections rearranged in the junction topology on the right side, corresponding to one baryon and one anti-baryon. Figure taken from [CS15, Fig. 7].

The need for these two parts of event generators comes from phenomenological analysis. Each simulates non perturbative effects and have been found to provide a better fit to data when included.

3.4 Hadronisation

Hadronisation refers to the process that ensures all final state particles are in a colour singlet state. It turns all partons generated by any of the previous steps into stable hadrons.

The hadronisation, also referred to as fragmentation, routine in PYTHIA 8 is based on the popular Lund String model [And+83] [Sjö84]. It is based on the fact that the potential between partons in a colour singlet state can be described by a linear potential,

$$V(r) = \kappa r, \tag{3.10}$$

with $\kappa \sim 1\text{GeV/fm}$.

This assumption is supported by lattice QCD calculations, bag model results, charmonium spectroscopy and Regge phenomenology [Sjö84] [Eks15]. Therefore, partons that experience an attractive potential can be construed as if they were connected by a string that stretches and stores a linear energy density.
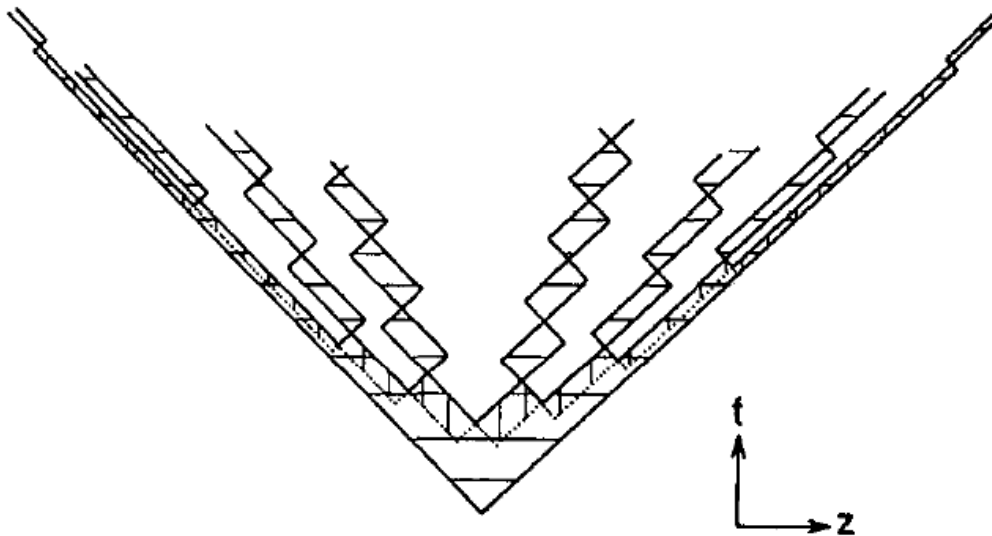
Figure 3.6: Representation of the Lund String fragmentation model. Horizontal lines denote colour string connections between the partons whose trajectories occupy adjacent diagonal lines. Figure taken from [Sjö84, Fig. 9].

Strings are massless objects that extend from a triplet charge to an anti-triplet charge. When both charges get separated, the energy of the string increases until a $q\bar{q}$ pair appears through quantum tunneling. The tunneling probability is easily described by quantum mechanics and is of the form

$$\mathcal{P} \propto \exp\left(\frac{\pi p_{\perp,q}^2}{\kappa}\right) \exp\left(-\frac{\pi m_q^2}{\kappa}\right), \tag{3.11}$$

which exponentially suppresses the creation of heavy quarks as well as quarks with large transversal momentum.

This new pair is connected through strings to the previous pair in such a way that two colour singlets are formed, following the *Colour Confinement Theorem*. The new quarks initially have zero or very low momentum, so they will begin moving as they begin to feel the force from their corresponding string, which will stretch, again raising the amount of energy stored in that string, which may or may not cause a new split. This process is repeated until no further splits occur. At that point, all partons are in colour singlet states forming hadrons which will be the observed particles.

Chapter 4

Analysis

4.1 Setup

The main objective of this work is to reproduce or compare BFKL predictions with the experimental results from the CMS Collaboration at the LHC for 13TeV pp collisions [CT20]. In particular we focus on the Mueller-Tang analysis (cf. Sec. 2.4), with the peculiarity that instead of looking for particles between jets, only *charged particles* are considered due to experimental concerns. The kinematical limits imposed on the leading jets in p_{T} are

$$p_{\mathrm{T,jet}} > 40\mathrm{GeV}, \tag{4.1a}$$

$$1.4 < |\eta_{\mathrm{jet\text{-}1,jet\text{-}2}}| < 4.7, \tag{4.1b}$$

$$\eta_{\mathrm{jet\text{-}1}}\eta_{\mathrm{jet\text{-}2}} < 0. \tag{4.1c}$$

Finding the jets was done via anti-kt algorithm [CSS08] with `SlowJet`, which is already integrated in PYTHIA 8, and setup as

$$\mathrm{Algorithm} = \mathrm{anti\text{-}kt} \tag{4.2a}$$

$$\mathrm{Max}\ \eta = 5.2 \tag{4.2b}$$

$$\mathrm{Radius} = 0.4 \tag{4.2c}$$

$$\mathrm{Min\ jet}\ p_T = 10\mathrm{GeV} \tag{4.2d}$$

$$\mathrm{Select} = \mathrm{only\ visible} \tag{4.2e}$$

$$\mathrm{Masses} = \mathrm{on\text{-}shell} \tag{4.2f}$$

In order to be counted towards the charge multiplicity, the particles in the gap region need to satisfy

$$p_{\mathrm{T}} > 200\mathrm{MeV}, \tag{4.3a}$$

$$|\eta| < 1. \tag{4.3b}$$

The lower rapidity limits are such that jets with radius $\mathcal{R} = 0.4$ will not spill into the rapidity gap region.

There are some parts in this chapter where other experiments are mentioned, mainly pp collisions at the LHC with 7 TeV CME [Col17] and $p\bar{p}$ collisions at the Tevatron at

1.8 TeV CME [aC98]. Our setup always changes to match those from the experimental results with the only exception of the jet finder for Tevatron in which the Cone jet finder algorithm was used experimentally while we still use the anti-kt one.

The main observable is the *fraction of colour singlet exchange*, or f_{CSE}, which is defined as

$$f_{\text{CSE}} = \frac{N_{\text{CSE}}}{N_{\text{CSE}} + N_{\text{non-CSE}}}. \qquad [4.4]$$

Experimentally it is not straightforward to differentiate between CSE and non-CSE events, so an auxiliary definition is devised: making use of the charge multiplicity distribution in the gap region, the background distribution is extracted by fitting a *Negative Binomial Distribution* (NBD) in the region $3 < N_{\text{tracks}} < 35$ and then subtracting the background from the total distribution, so

$$f_{\text{CSE}} = \frac{N_{\text{tot}} - N_{\text{non-CSE}}}{N_{\text{tot}}}. \qquad [4.5]$$

This represents the biggest divide between the analysis performed in experiments and ours since we are able to just differentiate between the two during event generation but cannot perform such a fit for the whole distribution. The problem lies in that MPI models are considered to be less reliable than theoretical calculations [KMR18], so instead a *Gap Survival Probability* (SP) parameter will be used. SP acts as a constant that resizes the contributions from the lowest multiplicity bins of the CSE events. Unfortunately there is no telling how the cross section is redistributed so it is not possible to perform the NBD fit to reproduce the experimental analysis. The final formula used for the fraction of CSE events by us is thus

$$f_{\text{CSE}} = \frac{N_{<3,\ \text{CSE}} \cdot SP}{N_{\text{CSE}} + N_{\text{non-CSE}}}. \qquad [4.6]$$

Finally, this observable is studied with respect to three independent variables: transverse momentum of the second leading jet, $p_{\text{T,jet-2}}$, and the differences in rapidity and azimuthal angle between the two leading jets, $\Delta\eta_{\text{jj}}$ and $\Delta\phi_{\text{jj}}$. Rapidity and azimuthal angle are measured from the centre of each jet.

The tune CP1 based on NNPDF3.1 [Col+17] from the CMS Collaboration for PYTHIA 8 was used [Col19].

4.2 Main Results

4.2.1 Main 13 TeV

One thing to note is that applying the tune *as is* for the CSE events was found to reduce f_{CSE} by a factor of about 2 without changing the dynamical behaviour. In order to obtain a better agreement with data only one parameter is required to change, `SpaceShower:rapidityOrder` which is left turned on for CSE, leaving the CP1 tune intact for QCD events, which has `SpaceShower:rapidityOrder` off (cf. Sec. B.1). As can be seen in Sec. B.1, SP can also be raised. In the end it was not done this way in order to reduce statistical uncertainties.
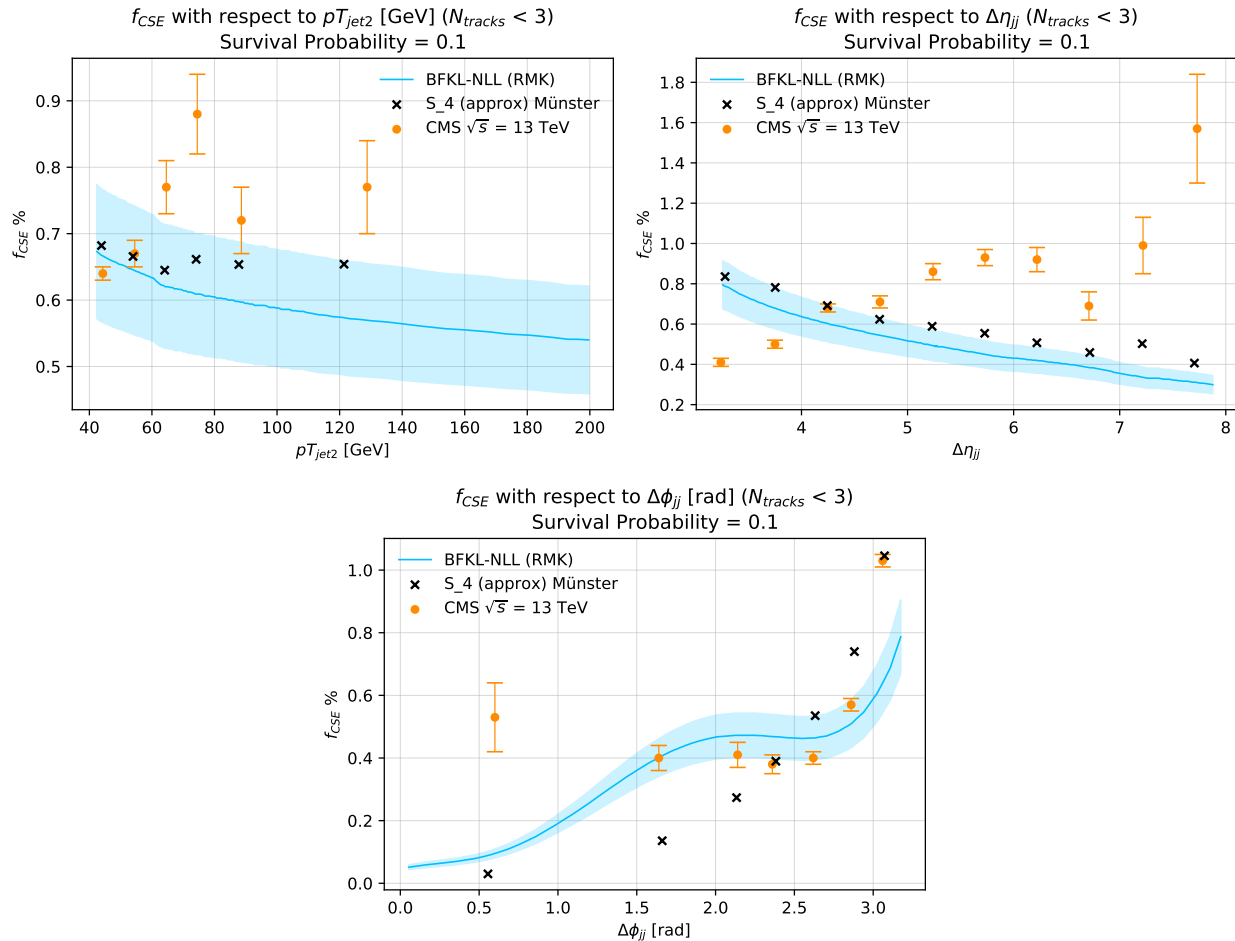
Figure 4.1: Differential plots of f_{CSE} with respect to $p_{\text{T,jet-2}}$, $\Delta\eta_{\text{jj}}$ and $\Delta\phi_{\text{jj}}$ using the tune CP1 from [Col19] but setting the parameter `SpaceShower:rapidityOrder = on` for CSE events. The experimental data is in orange circles, a previous prediction by RMK is in blue and our results using Eq. 3.2 are in black crosses. Corresponds with Tables 2, 3 and 4 or Figure 8 in [CT20].

Fig. 4.1 shows f_{CSE} with respect to the usual variables. Our results are labelled as "S_4 (approx) Münster", with "approx" referencing the fact that Eq. 3.2 was used instead of performing a full calculation of the cross section. All kinematical limits and bin sized have been taken from [CT20].

The first two plots with respect to $p_{\text{T,jet-2}}$ and $\Delta\eta_{\text{jj}}$ show good agreement with the previous theoretical prediction from the Royon, Market, Kepka (RMK) model, successfully reproducing previous results. The vertical displacement between results can be explained using a different value of SP; the important part is the dynamical behaviour. This is really important since it indicates that the implementation and setup used were correctly implemented. Let us discuss now the agreement with experimental data.

For the variable $p_{\text{T,jet-2}}$, no strong dependence was expected. Both experimental and theoretical predictions have only a mild correlation with the variable, although they seem to have opposite behaviours. The error bars on the experimental values advise caution however.

The variable $\Delta\eta_{jj}$ presents the most concerning disagreement between experiment and theory. Naïvely, in the Mueller-Tang analysis one would expect more forward jets to present a larger fraction of CSE events since more forward jets imply lower exchanged momentum t. Indeed, experimental data *mostly* agrees with this. Theoretical predictions, however, seem to contradict this intuition and point towards a decrease of CSE events as the exchanged momentum goes down.

RMK used a different MC event generator for showering, which is probably why the $\Delta\phi_{jj}$ plot is not equal. RMK used HERWIG in conjunction with NLOJet++ to make their predictions, which may explain the qualitative difference with our results from PYTHIA. The $\Delta\phi_{jj}$ distribution is the most affected by the choice of MC simulator since at parton level it is just a Dirac delta as $\delta(\phi_{\text{jet-1}} - \phi_{\text{jet-2}})$.

In regards to the difference with the experimental values it is possible that a full model of the things that are on shakiest ground are required. It was attempted to use NLO QCD to reproduce the data, however, no noticeable change in the $\Delta\phi_{jj}$ was observed (cf. Sec. 4.2.2). Alternatives not contemplated in this work include rescattering, NNLO QCD or alternative soft physics models other than SCI (cf. Sec. 4.8).

4.2.2 NLO QCD

The leftmost experimental point in the $\Delta\phi_{jj}$ in Fig. 4.1 seems to indicate that effects beyond LO might be important. In order to check for effects from fixed order NLO QCD POWHEG is used to generate such events to feed them via Les Houches Event Files [Boo+01] to PYTHIA 8 in order to generate the showering. For the NLO simulations the CP3 tune from the CMS Collaboration [Col19] is used as CP1 is only meant to be used for LO QCD.

Since event generation at NLO is much more costly than at LO, count number is reduced in comparison with other plots. For this reason Gaussian error bands ($\pm\sqrt{N}$) are included in the plots in Fig. 4.2.

Not much changes in terms of results, and the behaviour with respect to $\Delta\phi_{jj}$, which was the original motivation for moving to NLO, is nearly identical. My hypothesis is that we face a selection effect. Since there is a requirement for $p_{\text{T,jet}} > 40\text{GeV}$, when there are three hard emissions, the third jet will carry part of the momentum that would normally be carried by one of the leading jets. This in turn can prevent both leading jets from making the cut, creating a bias towards two jet events.

Other things that seem to point in that direction is that the NLO QCD distribution with respect to $\Delta\phi_{jj}$ is more peaked than at leading order (cf. Sec. B.3), pointing towards events with no back to back jets not making the cut. Three jet events are more likely to not be back to back in $\Delta\phi_{jj}$.

Lastly, the QCD cross section distribution with respect to $p_{\text{T,jet-2}}$ is a bit less pronounced for lower values in NLO than in LO, pointing again to lower $p_{\text{T,jet-2}}$ events being discriminated more strongly than at LO.

In terms of f_{CSE} the distributions seen in Fig. 4.2 are mostly unchanged. $\Delta\eta_{jj}$ still has the downwards trend seen at LO, although it seems to have a more moderate incline. This small shift can also be seen in the QCD cross section distribution with respect to $\Delta\eta_{jj}$ in Sec. B.3.
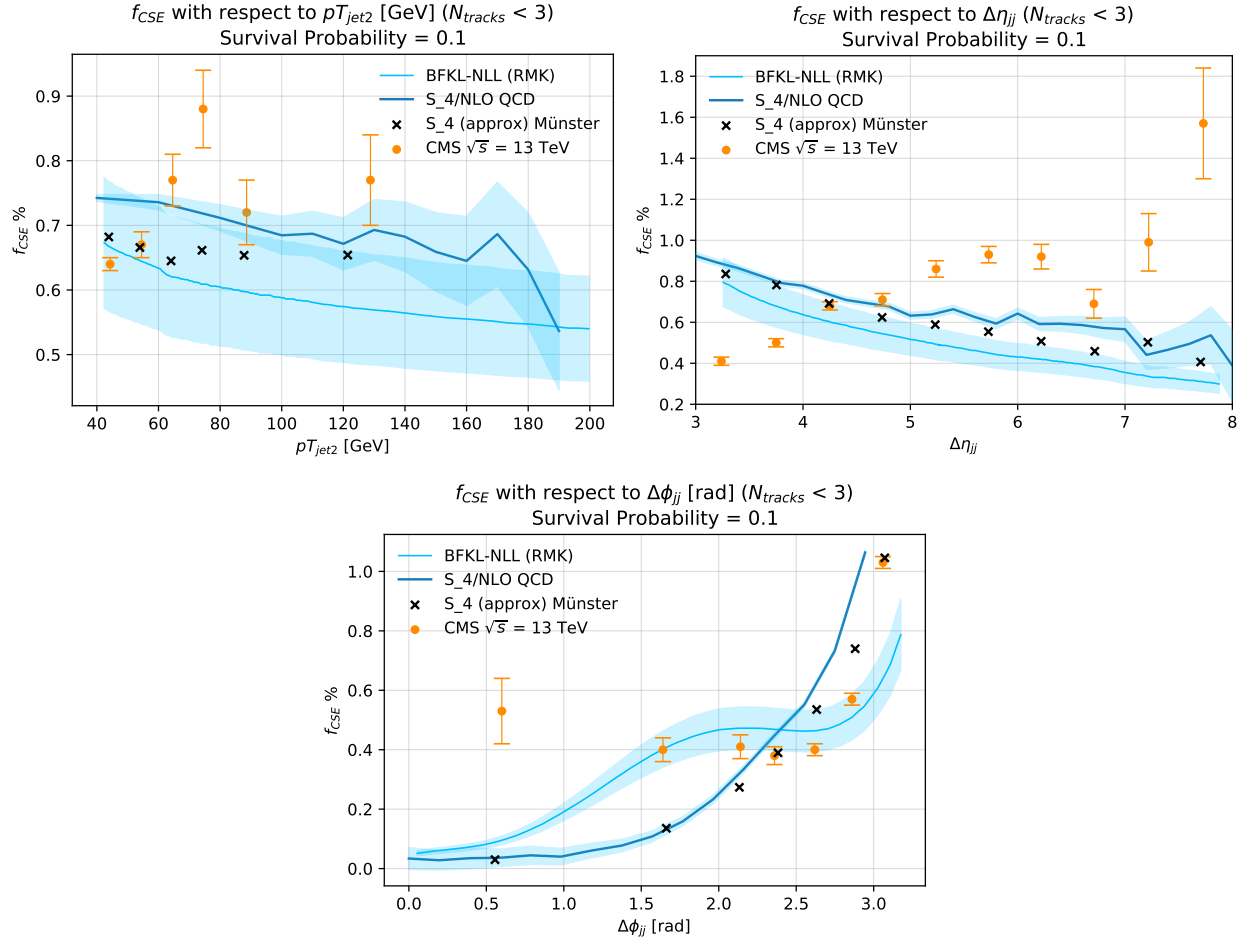
Figure 4.2: Differential plots of f_{CSE} with respect to $p_{\text{T,jet-2}}$, $\Delta\eta_{\text{jj}}$ and $\Delta\phi_{\text{jj}}$ using the tune CP1 from [Col19] but setting the parameter `SpaceShower:rapidityOrder = on` for CSE events. Experimental data represented with orange dots, a previous prediction by RMK in light blue, LO QCD indicated by black crosses and NLO QCD in dark blue line. Corresponds with Tables 2, 3 and 4 or Figure 8 in [CT20].

4.2.3   7 TeV

In order to see how well our predictions hold for lower values of CME the analysis from [Col17] is also replicated. It also involves studying Mueller-Tang jets to extract the fraction of CSE events. The kinematical limits are slightly different, although this should have no impact (cf. Sec. B.2), but the analysis on $\Delta\eta_{\text{jj}}$ is performed in three brackets of $p_{\text{T,jet-2}}$ instead of for all values as for 13 TeV, see Fig. 4.4. The kinematical limits changed are

$$|\eta_{\text{jet-1,jet-2}}| > 1.5 \tag{4.7a}$$

$$\text{Jet finder Radius} = 0.5 \tag{4.7b}$$

For these simulations we use ISR and SP = 0.1 [KMR10], which was the value predicted to correct parton level results (cf. Sec. 4.3), i.e. with no ISR. Predictions are largely unaffected because at 7 TeV the effect of ISR is much less noticeable than at 13 TeV.
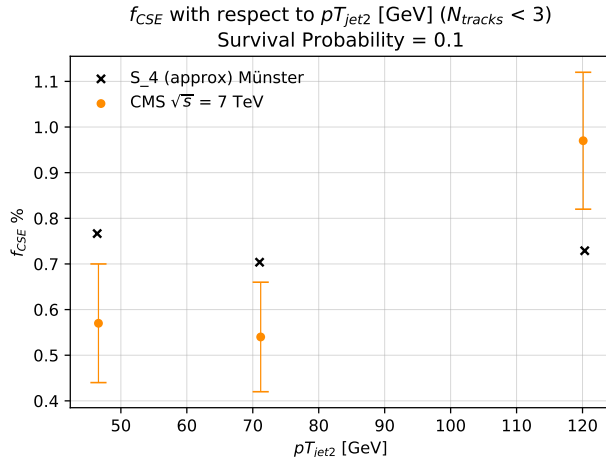
Figure 4.3: General differential f_{CSE} with respect to $p_{\text{T,jet-2}}$. Experimental data represented with orange dots and our results indicated by black crosses. Corresponds to Table 2 or Figure 9 in [Col17].

The results are consistent with what is found for the main analysis at 13 TeV; the order of magnitude obtained for f_{CSE} is correct, but the exact dynamical behaviour with respect to $\Delta\eta_{\text{jj}}$ shows the same disagreement.

4.2.4 Tevatron

Although these results can be studied on their own, the original motivation was to reproduce the predictions for Tevatron in [Che+09; KMR10], which used the same BFKL cross section parametrization that we use but achieve a much better agreement with experimental data than the one we observe in our own simulations. For this purpose the [aC98] experimental setup is replicated while following the simulation prescription of [Che+09; KMR10]. The results can be seen in Fig. 4.5. Notable differences with respect to other simulations in this work: no ISR or FSR is considered, SP has a value of 0.1 and the beams are $p\bar{p}$.

The most remarkable thing here might be the lack of agreement despite using the same BFKL cross section formula as the previous work. There are two main differences when it comes to how the results were obtained. First of all, they used HERWIG in combination with NLOJet++ while we use PYTHIA 8. HERWIG is also the MC used by the D0 Collaboration [aC98] when they observed that showering creates an upwards trend of f_{CSE} as $\Delta\eta_{\text{jj}}$ increases. The previous result using NLL BFKL also considered NLO QCD while we stayed at LO QCD, although this should not have an impact large enough on its own (cf. Sec. 4.2.2).

If we study these plots on their own, we find a similar situation to the main results: the order of magnitude of predictions matches that of experimental data, albeit with a bit more disagreement in the low-pT graph, but the dynamical behaviour does not agree.
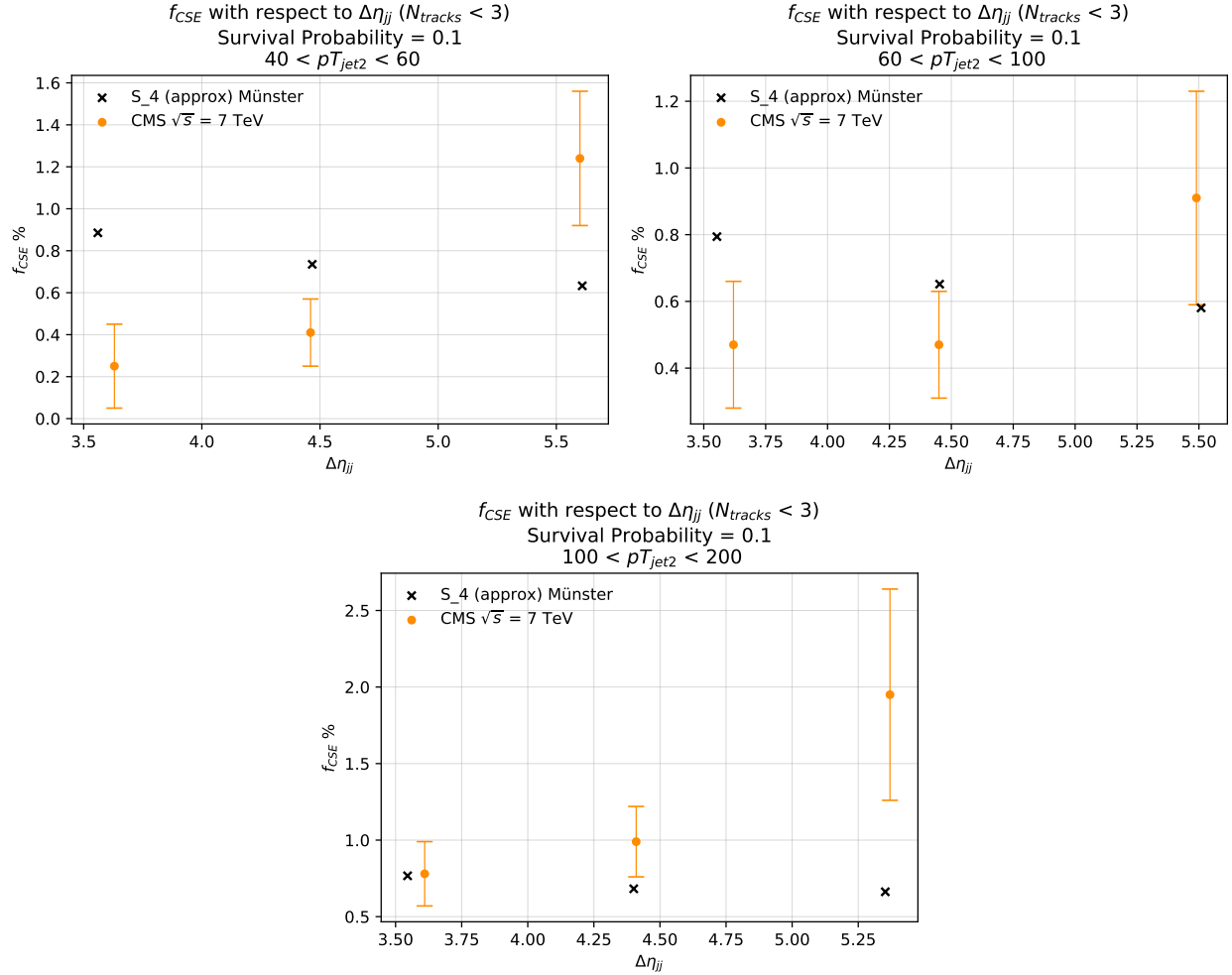
Figure 4.4: Double differential plots of f_{CSE} with respect to $\Delta\eta_{\text{jj}}$ for different ranges of $p_{\text{T,jet-2}}$. Experimental data represented with orange dots and our results indicated by black crosses. Corresponds with Table 3 or Figure 10 in [Col17].

4.3 Survival Probability

SP predictions [KMR10] pointed towards a value of 0.03 at the LHC. In this work, however, a value of 0.1 is used since it provides a better fit to experimental data. This seeming contradiction can be resolved by examining the definition of SP a bit more carefully.

The survival probability is the total probability that not one of several effects which are not simulated destroys the gap. Therefore, the more gap-destroying effects are explicitly simulated, the less impact this parameter will have: if we assume two survival probabilities, SP_1 and SP_2, where SP_2 needs to compensate less effects than SP_1, then

$$\left. \begin{array}{l} SP_1 = \prod_i^n \left(1 - \mathcal{P}_i\right) \\ SP_2 = \prod_i^{n-1} \left(1 - \mathcal{P}_i\right) \end{array} \right\} \implies SP_2 > SP_1. \qquad [4.8]$$

The previous theoretical prediction was made using the Eikonal Model, which does not take into account any kind of ISR or FSR. Following the prescription that SP encapsulates all

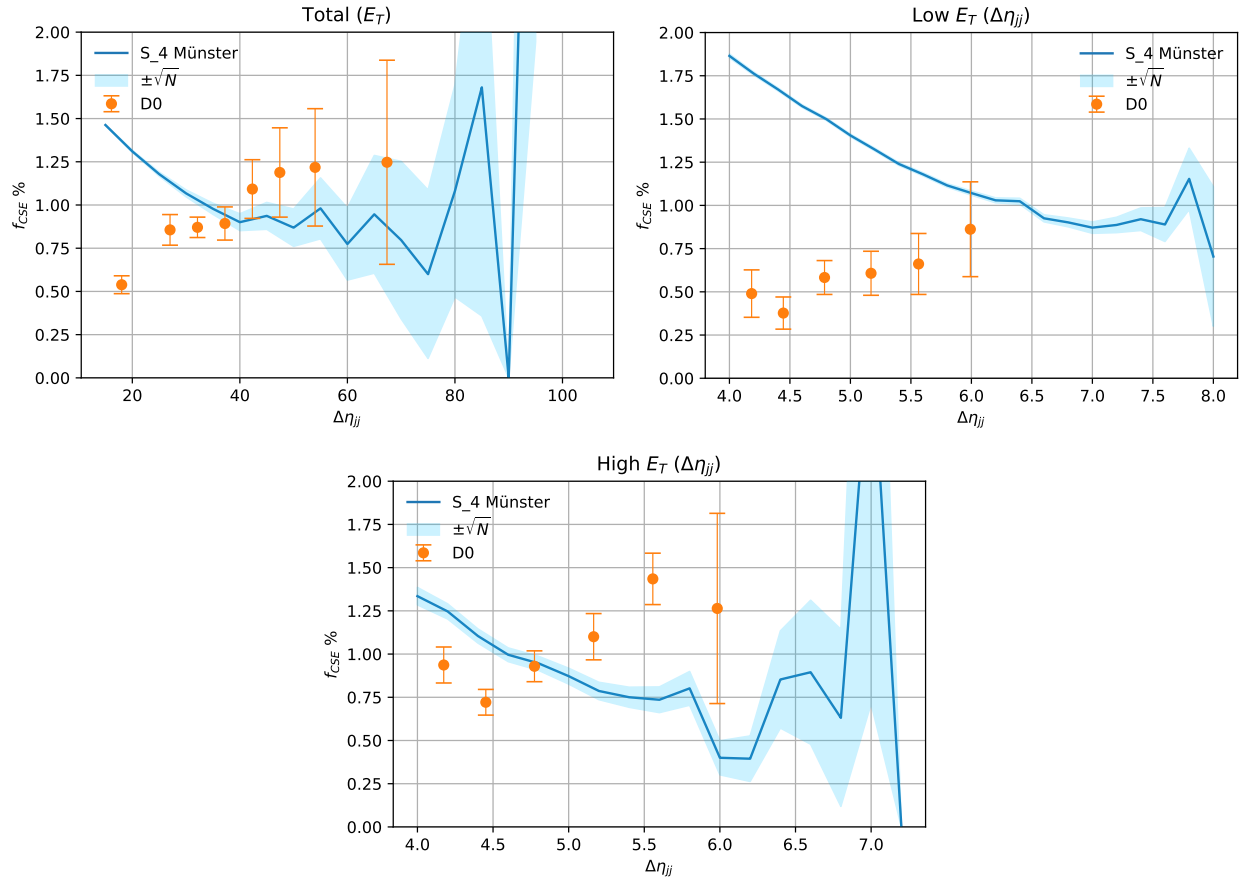Figure 4.5: Differential plots of f_{CSE} with respect to $p_{\mathrm{T,jet\text{-}2}}$ and $\Delta\eta_{\mathrm{jj}}$ for two brackets of $p_{\mathrm{T,jet\text{-}2}}$. Results obtained at parton level while reproducing Tevatron experimental conditions.
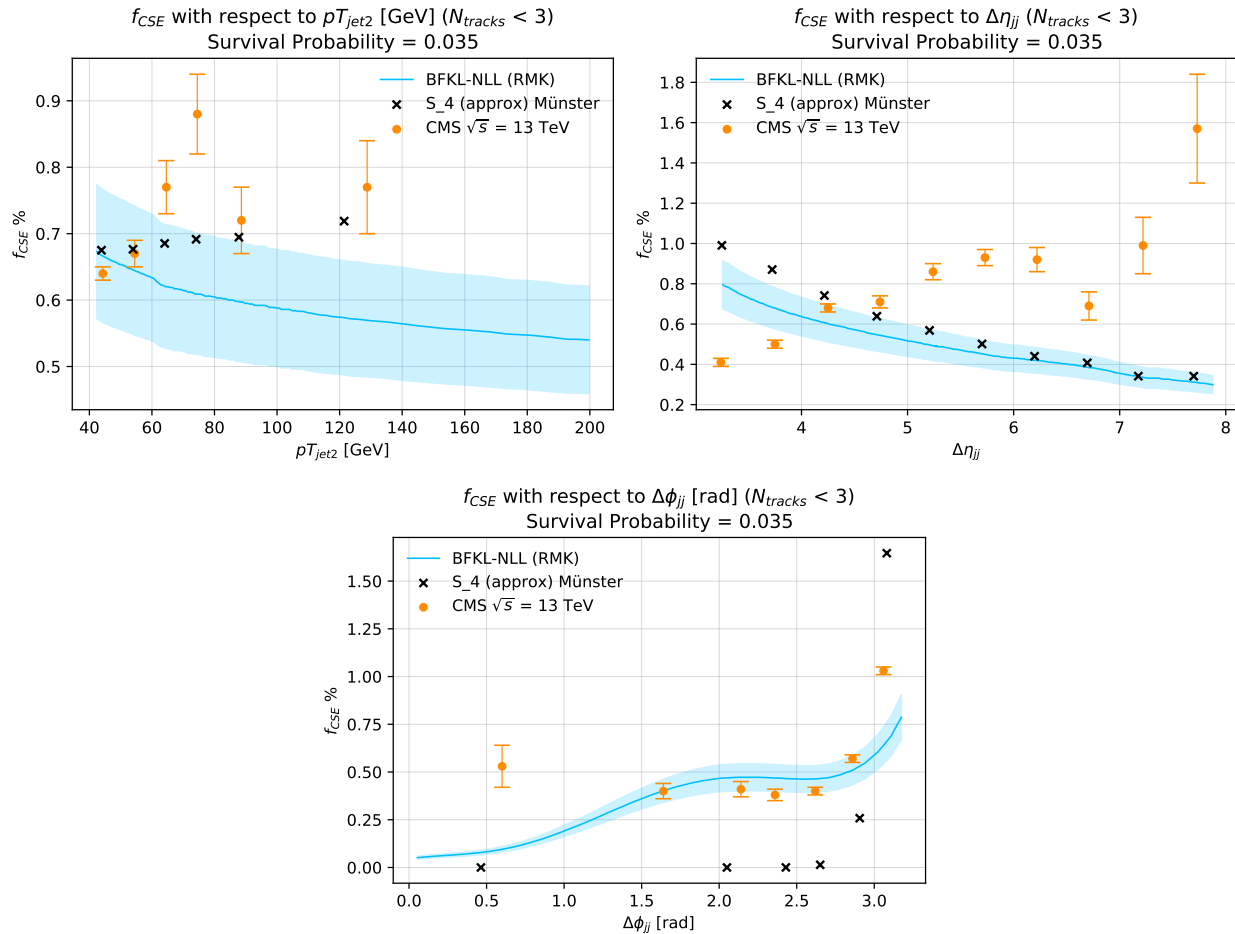
Figure 4.6: Differential plots of f_{CSE} with respect to $p_{\text{T,jet-2}}$, $\Delta\eta_{\text{jj}}$ and $\Delta\phi_{\text{jj}}$ after turning off ISR. The experimental data is in orange circles, a previous prediction by RMK is in blue and our results are in black crosses. The survival probability has been lowered to SP = 0.035. Corresponds with Tables 2, 3 and 4 or Figure 8 in [CT20].

effects that destroy the gap and should be increased if more effects are explicitly simulated, it is easy to figure out why we find a higher value for SP than predicted. Our simulations have been done with ISR turned on by default and *turning off ISR* for CSE made us adjust the value of SP to 0.035, as seen in Fig. 4.6, which is much more in line with previous predictions.

Ideally, if all effects were being taken care of the survival probability would be unnecessary, meaning that it would take a value of 1.

Two final remarks about turning ISR off: First of all, the $p_{\text{T,jet-2}}$ distribution is closer to the experimental values than before, and furthermore, it has reversed the trend, going from decreasing f_{CSE} as $p_{\text{T,jet-2}}$ increases to slightly increasing with $p_{\text{T,jet-2}}$. High $p_{\text{T,jet-2}}$ events probe the PDF at higher Q^2, which makes the algorithm take more steps (emissions) before reaching the cutoff. More emissions increase the probability that charged particles end up in the gap, reducing the *perceived* fraction of CSE events. Thus, turning off ISR generation causes that $p_{\text{T,jet-2}}$ behaviour to vanish.

The second notable thing is that the $\Delta\phi_{jj}$ distribution is very close to the partonic one, so FSR seems to have little impact in this observable, indicating that efforts to improve predictions may be better allocated in areas other than FSR.

The results with respect to $\Delta\eta_{jj}$ remain largely unaffected, so it is safe to discard ISR as the reason why the incorrect behaviour is obtained. FSR can also be indirectly discarded since it was shown in $\Delta\phi_{jj}$ to have very little effect in displacing the jets.

4.4 PDF Influence

There are many different PDF choices on the market, each with slight differences. In order to check how much PDF choice is impacting the results, a fake PDF is constructed. This PDF is completely flat, giving equal probability of finding all types of partons at all values of x when probed at any scale Q^2. Then f_{CSE} is calculated using the PDF indicated for the CP1 tune from the CMS Collaboration [Col19] and this fake PDF. Additional plots can be seen in Sec. B.2.
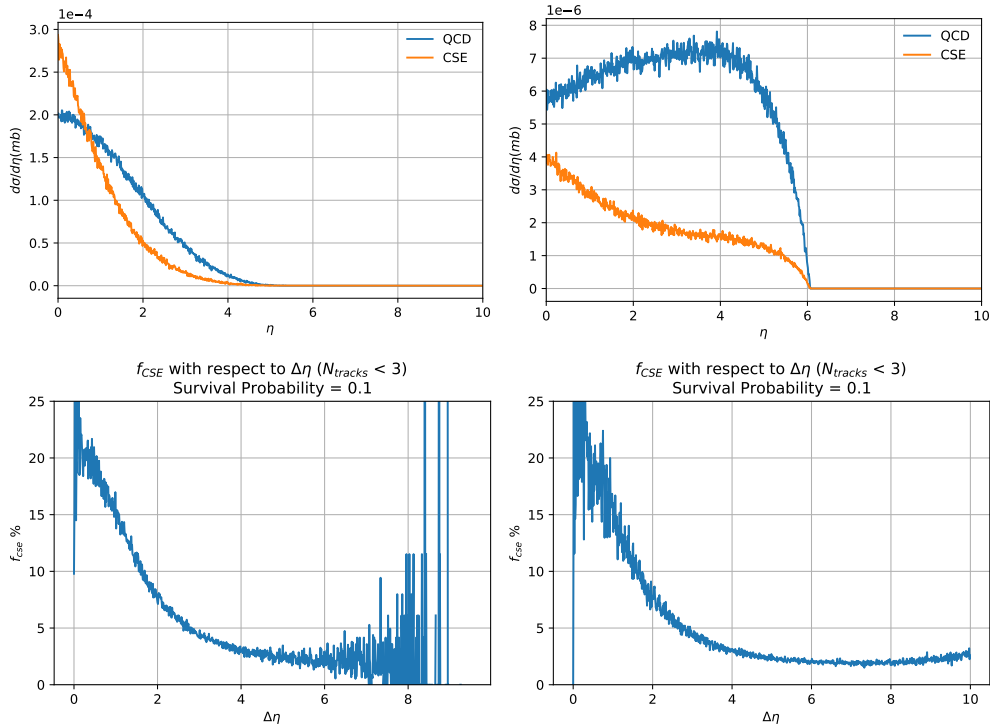
Figure 4.7: Cross section and f_{CSE} distributions with respect to $\eta = \Delta\eta_{jj}/2$. At the top are the individual CSE and QCD distributions, while the bottom plots have the differential f_{CSE} obtained using the results on top. On the left are the results using a normal PDF, while the results on the right were obtained using a flat PDF. Kinematical conditions are the same as in the standard simulations.

Except for statistical uncertainties both the flat and normal PDFs result in the same f_{CSE}. Not only that, but *changing the kinematical limits also leaves f_{CSE} invariant*(cf. Sec. B.2). A closer inspection of the definition of f_{CSE} reveals how certain changes may cancel out.

If we have an observable $f(x)$ that is the ratio between two different distributions with the same variable dependency, $H(x)$ and $G(x)$, such that

$$f(x) = \frac{H(x)}{G(x)}, \qquad [4.9]$$

then it is possible to apply variations dependent of another variable, y, that is independent from x, and leave our initial observable invariant, as long as both $H(x, y)$ and $G(x, y)$ have the same factorizable dependence with y.

To see this, suppose that $H(x, y) = h_1(x)h_2(y)$ and $G(x, y) = g_1(x)g_2(y)$. Because we are only observing with respect to the x variable it is necessary to integrate over all possible values for y for each distribution:

$$
\begin{aligned}
f(x) &= \frac{\int H(x, y)\, dy}{\int G(x, y)\, dy} \\
&= \frac{h_1(x) \int h_2(y)\, dy}{g_1(x) \int g_2(y)\, dy}.
\end{aligned}
\qquad [4.10]
$$

If we were to cull the phase space, such that $\Delta(x, y)$ gives the contribution of each point towards the total then

$$
\begin{aligned}
f(x) &= \frac{\int H(x, y)\Delta(x, y)dy}{\int G(x, y)\Delta(x, y)dy} \\
&= \frac{h_1(x) \int h_2(y)\Delta(x, y)dy}{g_1(x) \int g_2(y)\Delta(x, y)dy}.
\end{aligned}
\qquad [4.11]
$$

If both $H(x, y)$ and $G(x, y)$ have the same dependency with respect to y, that is, $h_2(y) = g_2(y)$, then the observed distribution $f(x)$ is independent of the cull function $\Delta(x, y)$.

The previous formalism can be applied to the change in kinematical limits straight away, however, we need to prove that this can be applied to PDFs. First of all we need to move to parton level variables.

Now, using rapidity properties (differences of rapidities are Lorentz invariant and changes of frames of references equates to addition), considering that in the CM frame of reference $\hat{\eta}^1_{CM} = -\hat{\eta}^2_{CM} = \hat{\eta}_{CM}$, and that the boost required to go from CM frame to detector frame is γ, the laboratory frame variables are defined as

$$
\begin{aligned}
\eta^1 &= \hat{\eta}_{CM} + \gamma \\
\eta^2 &= -\hat{\eta}_{CM} + \gamma
\end{aligned}
$$

$$
\begin{aligned}
\hat{\eta}_{CM} &= \frac{1}{2}(\eta^1 - \eta^2) \\
\gamma &= \frac{1}{2}(\eta^1 + \eta^2).
\end{aligned}
$$

It is easy to argue that $\hat{\eta}_{CM}$ and γ must be independent variables. If they were not, all the processes dependent on the center of mass energy would be altered simply by changing our frame of reference, without changing the system itself. Representing the independent

variables in vector form verifies that the transformation from $(\hat{\eta}_{CM}, \gamma) \rightarrow (\eta^1, \eta^2)$ preserves the independency:

$$Base := \begin{bmatrix} \hat{\eta}_{CM} \\ \gamma \end{bmatrix} \implies \vec{\eta^1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \vec{\eta^2} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad [4.12]$$

Since $\vec{\eta^1} \cdot \vec{\eta^2} = 0$ we still have linearly independent variables.

Notice how the variable $\Delta\eta_{jj}$ is $\Delta\eta$ at hadron level, which in turn is related to $\hat{\eta}_{CM}$ by $\Delta\eta = 2\hat{\eta}_{CM}$. Additionally, if we are talking in the context of detector variables we need to take twice the absolute value, as detectors cannot distinguish between forward and backward scattering jets. We have that for a hadronic process:

$$d\sigma_{\text{Hadronic}} = \sum_{ij} \int dx_1 \, dx_2 \, f_i(x_1) \, f_j(x_2) \, d\hat{\sigma}_{\text{partonic}}(\hat{s}, \hat{t}). \qquad [4.13]$$

For the following change of variables

$$dx_1 dx_2 d\hat{t} = \frac{\hat{s}}{s} d\eta_1 d\eta_2 d(p_T^2) = \frac{2\hat{s}}{s} d\Delta\eta d\gamma d(p_T^2), \qquad [4.14]$$

the Jacobian determinant is,

$$J\left[\eta_1, \eta_2 \rightarrow \Delta\eta, \gamma\right] = \begin{vmatrix} \frac{d\eta_1}{d\Delta\eta} & \frac{d\eta_1}{d\gamma} \\ \frac{d\eta_2}{d\Delta\eta} & \frac{d\eta_2}{d\gamma} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} = 2. \qquad [4.15]$$

Putting everything together,

$$\frac{d\sigma_H}{d\Delta\eta} = \sum_{ij} \int \frac{2\hat{s}}{s} \, d\gamma \, d(p_T^2) \, f_i(x_1) \, f_j(x_2) \, \frac{d\hat{\sigma}}{d\hat{t}}(\hat{s}, \hat{t}). \qquad [4.16]$$

The variable $\hat{t}$ necessarily has to be independent of the boost with respect to the laboratory system, γ. Whether the variable $\hat{s}$ is also not dependent on the boost is not so trivial.

If we have a $2 \rightarrow 2$ process then the relation $p_1 + p_2 = p_3 + p_4$ has to hold. Taking into account the relations:

$$P^\mu = (m_T cosh(\eta), \vec{p}_T, m_T sinh(\eta)), \qquad \text{with } m_T = \sqrt{m^2 + p_T^2}.$$

we arrive at the following relation between Bjoken-x and the rapidity variables:

$$x_1 = \frac{p_T}{\sqrt{s}} \left(e^{\eta_3} + e^{\eta_4}\right) = \frac{2p_T}{\sqrt{s}} e^\gamma cosh(\hat{\hat{\eta}}) \qquad [4.17a]$$

$$x_2 = \frac{p_T}{\sqrt{s}} \left(e^{-\eta_3} + e^{-\eta_4}\right) = \frac{2p_T}{\sqrt{s}} e^{-\gamma} cosh(\hat{\hat{\eta}}) \qquad [4.17b]$$

Now it should be clearer that $\hat{s}$ loses its dependency with the boost when we operate with our chosen variables, as one x's contribution cancels the other as

$$\hat{s} = x_1 x_2 s = 4p_T^2 e^{\gamma-\gamma} cosh^2(\hat{\eta}) \qquad [4.18a]$$

$$= 4p_T^2 cosh^2(\hat{\eta}) \qquad [4.18b]$$

$$= 2p_T^2 (cosh(\Delta\eta) + 1). \qquad [4.18c]$$

Similarly $\hat{t}$ variable is found to also be independent from γ as

$$\hat{t} = (p_1 - p_3)^2 = -p_T^2(1 + e^{-2\hat{\eta}}) \qquad [4.19\text{a}]$$

$$= -p_T^2(1 + e^{-\Delta\eta}) \qquad [4.19\text{b}]$$

$$= -2p_T^2 e^{-\hat{\eta}} cosh(\hat{\eta}) \qquad [4.19\text{c}]$$

Therefore, if we take as valid that the invariant partonic cross section, $d\hat{\sigma}/d\hat{t}$, only depends on $\hat{s}$ and $\hat{t}$, then it can be factorized out of the $d\gamma$ integral.

There is still another integral to be resolved, the $d(p_T^2)$ one. One can approach this two different ways. First one is to say that we are only observing a narrow section of the possible p_T space, and therefore we could take it as a constant, making it possible to factorize everything again. In this case the cancellation would be only approximate.

The alternative is realising that as $2 \to 2$ processes have a $1/p_T^4$ dependency, regardless of them being CSE or normal processes, then the condition that dependency has to be the same is automatically fulfilled.

4.5 Process Initiators

In this section, the cross section distributions are separated based on which particles enter in the initial state. Only three kinds of process initiators are considered, gg, gq and qq. The initiators are grouped based on the closest match with that classification, so a process $q\bar{q} \to gg$ would be included in the qq category. Fig. 4.8 shows the distributions with respect to N_{tracks}, $p_{\text{T,jet-2}}$, $\Delta\eta_{\text{jj}}$ and $\Delta\phi_{\text{jj}}$.

For BFKL with LO impact factors, the cross section for all three processes should have the same dynamical behaviour and differ only in a colour factor. Indeed it is easy to see that gg is approximately $9/4$ times as large as qg and $(9/4)^2$ times as large as qq. What we see however, is not the pure cross section, but its convolution with the PDFs.

Previous works [aC98] found a strong correlation between rising f_{CSE} and qq initiated processes as can be seen in Fig. 4.9. Since the cross sections couple equally to gg, qg and qq, this result has to be caused by PDFs.

From Eq. 4.18 one sees that as $\Delta\eta_{\text{jj}}$ increases, $x_1 x_2$ also increases. Gluon distribution functions decrease rapidly for large values of x, which makes f_{CSE} drop since it is being divided by the inclusive cross section, which takes into account all contributions. In contrast, the contributions that have at least one q get a boost as their proportional contribution raises.

At the end of the day, fluctuations between the three have to cancel, since PDF dependence has to cancel in the ratio as was discussed in Sec. 4.4. Thanks to Cristian Baldenegro for the fruitful discussion about these plots.

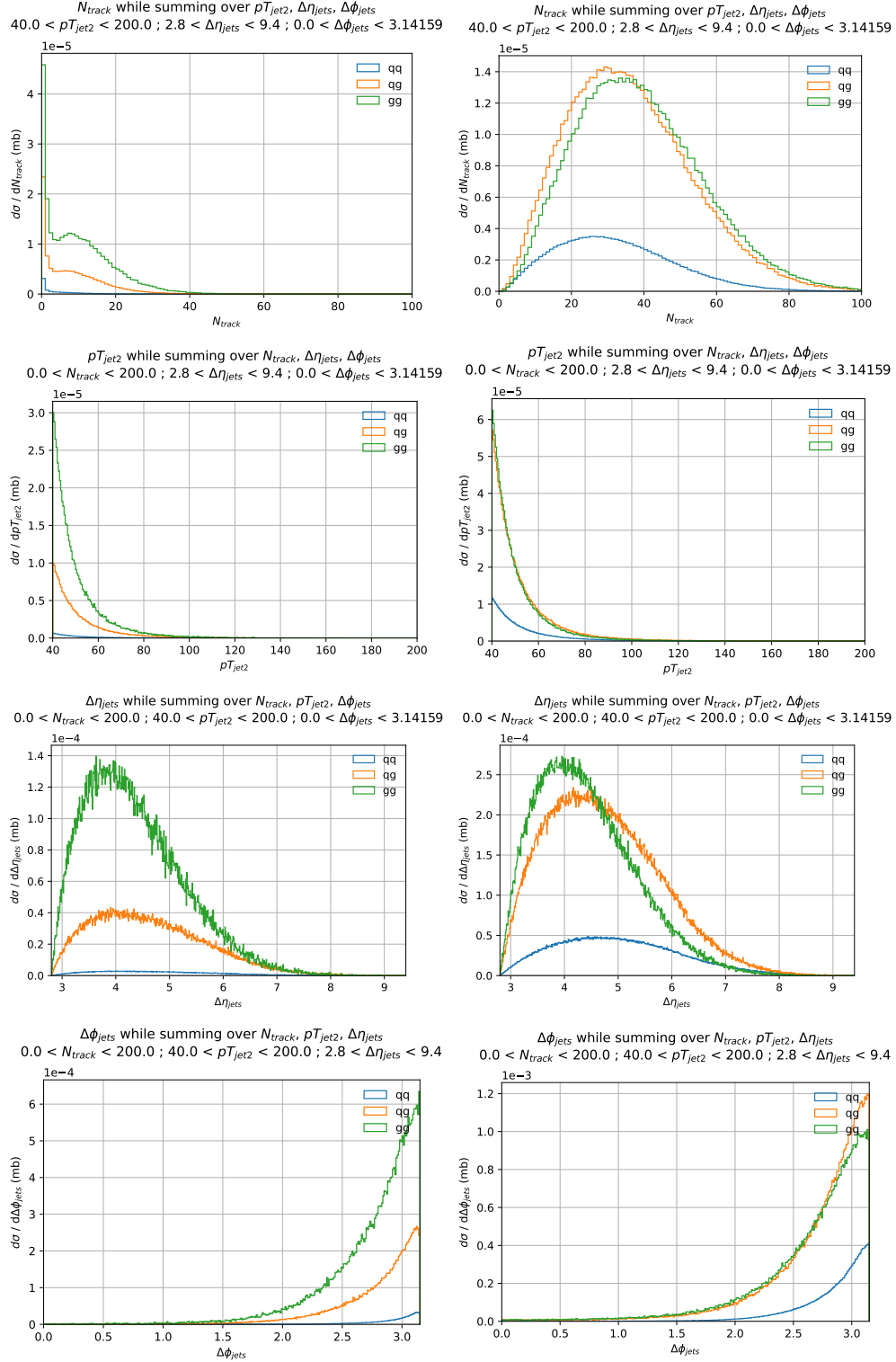Figure 4.8: Differential cross section with respect to N_{tracks}, $p_{\text{T,jet-2}}$, $\Delta\eta_{\text{jj}}$ and $\Delta\phi_{\text{jj}}$ displaying individual contributions from different initiators for CSE on the left and Born-level QCD on the right.
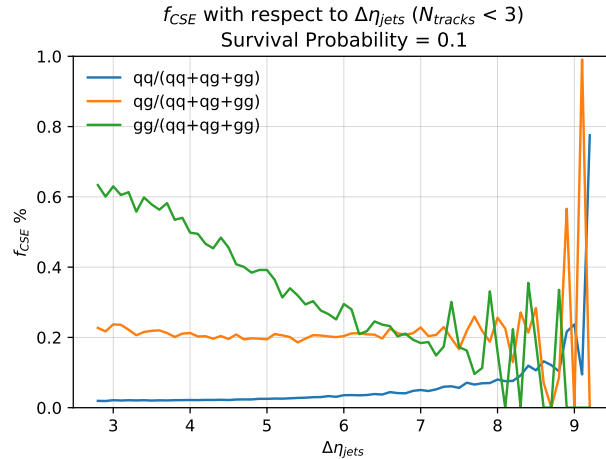
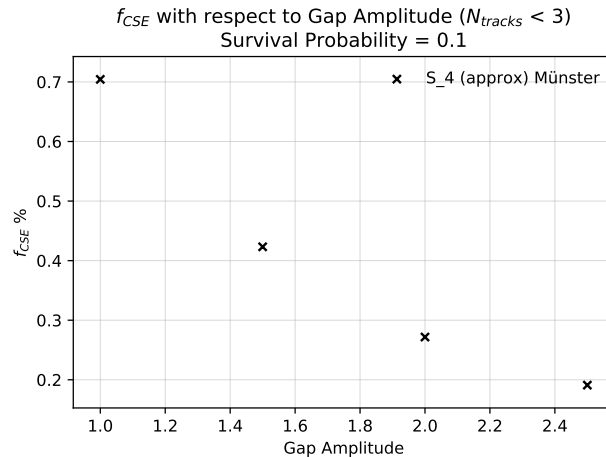Figure 4.9: Individual contributions to f_{CSE}, taking the denominator to be the inclusive cross section.

Figure 4.10: f_{CSE} as a function of gap amplitude.

4.6 Variable Gap Amplitude

Given Eq. 2.14, the possibility of forcing events to be more forward was considered. In theory this should increase the number of CSE events, as the kinematical limits push events further and further into the Regge region. In order to carry this out, the gap amplitude from Eq. 4.1c was extended and total f_{CSE} plotted against it, always taking care of keeping the jets outside of the gap region. As can be seen in Fig. 4.10 the result was the opposite of what was expected.

ISR was found to be mostly isotropic between the jets which means that as the gap region is increased, so is the amount of ISR particles that get counted inside of it. To the best of my knowledge this was also seen in experimental data (in private discussions with Christophe Royon and Cristian Baldenegro).

Alternatively, increasing just the minimum distance between the jets and the gap, while
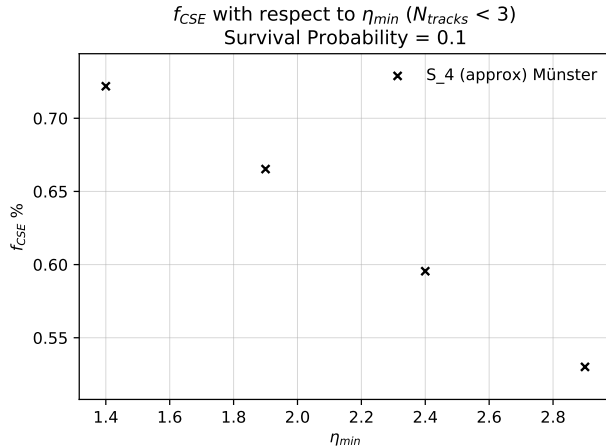
Figure 4.11: f_{CSE} as a function of $\eta_{\mathrm{jet,min}}$.

leaving the gap area intact to prevent flooding from ISR, was considered in Fig. 4.11. This is, increasing the lower limit for η_{jet} from Eq. 4.1c while leaving η from Eq. 4.3b intact.

This time f_{CSE} decreases simply because, to all effects, this is roughly equivalent to cutting the left part of a regular $\Delta\eta_{\mathrm{jj}}$ plot and then averaging over all values. Since in our previous simulations f_{CSE} with respect to $\Delta\eta_{\mathrm{jj}}$ decreases as $\Delta\eta_{\mathrm{jj}}$ rises, averaging only for high values of $\Delta\eta_{\mathrm{jj}}$ naturally results in lower values of total f_{CSE}.

4.7 Sources of Error

Three sources of error *from the MC* are considered: factorization scale, renormalization scale and PDF replicas. In theory observables should not depend on either factorization or renormalization scales. However that is only strictly true if *all* diagrams to all orders are considered. Since this is not really feasible, a multiplicative factor for each is introduced and varied to verify the degree of independence of f_{CSE} with respect to these scales. The factorization scale enters at the PDF evaluation as the the probing scale Q^2. The renormalization scale enters in the renormalization scheme for the strong coupling constant. The multiplicative factor for the factorization scale is denoted by μ_F and the one for the renormalization scale by μ_R, such that

$$\mu_F \rightarrow f_i(x, \mu_F\, Q^2), \qquad\qquad [4.20\mathrm{a}]$$

$$\mu_R \rightarrow \alpha_s(\mu_R\, Q^2). \qquad\qquad [4.20\mathrm{b}]$$

The PDF replicas come because extracting the values from scattering data carries uncertainty. In order to give a measure of this uncertainty a hundred replicas are given, each with slight variations with respect to the central one, which is what is used by default.

Despite what was calculated in Sec. 4.4, the error from PDF does not cancel. This is because it was calculated assuming independent variations in numerator and denominator. That is, propagating the resulting standard deviation through the f_{CSE} formula. The scaling errors are calculated by making the multiplicative factors μ_F and μ_R take the values 4 or 1/4

without mixing scale variations, i.e. if one changes the other stays with the default value of one. As we expected, the uncertainty coming from the renormalization scale *almost* vanishes. However, there is a very notable lack of cancelation for factorization scale uncertainties.
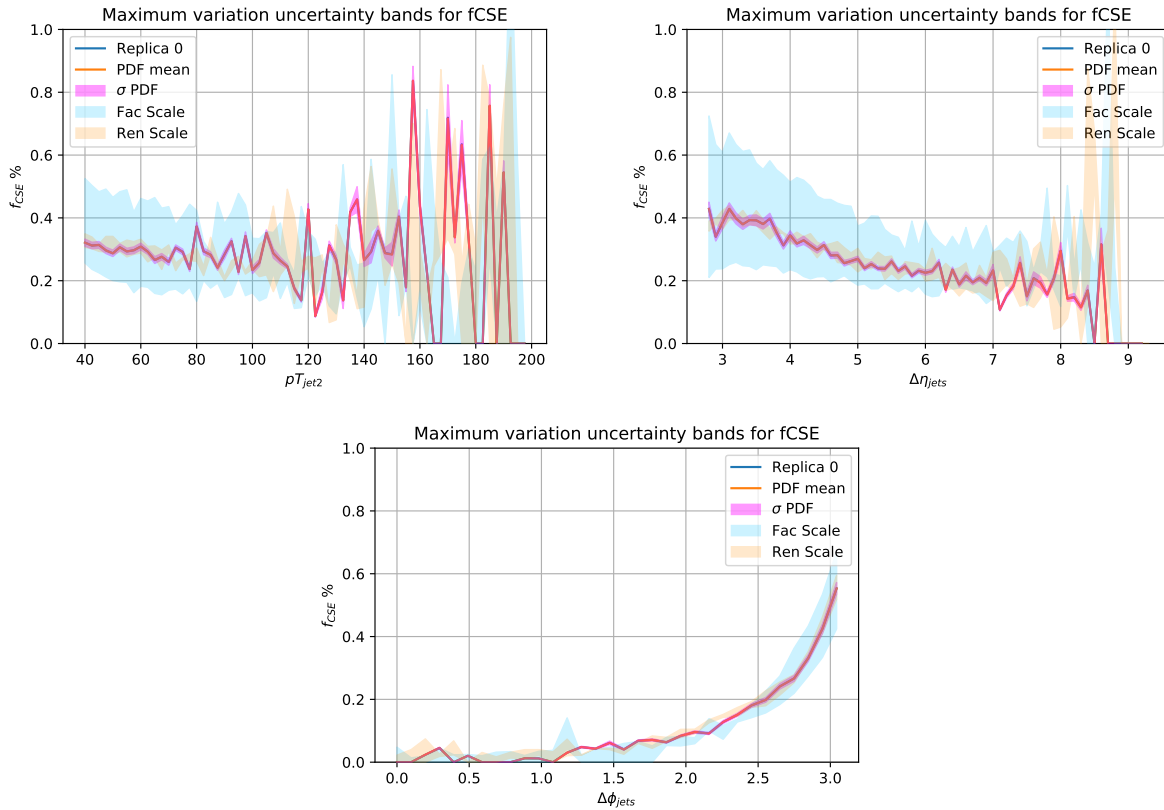

Figure 4.12: Differential plots of f_{CSE} with respect to $p_{\mathrm{T,jet\text{-}2}}$, $\Delta\eta_{\mathrm{jj}}$ and $\Delta\phi_{\mathrm{jj}}$ displaying variations with respect to the renormalization and factorization scales, as well as PDF uncertainties.

The reason why the factorization scale variation does not cancel like the renormalization one does is that it does not symmetrically affect numerator and denominator. After the partonic hard process PYTHIA 8 is at a scale $\mu_F\,Q^2$ and emits new particles until it reaches lower cutoff scale Q_0^2. By increasing μ_F, more emissions take place before halting the routine. This makes it so that high multiplicity events are the ones most affected by this change, as can be seen in Fig. 4.13.

However, increased emissions are not enough to explain the large variation seen. A secondary effect is also taking place: interference between the jet finder and ISR particles. Increasing $\mu_F\,Q^2$ not only increases particle multiplicity, but it also increases the momentum carried by said emissions. When those emissions fall within the cone of a jet, their momentum is added to it, which subsequently makes jets overcome the $p_{\mathrm{T,jet}} > 40\mathrm{GeV}$ cut. Since cross sections fall as $1/p_{\mathrm{T,jet}}^4$, a small displacement from the left side of the distribution towards higher $p_{\mathrm{T,jet\text{-}2}}$ values adds a lot of cross section to the events that make the cuts (cf. Fig. 4.8, $p_{\mathrm{T,jet\text{-}2}}$ plot).

Since the numerator in the f_{CSE} definition represents events with no particles in the
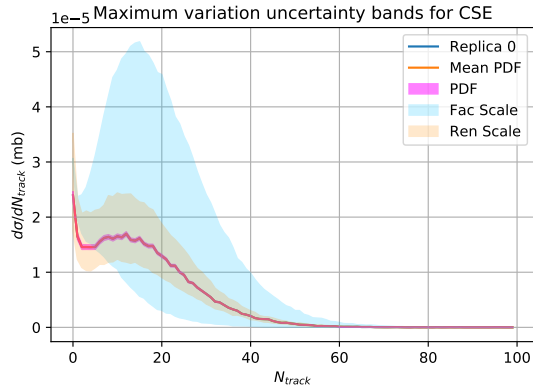
Figure 4.13: N_{tracks} distribution for CSE events and how it is affected by different variations from scales and PDF.

gap region, it is not affected, creating a discrepancy and preventing the cancellation of the variation. Turning ISR off confirms this hypothesis since then the factorization scale uncertainty cancels to a degree comparable with the renormalization scale (cf. Sec. B.4).

4.8 Soft Colour Interactions

The *Soft Colour Interactions* (SCI) [EEI17; EIR95; Rat98] model is noteworthy for being the only model to get the correct $\Delta\eta_{\text{jj}}$ in [Col17]. SCI is an alternative routine for the colour reconnection routine in PYTHIA; developed initially for PYTHIA 6 it is now integrated into PYTHIA 8 as part of this work (for implementation details see Appendix C). Whereas PYTHIA 8's default has the specific aim of reducing string length, SCI takes a different approach and treats all possible reconnections as equals, with a fixed probability of reconnecting any two dipoles that depends only on the overall event dynamics.

The basic element of SCI is the probability,

$$\mathcal{P} = p_0 \, \log(1 - \hat{t}/400)^b, \tag{4.21}$$

where p_0 and b are parameters of the model and $\hat{t}$ is the dynamical variable of the event, the transferred momentum for the $2 \rightarrow 2$ process. While there is some leeway in the values of both parameters, it is recommended to only change p_0, since b has been fit to data.

The default values given by the creators of the model are

$$p_0 = 0.7,$$
$$b = 0.1.$$

Using these values is how Fig. 4.14 was obtained. The events generated with this model are nearly guaranteed to destroy the gap instead of resulting in an effective gap survival of 0.1, as well as producing events with too many charged particles. For comparison, typically experimental charged multiplicity distributions peak at $N_{\text{tracks}} \sim 20$.

Trying different values of p_0 does not change the obtained distribution except for extremely small values of around $p_0 \sim 0.001$. To explain this we need to consider that for our
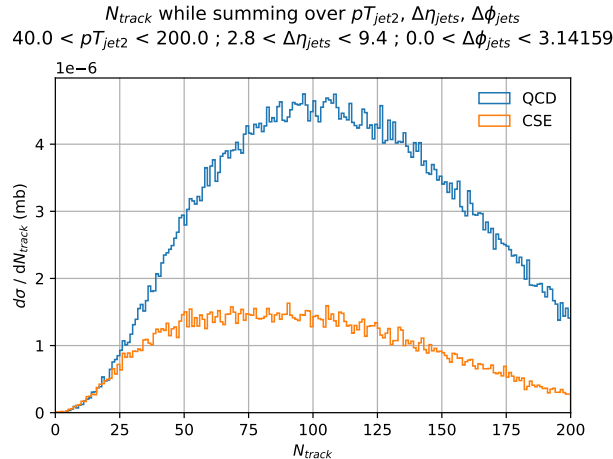
Figure 4.14: N_{tracks} distributions for CSE and QCD events using SCI with default parameter values. This plot displays the saturation result.

experimental setup $\hat{t} > 1600\,\text{GeV}^2$, which sets the factor $\log(1 - \hat{t}/400) \sim 1.05$ and growing very slowly. Therefore the probability of reconnection is, to a good approximation,

$$\mathcal{P} \sim p_0.$$

Because of how the model is built, if there are n colour dipoles, $n(n-1)/2$ attempts at reconnecting dipoles will be made, resulting in the average number of reconnections

$$\mathcal{P}\,\frac{n(n-1)}{2} \sim p_0\,\frac{n(n-1)}{2}. \tag{4.22}$$

We can now define the ratio of colour reconnections to number of dipoles, which gives a rough estimate of how many times each dipole is reconnected, as

$$R \sim p_0\frac{n-1}{2}. \tag{4.23}$$

Now it is easy to see that the number of particles, and therefore possible colour dipoles, scales with CMe, and so does the number of times each dipole gets reconnected. The reason p_0 has little to no effect is that in a typical event at 13 TeV, $n \sim 100$ and *each dipole* gets reconnected $\sim 50\,p_0$, resulting in the final state colour topology being completely random for the default value of $p_0 = 0.7$.

Comparing with PYTHIA 8's default colour reconnection, about $20 - 40$ reconnections happen for a typical event with number of dipoles $n \sim 100$. From Eq. 4.22, we obtain a value of $p_0 \sim 0.005$, which matches the critical value observed to exit the saturation regime.

Another difference is noted; for *pp* collisions at 13 TeV, PYTHIA 8's default CR reduces the amount of charged particles produced while SCI seems to always increase it. This may point towards one of the possible venues of improving on the model.

Special thanks to Andreas Ekstedt, Rikard Enberg and Gunnar Ingelman for extensively discussing their model with us.

4.9 Default Colour Reconnection + MPI

In order to have a direct comparison with SCI, the default colour reconnection scheme along with PYTHIA 8 MPI is also tested. The interest in finding a good description of data with MPI instead of using SP goes beyond finding a good agreement with data. Performing the same analysis as the one performed in experiments (cf. Sec. 4.1) is not feasible while using a static SP. Obtaining the complete charged multiplicity distribution enables us to perform the same kind of fit as the experimentalist use, as seen in Fig. 4.15 (cf. Sec. 4.1).
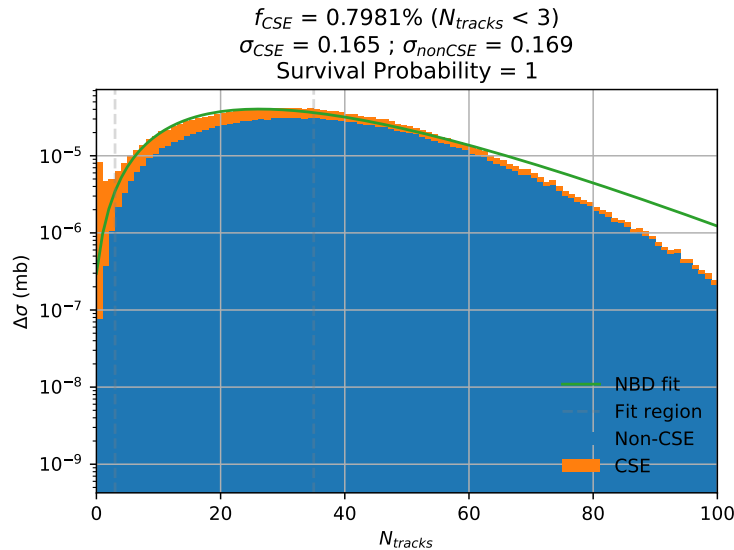

Figure 4.15: N_{tracks} joint distribution for CSE and QCD events in log scale, along with best fit to Negative Binomial Distribution, showing excess gap events with respect to background.

Now with the analysis being identical to the experimental one the results are shown in Fig. 4.16. By far the most notable thing from this analysis is that using PYTHIA 8 internal machinery for MPI, a reasonable result is obtained at all. The only remaining difference now between the simulation of CSE and regular QCD events is the parameter `SpaceShower:rapidityOrder` (on for CSE and off for QCD, cf. Sec. B.1), which is worth mentioning that it is not mandatory or even standard practice for PYTHIA tunes. Anyhow, rapidity ordering was not found to change the dynamic behaviour of f_{CSE}, but merely reduced it by half.

While this does not assure that MPI and CR in PYTHIA 8 is ready to be used as a substitute for theoretical predictions of SP, it is worth considering just for the benefit of reproducibility of analysis and having a more dynamic predictor than just a constant SP.

In regards to that last point, based on theoretical predictions SP is not expected to have a strong dependence on kinematical variables, only on CMe. However, the change required in order to reverse the $\Delta\eta_{\text{jj}}$ tendency is not a strong one, and small deviations can contribute towards better predictability. Indeed, the $\Delta\eta_{\text{jj}}$ behaviour seems a bit less decreasing, so while dynamical SP might not be the full answer, it is possibly part of it.
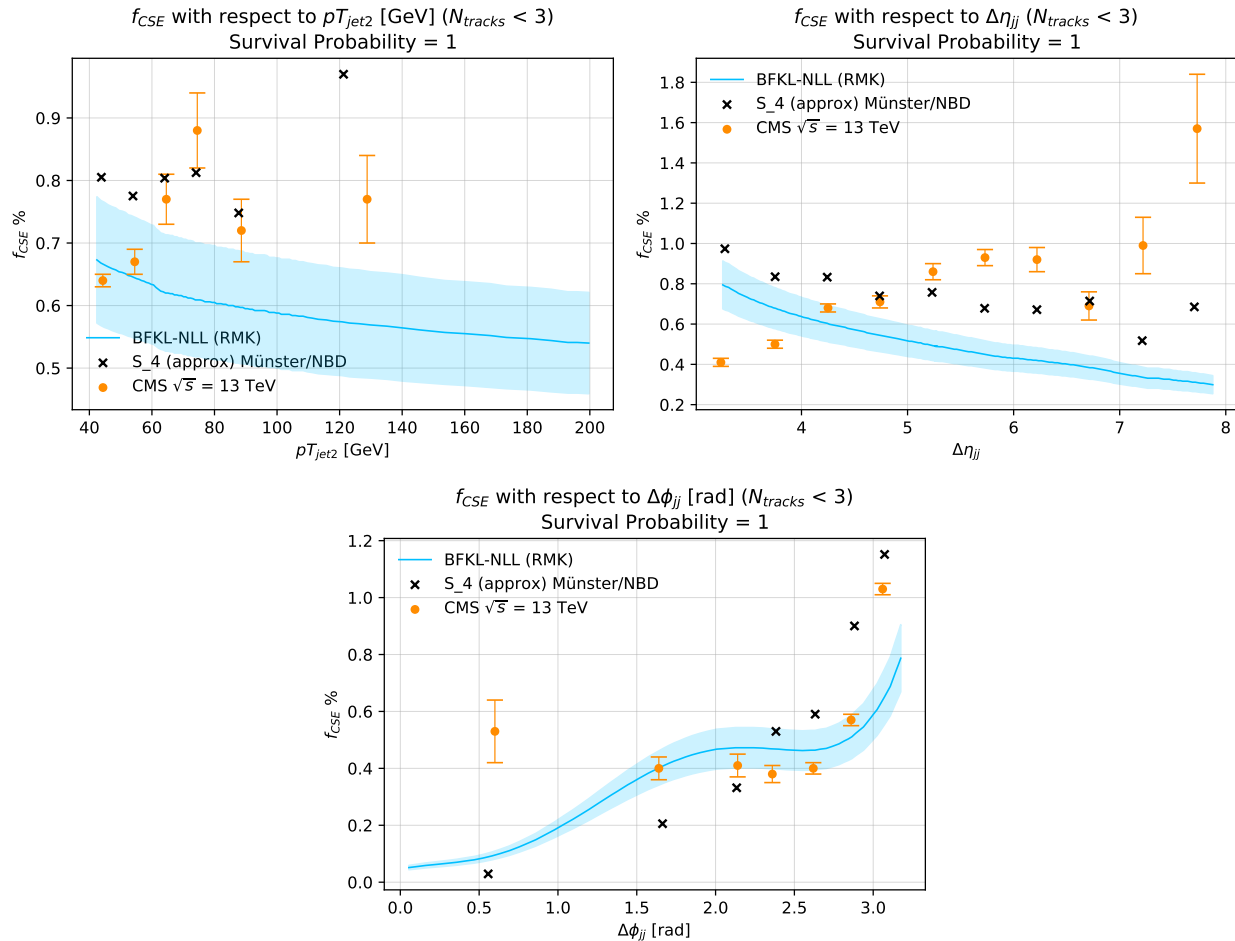
Figure 4.16: Differential plots of f_{CSE} with respect to $p_{\mathrm{T,jet\text{-}2}}$, $\Delta\eta_{\mathrm{jj}}$ and $\Delta\phi_{\mathrm{jj}}$ using the tune CP1 from [Col19] but setting the parameter `SpaceShower:rapidityOrder = on` for CSE events. The experimental data is in orange circles, a previous prediction by RMK is in blue and our results using PYTHIA 8 MPI and CR routines are in black crosses. Corresponds with Tables 2, 3 and 4 or Figure 8 in [CT20].

4.10 Challenges

It is worth remarking that literature studies on BFKL seem to be heavily influenced by the MC used. The authors of the original D0 Collaboration for Tevatron at 1.8 GeV [aC98] noted that showering in HERWIG altered the $\Delta\eta_{\mathrm{jj}}$ behaviour, getting an upwards trend, while the parton level results obtain the same downwards trend we observe. They mentioned that "parton showering for standard QCD processes reduces the jet $\Delta\eta$ separation at the jet level relative to the parton level". While this effect is observed when trying to reproduce Tevatron results (cf. Sec. B.5), it is not enough to reverse the trend.

Previous works by [KMR10; Che+09] using HERWIG in conjunction with NLOJet++ could very accurately reproduce f_{CSE} data from extracted from Tevatron by the D0 Collaboration [aC98], but obtained much worse results for [CT20] using PYTHIA 8 for QCD and HERWIG for CSE. This work, using PYTHIA 8 for CSE and QCD and the same

parametrization for the BFKL cross section as [KMR10; Che+09] was unable to reproduce the f_{CSE} predictions for the data from the D0 Collaboration at Tevatron [aC98], but agreed to a good extent with the same model's (RMK) predictions for LHC at 13 TeV [CT20].

The SCI model, using PYTHIA 6 obtained the correct behaviour with respect to $\Delta\eta_{\text{jj}}$ for LHC at 7 TeV [Col17], but when implemented into PYTHIA 8 did not get a good agreement with data without further tuning for LHC at 13 TeV.

The variability introduced by different Monte Carlo event generators, even when using the same underlying hard physics (namely BFKL), is an important factor that should be kept present when analysing generated data. Even changing the setup, such as CME requires the use of different tunes, which may alter in no obvious ways how the different routines inside the MC generator shape the distributions, thus altering the dynamical behaviour observed. Perhaps the most convincing argument towards the importance of *minute* details of different MC is that the difference need not be very significant at all, since just a small deviation is enough to reverse $\Delta\eta_{\text{jj}}$ behaviour for the Mueller-Tang analysis (cf. Sec. B.5)

Chapter 5

Summary and Outlook

This work produced the first results from an implementation of BFKL at NLL with LO impact factors into PYTHIA 8 successfully reproducing a previous result for f_{CSE} as well as confirming a previous theoretical prediction for the *Survival Probability*. Analysis also brings attention to the resulting distribution of f_{CSE} with respect to $\Delta\eta_{\mathrm{jj}}$ as concerning, since they seem to contradict both theoretical intuitions and experimental data.

Attention has also been brought upon the standard *Multi-Parton Interactions* and *Colour Reconnection* routines in PYTHIA 8, which were originally discarded in accordance with standard practice for BFKL studies but has shown promising results. More work needs to be done in order to validate the model for future analysis, such as its robustness in other setups, like LHC at 7 TeV, or how much it is affected by different tunes.

As a follow up, NLO contributions for fixed order QCD using POWHEG were computed and along with *Parton Distribution Functions*, the latter through both analytical and phenomenological analysis, have been discarded as causes for strong deviations from experimental observations in MC predictions.

Different parts of PYTHIA 8 have been studied and their impact and effects on observables explored, both in the main text as well as in Appendix B. In particular, special attention is required to be put into the *Initial State Radiation* routine, as it has many important repercussions, from producing large factorization scale uncertainties, to being the main influence in $\Delta\phi_{\mathrm{jj}}$ and, to a lesser extent, $p_{\mathrm{T,jet\text{-}2}}$ distributions. Additionally, ISR seems to spoil the prospect of enlarging the gap region, as it being nearly isotropic between the jets makes charged multiplicities grow with the gap amplitude.

An alternative model for the Colour Reconnection routine in PYTHIA 8, the *Soft Colour Interactions* model, has been implemented from the previous FORTRAN code. The basic model seems to saturate and produce unsatisfying results for LHC at 13 TeV without retuning. Some steps towards integrating the new junction colour topologies from PYTHIA 8 have already been taken, and in the future they could be fully integrated.

Finally, a short historical context is given. It serves to display the difficulties in studying BFKL dynamics while using Monte Carlo simulators, as they may be having a stronger influence in results than previously anticipated, not because of large deviations from physical events, but rather because of how sensible our observable is to small variations.

In regards to next steps, they can be divided into two broad categories: reproducing previous results and improving soft physics. In order to reproduce previous results [Che+09; KMR10] that showed a much better agreement with experimental data, both CSE and QCD contributions should be examined, validating them, for example, with individual BFKL cross section predictions from [KMR10] and other NLO QCD calculations for dijet events. Exploring the reasons why our predictions do not match theirs is key to furthering our understanding of BFKL phenomenology.

As for improving soft physics, the default MPI+CR routines in PYTHIA 8 are a great place to start. The model should be tested in different scenarios for which we already have data, and compare against theoretical predictions for SP. The susceptibility to different tunes is also worth exploring. Getting a robust model for soft interactions will not only close the gap between experimental and theoretical analysis but also provide a dynamical "SP" that may contribute towards obtaining better predictions.

Appendix A

Useful Relations: $2 \to 2$ kinematics

Here we present some kinematical relations for $2 \to 2$ processes that may be of assistance when studying them. For more detailed descriptions of these relations and how to derive them we refer to [Ams+08].

Mandelstam variables:

$$a + b \to c + d$$

$$s = (p_a + p_b)^2$$
$$t = (p_a - p_c)^2$$
$$u = (p_a - p_d)^2$$

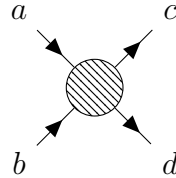Figure A.1: Mandelstam variables for $2 \to 2$ processes.

Rapidity and pseudorapidity:

$$\left.\begin{array}{rl} \text{rapidity} \quad y & = \dfrac{1}{2}\ln\dfrac{E + p_z c}{E - p_z c} \\[2ex] \text{pseudorapidity} \quad \eta & = -\ln\left[\tan\dfrac{\theta}{2}\right] \end{array}\right\} \quad m \ll E \implies y \simeq \eta$$

Bjorken-x:

$$x_1 = \frac{p_T}{\sqrt{s}}\left(e^{y_3} + e^{y_4}\right) = \frac{2p_T}{\sqrt{s}}e^{\gamma}cosh(\hat{\eta})$$

$$x_2 = \frac{p_T}{\sqrt{s}}\left(e^{-y_3} + e^{-y_4}\right) = \frac{2p_T}{\sqrt{s}}e^{-\gamma}cosh(\hat{\eta})$$

$$x_1 = e^{2\gamma}x_2$$

Partonic variables:

$$\hat{s} = x_1 x_2 s = 4p_T^2 e^{\gamma - \gamma} cosh^2(\hat{\eta})$$
$$= 4p_T^2 cosh^2(\hat{\eta})$$
$$= 2p_T^2 (cosh(\Delta\eta) + 1)$$
$$= x_1^2 e^{2\gamma} s$$
$$= x_2^2 e^{-2\gamma} s$$

$$\hat{t} = (p_1 - p_3)^2 = -p_T^2(1 + e^{-2\hat{\eta}})$$
$$= -p_T^2(1 + e^{-\Delta\eta})$$
$$= -2p_T^2 e^{-\hat{\eta}} cosh(\hat{\eta})$$

$$-\frac{\hat{s}}{\hat{t}} = 1 + e^{2\hat{\eta}}$$

Observable variables:

$$\Delta\eta = 2\hat{\eta}, \text{(only absolute value observable)}$$
$$p_T^2 = \frac{-\hat{t}}{1 + e^{-2\hat{\eta}}} = \frac{-\hat{t}}{\hat{s}}(\hat{s} + \hat{t})$$
$$= \frac{-\hat{t}e^{\hat{\eta}}}{2cosh(\hat{\eta})}$$
$$\Delta\phi = \pi, \text{(at parton level)}$$

Appendix B

Additional Plots

This appendix contains plots and results that did not fit neatly in the main text for being either redundant or because they complement only one or two sentences. Either way it can be useful to have them here for future reference.

B.1 SpaceShower:rapidityOrder Influence

It was found that just applying the new tune for all parts of the simulation produced a value of f_{CSE} that was too low with respect to before the tune change, as can be seen in Fig. B.1.

The parameter `SpaceShower:rapidityOrder` is identified as the culprit for the excess of charged particles in the gap, as can be seen in Fig. B.2. When it is turned on, the emissions from ISR are forced to be ordered in rapidity, which reduces the phase space for additional emissions. This was the reason given in [Col19] to turn it off. A secondary effect this has is that now ISR emissions are more likely to enter the gap region, exacerbating the gap destruction effect from ISR.

The HardQCD part of the simulation is left untouched since the tune was originally fit to accurately describe those types of events in the first place.

The easiest solution is to just keep the CMS tune, but have `SpaceShower:rapidityOrder` turned on for CSE events. However, another approach is also tested. Following the description of SP in Sec. 4.3, if one explicitly simulates all possible emissions in phase space the survival probability has to be raised accordingly. A value of $SP = 0.2$ is found to give a good fit, as can be seen in Fig. B.3. The main advantage of this approach is that the simulation parameters for both parts are more similar, although there is no strong theoretical foundation for a value of SP = 0.2.

From these two approaches to recover the original predictions, keeping the parameter `SpaceShower:rapidityOrder` turned on for CSE events is chosen since it does not have a sizable impact on dynamical behaviour, which is our main concern, but it vastly improves the statistics of predictions.
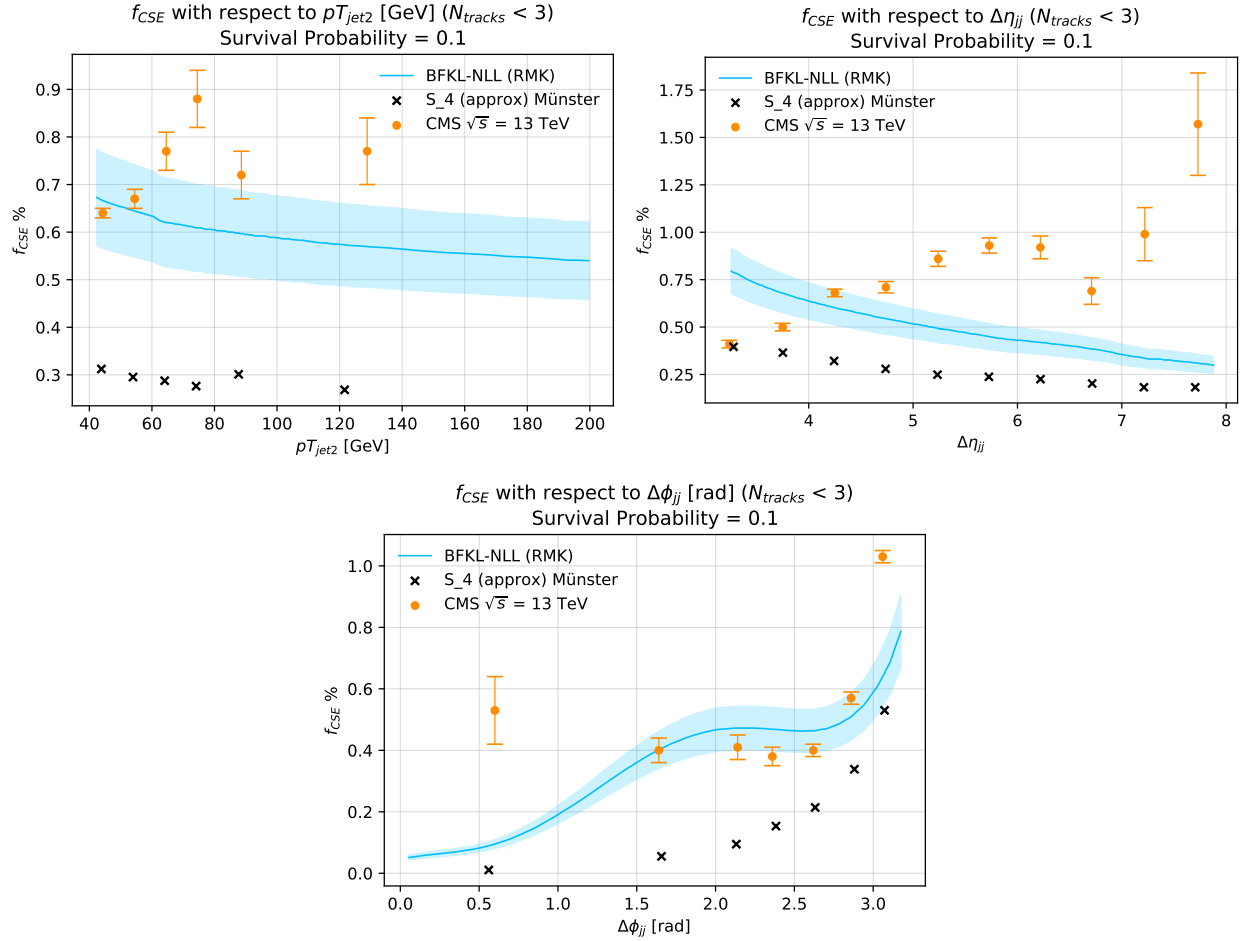
Figure B.1: Differential plots of f_{CSE} with respect to $p_{\text{T,jet-2}}$, $\Delta\eta_{\text{jj}}$ and $\Delta\phi_{\text{jj}}$ after applying the new tune from [Col19]. Corresponds with Tables 2, 3 and 4/Figure 8 in [CT20].

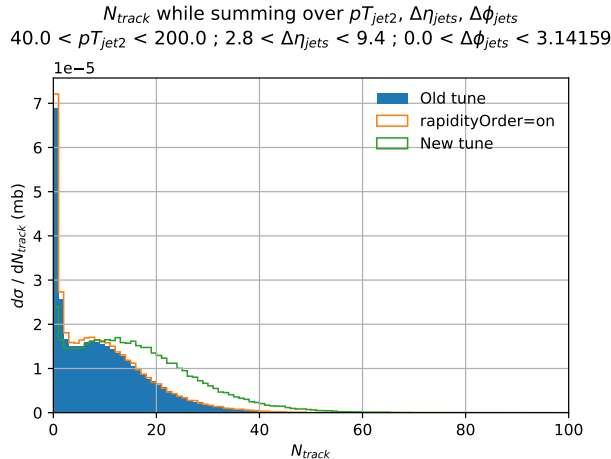Figure B.2: N_{tracks} distribution for CSE events and how it is affected by the parameter `rapidityOrder`. "New tune" has `rapidityOrder` off, "rapidityOrder=on" is the CP1 CMS tune with `rapidityOrder=on` and "Old tune" is the Monash tune, `Tune:pp=14` in PYTHIA 8, which also has `rapidityOrder=on`.

B.2 Influence from Kinematical Limits

In this section the possibility that the $\Delta\eta_{\text{jj}}$ trend is a selection effect of the kinematical limits imposed is explored. For this reason, two PDFs are compared, a normal PDF and a flat PDF, so that all events have equal probability. The distributions are at parton level to eliminate additional effects.

First the kinematical limit for $\Delta\eta_{\text{jj}}$ are removed so the distributions run from 0 to 10. Furthermore we also take away the imposed condition of having jets on opposite hemispheres we obtain Fig. B.4, which still maintains the same f_{CSE} distribution for both cases.

These results should not be considered physical, since they are well beyond the kinematical regime we require to make predictions. Nonetheless, they serve to illustrate the mathematical property exposed in Sec. 4.4 as it relates to kinematical impositions. Just to reiterate, *changing PDF or kinematical limits has no impact on f_{CSE}*. at parton level, and the changes they induce are heavily suppressed at hadron level.

B.3 NLO QCD Dynamics

The differential cross sections for QCD NLO and QCD LO with respect to N_{tracks}, $p_{\text{T,jet-2}}$, $\Delta\eta_{\text{jj}}$, $\Delta\phi_{\text{jj}}$ and $p_{\text{T,jet-3}}$ can be seen in Fig. B.5.

There is a slight shift in the $\Delta\eta_{\text{jj}}$ distribution, which is what we were looking for in the first place. Although the shift was not enough to correct the predictions it can be useful to take into account to compound with other similar effects.

Worth noting is that the total cross section seems to decrease for NLO and that the $\Delta\phi_{\text{jj}}$ distribution is more "peaked" at $\Delta\phi_{\text{jj}} \sim \pi$. The fact that there seem to be less third jets with high p_{T}, see Fig. B.6, point towards a selection effect. As the third jet carries momentum away from the two leading ones, less pairs fulfil the condition that $p_{\text{T,jet}} > 40\text{GeV}$.
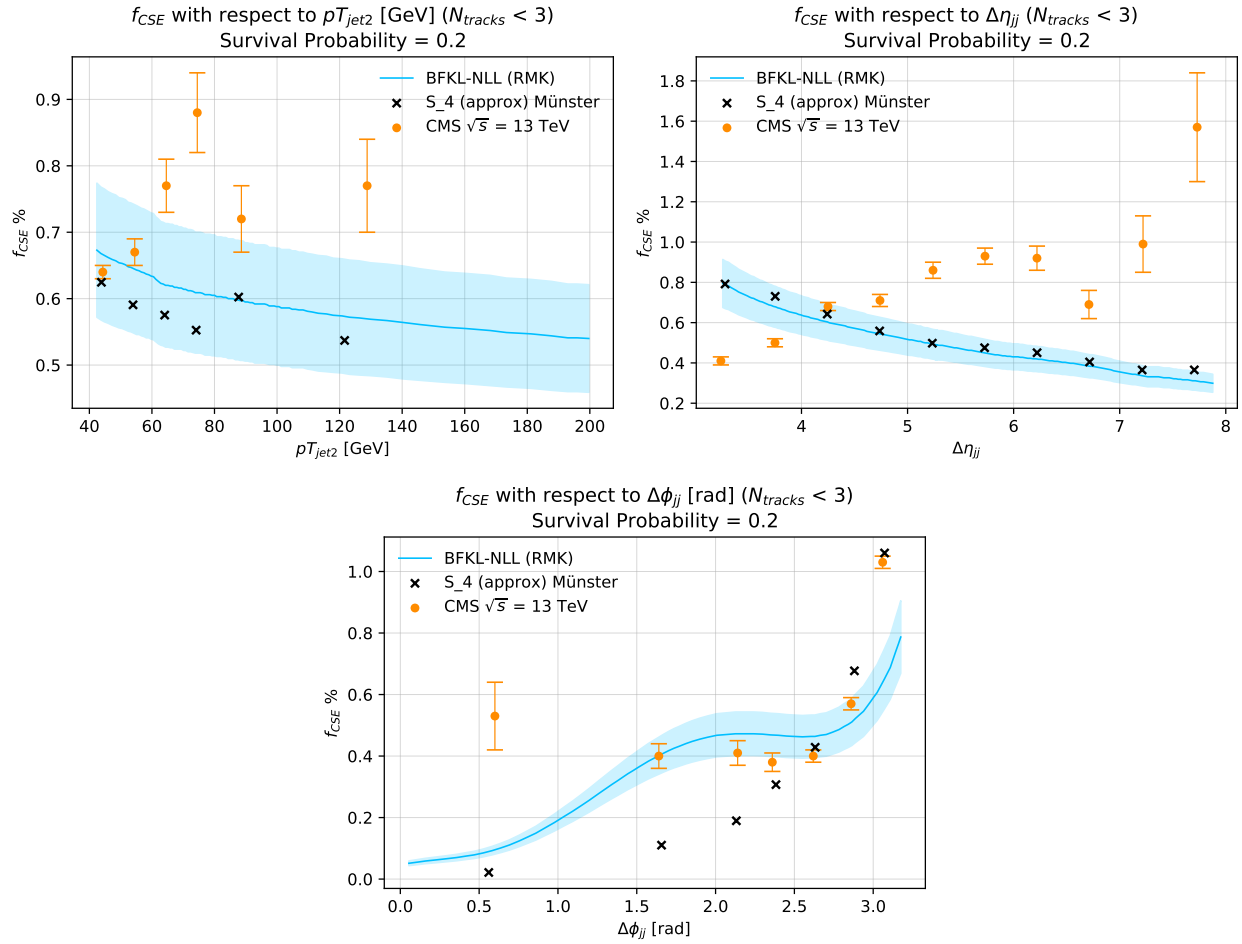
Figure B.3: Differential plots of f_{CSE} with respect to $p_{\mathrm{T,jet\text{-}2}}$, $\Delta\eta_{\mathrm{jj}}$ and $\Delta\phi_{\mathrm{jj}}$ after applying the new tune from [Col19] but keeping the parameter `rapidityOrder=off` for CSE. The survival probability has been raised to $SP = 0.2$. Corresponds with Tables 2, 3 and 4/Figure 8 in [CT20].
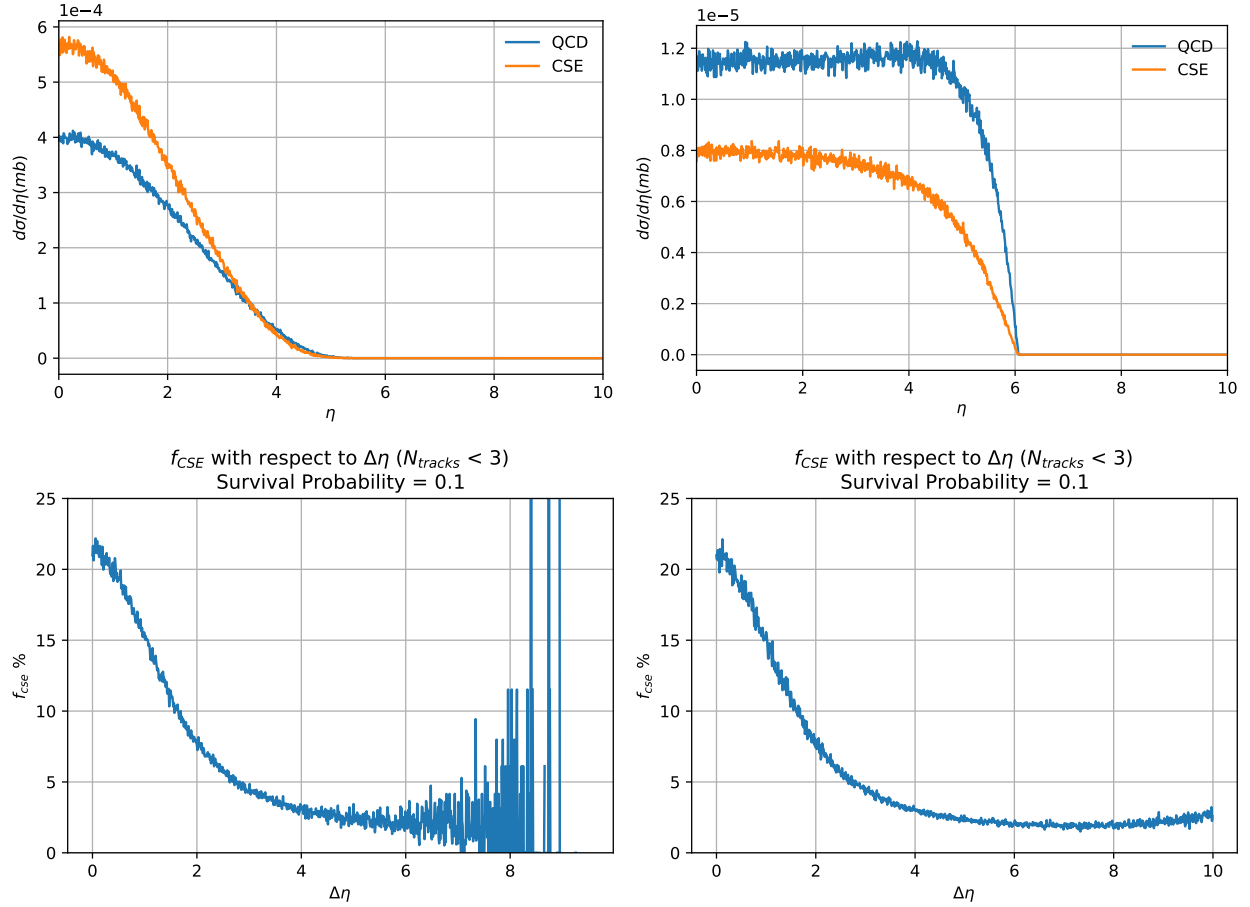
Figure B.4: Cross section and f_{CSE} distributions with respect to $\Delta\eta_{\mathrm{jj}}$. At the top are the full distributions, while the bottom plots have the differential f_{CSE} obtained from the distribution on top. On the left are the results using a normal PDF, while the results on the right were obtained using a flat PDF. For these plots no limits on η_{jet} have been imposed, which includes the condition $\eta_{\mathrm{jet\text{-}1}}\eta_{\mathrm{jet\text{-}2}} < 0..$

Figure B.5: Differential cross sections for LO and NLO QCD processes with respect to N_{tracks}, $p_{\text{T,jet-2}}$, $\Delta\eta_{\text{jj}}$, $\Delta\phi_{\text{jj}}$. Kinematical limits are those of Sec. 4.1 and the CP3 [Col19] was used.
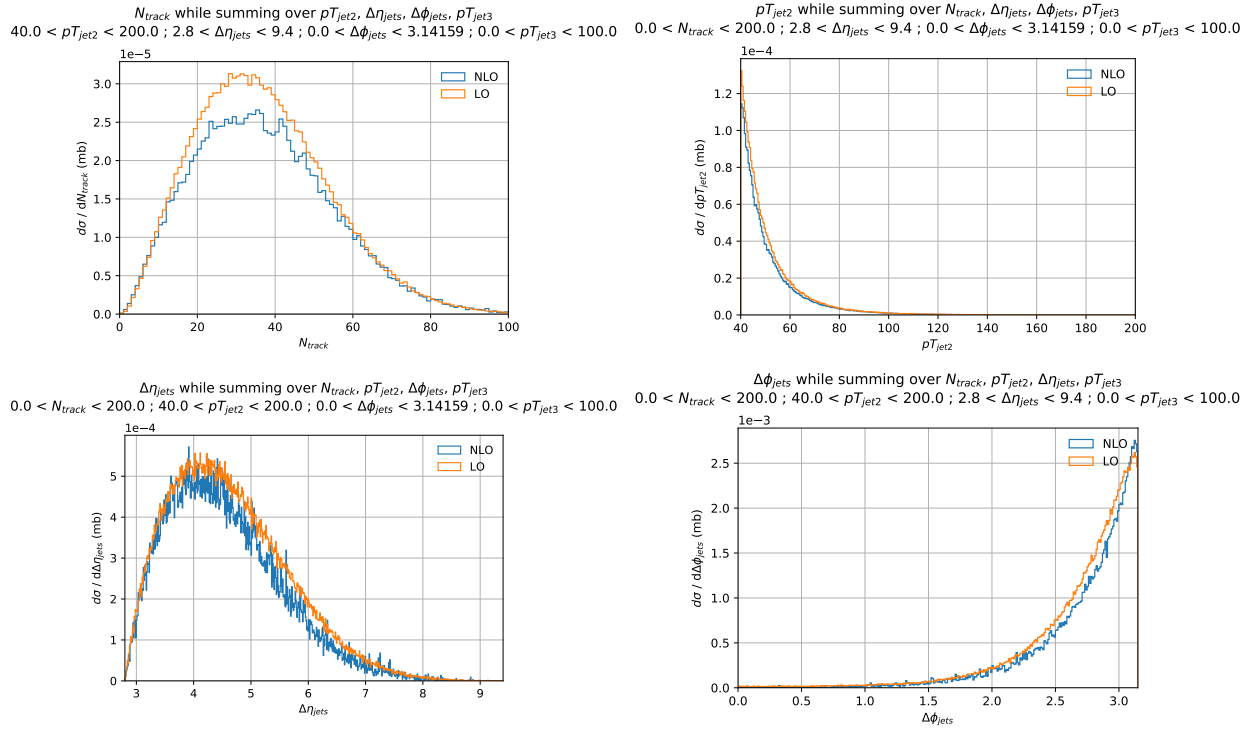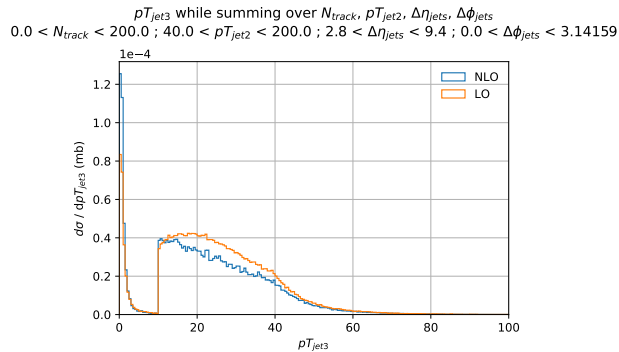
Figure B.6: Differential cross sections for LO and NLO QCD processes with respect to $p_{\text{T,jet-3}}$. Kinematical limits are those of Sec. 4.1 and the CP3 [Col19] was used.

This is also the reason why $\Delta\phi_{\text{jj}}$ seems more "peaked". Only when there is no third jet will $\Delta\phi_{\text{jj}} = \pi$. After that emissions carry more and more momentum, making the whole distribution fall faster than the LO case.

B.4 ISR Impact on Factorization Scale Uncertainty

Since the increase in produced particles from changing the factorization scale is not enough to account for the lack of cancellation. The reason why the explanation in Sec. 4.7 is based on a selection effect is that the emissions were observed to have increased their average energy by a factor *roughly* equal to the variation in the factorization scale. ISR looked like the most probable reason, since FSR is nearly collinear with the jets, so it would have an equal contribution for CSE and QCD, thus cancelling, and MPI has a small influence in this case (remember that all secondary interactions are forced to be softer than the main one).

The final clue that pointed towards ISR is that the variation in the N_{tracks} distribution disproportionally affects events with a large number of charged particles in the gap (see Fig. 4.13).

In order to verify whether our explanation for why factorization scale uncertainty do not cancel, we will turn ISR off and see how it influences the factorization scale uncertainties in Fig. B.7.

After turning ISR off the cancellation is nearly as good as with the renormalization scale or independent variation PDFs, which seems to validate our reasoning.

B.5 Showering Effect on $\Delta\eta_{\text{jj}}$ Distributions

As mentioned before, in [aC98] they point to the showering of events as responsible for the change in behaviour of the $\Delta\eta_{\text{jj}}$ distribution. In particular they are talking about the regular QCD part of the simulation, which they did in HERWIG. However we have the ability to shower both CSE and QCD, so next we will see if the same effect is observed. In order to see it, in Fig. B.8, a comparison between parton level and after showering distributions is presented.
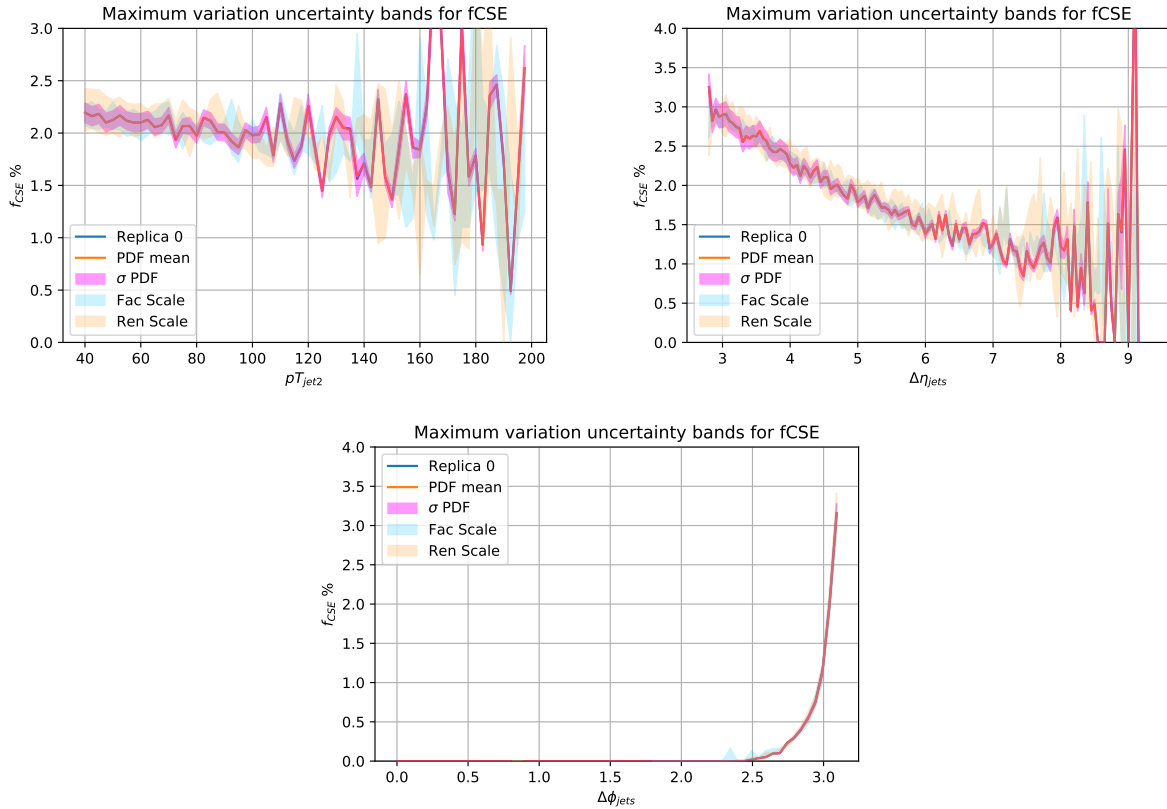
Figure B.7: Differential plots of f_{CSE} with respect to $p_{\mathrm{T,jet\text{-}2}}$, $\Delta\eta_{\mathrm{jj}}$ and $\Delta\phi_{\mathrm{jj}}$ displaying variations with respect to the renormalization and factorization scales, as well as PDF uncertainties when ISR is turned off, displaying the ISR dependency of the factorization scale uncertainty.
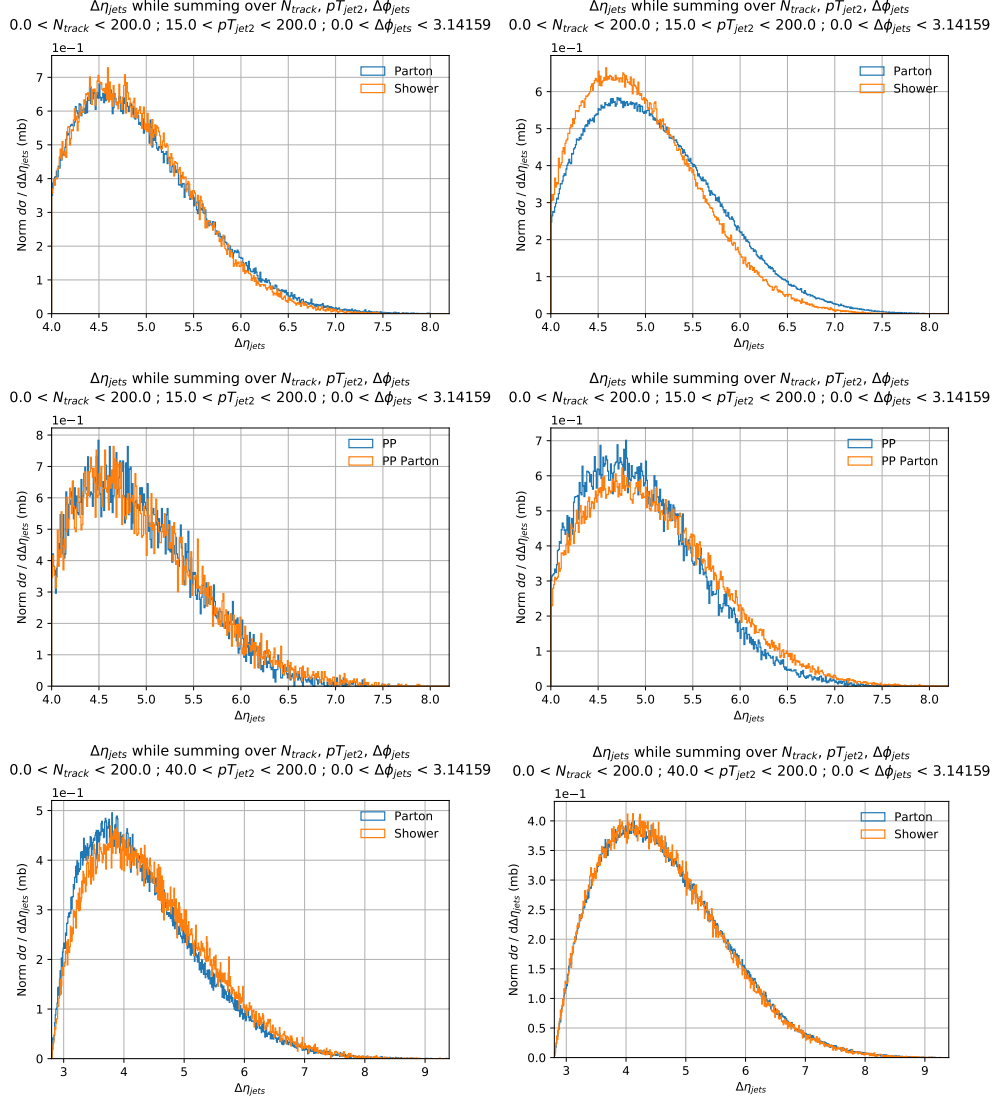
Figure B.8: Differential cross sections with respect to $\Delta\eta_{jj}$ displaying both parton and shower distributions. On the left are the CSE events, on the right QCD, on top Tevatron proton-antiproton events, on the middle Tevatron proton-proton events and on the bottom CMS proton-proton. All distributions have been normalized for ease of comparison of dynamical behaviour.

Three settings will be considered: Tevatron [aC98], at 1.8 TeV with $p\bar{p}$ and pp beams, and LHC [CT20] at 13 TeV with pp beams. This way, if the effect were due to $p\bar{p}$ or just a peculiarity of events at 1.8 TeV we will be able to tell.

As it was said in [aC98], after showering the regular QCD processes see their distribution shifted towards smaller values of $\Delta\eta_{\mathrm{jj}}$, while CSE processes barely change (top row of Fig. B.8). It is possible that this is an effect caused due to the events being $p\bar{p}$. If we now change the beams to proton-proton, as in [CT20], the effect is still there to pretty much the same extent. However, after also altering the centre of mass energy to match [CT20], the QCD distributions no longer move, and instead the CSE distribution changes, but to higher values of $\Delta\eta_{\mathrm{jj}}$. For the purposes of calculating f_{CSE}, QCD shifting to lower values is the same as CSE shifting to higher ones.

Therefore, the effect noted by [aC98] has been also observed in our setup. However this effect was not enough to reverse the behaviour in our simulations, unlike in their simulations.

The second important takeaway from these plots is that both the QCD and BFKL distributions are extremely similar in $\Delta\eta_{\mathrm{jj}}$, which means that looking for what causes the downwards trend in this variable amounts to trying to find effects that may displace the peak of the distributions by < 0.5 units of rapidity.

Appendix C

SCI Implementation

The Soft Colour Interactions model [EIR95; Rat98] was implemented into PYTHIA 8. Two versions are displayed here, one that is purely a code translation from the FORTRAN code that implemented SCI into PYTHIA 6 provided by Ekstedt, Enberg, and Ingelman, and a second implementation in the style of PYTHIA 8. The reason why two versions are provided is that, while the do the same, it is much easier to spot mistakes in the FORTRAN one, while it is easier to optimize the full C++ implementation.

The main challenge has to do with the changes in colour structures considered by each iteration of PYTHIA. PYTHIA 6 works at leading colour, while PYTHIA 8 took it one step further and works in next to leading colour, which means that it considers junction structures(cf. Sec. 3.3).

Junctions have been dealt with by simply colour tracing the final configuration and updating the `endCol` fields. There is also the possibility of completely integrating junctions in the routine, since junctions are also colour-charge-sinks and colour-charge-emitters (much like quarks are). It would just take a bit of care into how to deal with them.

C.1 Code Translation

Here is the first implementation, made just translating the original code from FORTRAN into C++ and dealing with junctions at the end.

```cpp
// Original SCI, i.e. replicating PYTHIA 6 PYFSCR routine. CC -> Carbon Copy

bool ColourReconnection::reconnectSCI_CC( Event& event, int iFirst){
  double prob;
  double tHat = infoPtr->tHat();
  // Number of colour reconnections. Serves as indices in the arrays
  // It's actually twice the number of dipoles
  int nCR = 0; // So that we can work with 0-indexing
  // Maximum number of particles that can take part in the reconnection
  int oldSize = event.size();
  // Minimum (first) colour tag used -> from html documentation
  int ctmin = 100;
  // Highest (last) colour tag used and copy for CR
  int maxCT = event.lastColTag();
```

```
16      int nCT = maxCT;
17      // Vectors of indexes and partner indexes
18      vector<int> iCol(8000), jCol(8000);
19      // Vector for colour types (true, false) -> (colour, anti-colour)
20      vector<int> typeCol(8000);
21
22      // Temporary colour tag storage
23      int CTi, CTj;
24
25      // Dipole log in
26      for (int CT = ctmin; CT < maxCT; CT++){
27        // Loop over particles and find the "ends" of the colour line
28        int iC=0, iA=0;
29        // Make ind start from iFirst?
30        for (int ind=0; ind < oldSize; ind++){
31          if (!event[ind].isFinal()) continue; // Needed? Yes?
32          if (event[ind].col() == CT) iC = ind;
33          if (event[ind].acol() == CT) iA = ind;
34        } // End of finding ends of dipole
35
36        // Log the dipole into our vectors
37        if (iC != 0 && iA != 0){
38          // Add coloured parton
39          iCol[nCR] = iC;
40          jCol[nCR] = nCR+1;
41          typeCol[nCR] = 1;
42          nCR++;
43          // Add anticoloured partner
44          iCol[nCR] = iA;
45          jCol[nCR] = nCR-1;
46          typeCol[nCR] = 2;
47          nCR++;
48        } // End of logging in the dipole
49      } // End of Colour Tagging loop
50
51      // Dipole reconnection
52      int NCOLR = 0; // Counter
53      int jColI, jColJ; // Index holders for the reconnection
54      for (int i=0; i < nCR-1; i++){
55        for (int j=i+1; j < nCR; j++){
56          // Check type of colour is equal
57          if (typeCol[i] != typeCol[j]) continue;
58
59          // Don't allow gluons to reconnect to themselves
60          if (iCol[i] == iCol[ jCol[j] ]
61              || iCol[j] == iCol[ jCol[i] ]) continue;
62
63          // Perform probability check
64          // Probability is static -> Can be defined out of loops
65          prob = p0 * pow(log(1. - tHat/400.), 0.1);
66          if (rndmPtr->flat() < prob){
67
68            // typecol[i] has to be equal to typecol[j]
69            if (typeCol[i] == 1){
```

58

```
70          CTi = event[ iCol[i] ].col();
71          CTj = event[ iCol[j] ].col();
72
73          event[iCol[jCol[j]]].acol(CTi);
74          event[iCol[jCol[i]]].acol(CTj);
75        } else {
76          CTi = event[ iCol[i] ].acol();
77          CTj = event[ iCol[j] ].acol();
78
79          event[iCol[jCol[j]]].col(CTi);
80          event[iCol[jCol[i]]].col(CTj);
81        }
82
83        // Update partner arrays
84        // It affects subsequent reconnections
85        jColI = jCol[i];
86        jColJ = jCol[j];
87
88        jCol[i] = jColJ;
89        jCol[jColJ] = i;
90
91        jCol[j] = jColI;
92        jCol[jColI] = j;
93
94        // Update number of colour reconnections
95        NCOLR++;
96
97      } // End of probability check and reconnection
98    } // End internal particle loop
99  } // End loop over all pairs of particles
100
101  // Correct endCols in Junctions
102  int endCol;
103  for (int iJun=0; iJun < event.sizeJunction(); iJun++){
104    for (int leg=0; leg < 3; leg++){
105      endCol = event.colJunction(iJun, leg);
106      // Colour-Junctions output colours
107      if (event.kindJunction(iJun) % 2 == 0){
108        bool endReached = false;
109        while (!endReached){
110          for (int ind=0; ind < event.size(); ind++){
111            if (endCol == event[ind].acol()){
112              // Ignore colour-sextets (CT<0) for now
113              if (event[ind].col() > 0){
114                endCol = event[ind].col();
115                break;
116              } else {
117                // If "ind" does not link to another tag
118                endReached = true;
119              }
120            }
121          }
122          // If no further connections found
123          endReached = true;
```

```
124            }
125          }
126
127          // AntiColour-Junctions output anticolours
128          if (event.kindJunction(iJun) % 2 == 1){
129            bool endReached = false;
130            while (!endReached){
131              for (int ind=0; ind < event.size(); ind++){
132                if (endCol == event[ind].col()){
133                  // Ignore colour-sextets (CT<0) for now
134                  if (event[ind].acol() > 0){
135                    endCol = event[ind].acol();
136                    break;
137                  } else {
138                    // If "ind" does not link to another tag
139                    endReached = true;
140                  }
141                }
142              }
143              // If no further connections found
144              endReached = true;
145            }
146          }
147
148          // Assign endCol to the leg
149          event.endColJunction(iJun, leg, endCol);
150        } // End of leg for-loop
151      } // End of Junction correction
152    return true;
153    }
```

Listing C.1: SCI model translation from FORTRAN to PYTHIA 8 C++. This code is meant to be placed in the file ColourReconnection.cc

C.2 C++ Implementation

Here is the second implementation, made using PYTHIA 8's structures and dealing with junctions at the end.

```
1
2    // Do SCI with code in style of PYTHIA 8
3
4    bool ColourReconnection::reconnectSCI_v0( Event& event, int iFirst){
5      // Use BeamDipole simply because it has everything needed
6      vector<BeamDipole> Dips;
7
8      // Define parameters
9      double tHat, prob;
10     // Maximum number of particles that can take part in the reconnection
11     int oldSize = event.size();
12     // Highest (last) colour tag used and copy for CR
13     int maxCT = event.lastColTag();
```

```
14      int nCT = maxCT;
15
16      // Build dipole vector
17      for (int ind1=0; ind1 < oldSize-1; ind1++){
18        // Only take into account particles that exists
19        if (!event[ind1].isFinal()) continue;
20        // Take colour end of the dipole as "anchor" for dipole build
21        int ColT = event[ind1].col();
22        if ( !(event[ind1].col() > 0) ) continue;
23        // Find partner
24        for (int ind2=0; ind2 < oldSize; ind2++){
25          // Check the particle exists and is not the same particle
26          if (!event[ind2].isFinal()) continue;
27          if (ind1 == ind2) continue;
28          // Find the acol end of the colour line (tag)
29          if (event[ind2].acol() == ColT){
30            Dips.push_back( BeamDipole (ColT, ind1, ind2) );
31            // Allow only one dipole creation (would exclude junctions)
32            break;
33          } // End of adding dipole
34        } // End of finding partner
35      } // End of building dipole vector
36
37      // Dipole reconnection
38      int coli, colj, acoli, acolj;
39      tHat = infoPtr->tHat();
40      for (int i=0; i < Dips.size(); i++){
41        for (int j=0; j < Dips.size(); j++){
42          coli = Dips[i].iCol;
43          acoli = Dips[i].iAcol;
44          colj = Dips[j].iCol;
45          acolj = Dips[j].iAcol;
46          // Check only colour type of one end of the dipole
47          if (event[coli].colType() != event[colj].colType()) continue;
48
49          // Check so gluons don't reconnect to themselves
50          if (coli == acolj || acoli == colj) continue;
51
52          // Perform probability check
53          prob = p0 * pow(log(1. - tHat/400.), 0.1);
54          if (rndmPtr->flat() < prob){
55            // Generate new colour tags. Other standard methods don't do this
56            // and reconnect to the existing colour lines
57            nCT = event.nextColTag();
58            // Reconnect on the event record
59            event[coli].col(nCT);
60            event[acolj].acol(nCT);
61            // Update dipole vector with the reconnection
62            // This may only matter for future reconnections
63            Dips[i].col = nCT;
64            Dips[i].iAcol = acolj;
65
66            // Same process for the other side
67            nCT = event.nextColTag();
```

```
68            event[colj].col(nCT);
69            event[acoli].acol(nCT);
70            Dips[j].col = nCT;
71            Dips[j].iAcol = acoli;
72          } // End of probability check and reconnection
73        } // End internal dipole loop
74      } // End external dipole loop
75
76      // Correct endCols in Junctions
77      int endCol;
78      for (int iJun=0; iJun < event.sizeJunction(); iJun++){
79        for (int leg=0; leg < 3; leg++){
80          endCol = event.colJunction(iJun, leg);
81          // Colour-Junctions output colours
82          if (event.kindJunction(iJun) % 2 == 0){
83            bool endReached = false;
84            while (!endReached){
85              for (int ind=0; ind < event.size(); ind++){
86                if (endCol == event[ind].acol()){
87                  // Ignore colour-sextets (CT<0) for now
88                  if (event[ind].col() > 0){
89                    endCol = event[ind].col();
90                    break;
91                  } else {
92                    // If "ind" does not link to another tag
93                    endReached = true;
94                  }
95                }
96              }
97              // If no further connections found
98              endReached = true;
99            }
100         }
101
102         // AntiColour-Junctions output anticolours
103         if (event.kindJunction(iJun) % 2 == 1){
104           bool endReached = false;
105           while (!endReached){
106             for (int ind=0; ind < event.size(); ind++){
107               if (endCol == event[ind].col()){
108                 // Ignore colour-sextets (CT<0) for now
109                 if (event[ind].acol() > 0){
110                   endCol = event[ind].acol();
111                   break;
112                 } else {
113                   // If "ind" does not link to another tag
114                   endReached = true;
115                 }
116               }
117             }
118             // If no further connections found
119             endReached = true;
120           }
121         }
```

```
122
123          // Assign endCol to the leg
124          event.endColJunction(iJun, leg, endCol);
125        } // End of leg for-loop
126      } // End of Junction correction
127   } // End of reconnectSCI_v0
128   return true;
```

Listing C.2: SCI model implementation in PYTHIA 8 C++. This code is meant to be placed in the file ColourReconnection.cc

C.3 Additional Changes

Just implementing one of the previous routines into ColourReconnection.cc is not enough, as PYTHIA needs some additional setup before it can run the SCI model.

First of all, in ColourReconnection.h, `p0` and the SCI methods need to be added as private members of the `ColourReconnection` class.

```
1   // p0 member declaration
2   double p0;
3   // Do SCI reconnection scheme
4   bool reconnectSCI_v0(Event& event, int oldSize);
5   // "Modern" implementation of SCI
6   bool reconnectSCI_CC(Event& event, int oldSize);
```

Listing C.3: Modifications on ColourReconnection.h

Then in ColourReconnection.cc, `p0` has to initialized in the constructor of the Colour-Reconnection class. For this we will use one of the spare parameters included in PYTHIA 8. PartonLevel.cc and BeamRemnants.cc need to be modified accordingly to ensure SCI gets called only where it is supposed to.

```
1   // In ColourReconnection.init()
2   p0 = settings.parm("Main:spareParm1");
3
4   // In ColourReconnection.next()
5   // reconnectMode = 5 -> calls SCI only in PartonLevel.cc
6   // reconnectMode = 6 -> calls SCI in BeamRemnants.cc::addOld()
7   // to match the calls of default CR
8   else if (reconnectMode == 5 || reconnectMode == 6)
9     return reconnectSCI_XX(event, iFirst);
```

Listing C.4: Modifications on ColourReconnection.cc

Finally, when setting the program, the following settings need to be added so that PYTHIA correctly sets up.

```
1   // Set parameter in program
2   pythia.settings.mode("ColourReconnection:mode", CRMode, true);
3   pythia.readString("Main:spareParm1 = "+to_string(p0Value));
```

Listing C.5: Setup in main program.

References

[aC98]   B. Abbott et al and DZero Collaboration. *Probing Hard Color-Singlet Exchange in ppbar Collisions at root-s=630 GeV and 1800 GeV.* 1998. arXiv: `hep-ex/9809016 [hep-ex]`.

[Ams+08]   C. Amsler et al. "Review of Particle Physics". In: *Physics Letters B* 667.1 (2008). Review of Particle Physics, pp. 1–6. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/j.physletb.2008.07.018`. URL: `http://www.sciencedirect.com/science/article/pii/S0370269308008435`.

[And+83]   B. Andersson et al. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: `https://doi.org/10.1016/0370-1573(83)90080-7`. URL: `http://www.sciencedirect.com/science/article/pii/0370157383900807`.

[Boo+01]   E. Boos et al. "Generic User Process Interface for Event Generators". In: *2nd Les Houches Workshop on Physics at TeV Colliders*. Sept. 2001. arXiv: `hep-ph/0109068`.

[Che+09]   F. Chevallier et al. *Gaps between jets at hadron colliders in the next-to-leading BFKL framework.* 2009. arXiv: `0903.4598 [hep-ph]`.

[Col+17]   The NNPDF Collaboration et al. *Parton distributions from high-precision collider data.* 2017. arXiv: `1706.00428 [hep-ph]`.

[Col17]   CMS Collaboration. *Study of dijet events with a large rapidity gap between the two leading jets in pp collisions at $\sqrt{s} = 7$ TeV.* 2017. arXiv: `1710.02586 [hep-ex]`.

[Col19]   CMS Collaboration. *Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements.* 2019. arXiv: `1903.12179 [hep-ex]`.

[CS15]   Jesper R. Christiansen and Peter Z. Skands. "String formation beyond leading colour". In: *Journal of High Energy Physics* 2015.8 (Aug. 2015). ISSN: 1029-8479. DOI: `10.1007/jhep08(2015)003`. URL: `http://dx.doi.org/10.1007/JHEP08(2015)003`.

[CSS08]   Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-ktjet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: `10.1088/1126-6708/2008/04/063`. URL: `http://dx.doi.org/10.1088/1126-6708/2008/04/063`.

[CT20]   – CMS Collaboration and – TOTEM Collaboration. "Study of hard color singlet exchange in dijet events with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (Apr. 2020). URL: http://cds.cern.ch/record/2712121.

[Don+02]   Sandy Donnachie et al. *Pomeron Physics and QCD*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2002. DOI: 10.1017/CBO9780511534935.

[EEI17]   Andreas Ekstedt, Rikard Enberg, and Gunnar Ingelman. *Hard color singlet BFKL exchange and gaps between jets at the LHC*. 2017. arXiv: 1703.10919 [hep-ph].

[EIR95]   A. Edin, G. Ingelman, and J. Rathsman. *Soft Colour Interactions as the Origin of Rapidity Gaps in DIS*. 1995. arXiv: hep-ph/9508386 [hep-ph].

[Eks15]   Andreas Ekstedt. "The soft color interactions of diffractive scattering : A study of Rapidity gap formation through BFKL exchange". PhD thesis. Uppsala University, High Energy Physics, 2015, p. 79.

[Ell16]   Joshua Ellis. *TikZ-Feynman: Feynman diagrams with TikZ*. 2016. arXiv: 1601.05437 [hep-ph].

[Enb03]   R. Enberg. "Quantum Chromodynamics and Color Singlet Exchange in High Energy Interactions". PhD thesis. Uppsala U., 2003.

[FR97]   J. R. Forshaw and D. A. Ross. *Quantum Chromodynamics and the Pomeron*. Cambridge Lecture Notes in Physics. Cambridge University Press, 1997. DOI: 10.1017/CBO9780511524387.

[KL12]   Yuri V. Kovchegov and Eugene Levin. *Quantum Chromodynamics at High Energy*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2012. DOI: 10.1017/CBO9781139022187.

[KMR10]   O. Kepka, C. Marquet, and C. Royon. *Gaps between jets in hadronic collisions*. 2010. arXiv: 1012.3849 [hep-ph].

[KMR18]   V A Khoze, A D Martin, and M G Ryskin. "Multiple interactions and rapidity gap survival". In: *Journal of Physics G: Nuclear and Particle Physics* 45.5 (Mar. 2018), p. 053002. ISSN: 1361-6471. DOI: 10.1088/1361-6471/aab1bf. URL: http://dx.doi.org/10.1088/1361-6471/aab1bf.

[Lev98]   E. Levin. *An introduction to Pomerons*. 1998. arXiv: hep-ph/9808486 [hep-ph].

[MR09]   C. Marquet and C. Royon. "Azimuthal decorrelation of Mueller-Navelet jets at the Tevatron and the LHC". In: *Physical Review D* 79.3 (Feb. 2009). ISSN: 1550-2368. DOI: 10.1103/physrevd.79.034028. URL: http://dx.doi.org/10.1103/PhysRevD.79.034028.

[MT92]   Alfred H. Mueller and Wai-Keung Tang. "High-energy parton-parton elastic scattering in QCD". In: *Phys. Lett. B* 284 (1992), pp. 123–126. DOI: 10.1016/0370-2693(92)91936-4.

[Rat98]   Johan Rathsman. *A Generalised Area Law for Hadronic String Reinteractions*. 1998. arXiv: hep-ph/9812423 [hep-ph].

[Sch14]   Matthew D. Schwartz. *Quantum Field Theory and the Standard Model.* Cambridge University Press, Mar. 2014. ISBN: 978-1-107-03473-0, 978-1-107-03473-0.

[Sjö+14]   Torbjörn Sjöstrand et al. *An Introduction to PYTHIA 8.2.* 2014. arXiv: `1410.3012 [hep-ph]`.

[Sjö84]   Torbjörn Sjöstrand. "Jet fragmentation of multiparton configurations in a string framework". In: *Nuclear Physics B* 248.2 (1984), pp. 469–502. ISSN: 0550-3213. DOI: `https://doi.org/10.1016/0550-3213(84)90607-2`. URL: `http://www.sciencedirect.com/science/article/pii/0550321384906072`.

[SMS06]   Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. *PYTHIA 6.4 Physics and Manual.* 2006. arXiv: `hep-ph/0603175 [hep-ph]`.