LU TP 04-25
June 2004

Correcting the Colour Dipole Cascade with Fixed Order Matrix Elements in Deep Inelastic Scattering

Master of Science Thesis
Christoffer Åberg

Thesis Advisor: Leif Lönnblad

Abstract

For describing the multi-hadronic final state in deep inelastic scattering, there are mainly two different approaches: fixed order matrix elements and parton showers. Fixed order matrix elements can account for a few widely separated jets, whereas parton showers give a good description of inter- and intrajet structure. Thus the two approaches complement each other in different regions of phase space, so it seems beneficial to use both. In this thesis an algorithm to combine fixed order matrix elements with the colour dipole cascade — the parton shower algorithm implemented in Ariadne— is modified for deep inelastic scattering and tested.

Department of Theoretical Physics
Lund University
2004

Contents

1 Introduction

What is the world made of?

Though probably cliché at the beginning of a thesis in high energy physics, the author feels that it has some legitimacy in this context. For with the current commonly accepted theory of particle physics, the Standard Model, we are one step closer to achieving the goal Demokritos set out over two thousand years ago.

Neglecting the discussion of whether the description physics offer of the world has something to do with the underlying nature of it, the Standard Model answer to the question is quarks and leptons, interacting via the exchange of gauge bosons. It is a beautiful theory, and can be rather compactly formulated, but as is not unusual in physics, the equations cannot be solved exactly and generally.

It is therefore not surprising that everything is not satisfactorily understood yet. For example, it is common knowledge that the atoms surrounding us are made up of a nucleus and electrons, and that the nucleus in turn is made up of protons and neutrons — but what makes up the proton? How is its structure related to the quarks and leptons of the Standard Model?

This question can, at least to a certain extent, be answered by so-called deep inelastic scattering experiments. In much the same way as Rutherford once revealed the inner structure of the atom by bombarding it with α-particles, one can investigate the structure of the proton by scattering leptons off it.

In fact, experiments performed at the Stanford Linear Accelerator (SLAC) during the late 1960s were the first to reveal that the proton was not an elementary particle. The constituents were called partons, and this discovery was a big step towards the eventual acceptance of the quark as a real particle and not just a mathematical construct.

Such experiments in the beginning only detected the lepton, and did not care too much about what happened to the proton. In order to gain further insight, later experiments began taking a closer look at the final state. What one found was not quarks, but rather what is known as jets — collimated streams of composite particles.

To describe this satisfactorily is rather difficult. There are mainly two different approaches: fixed order matrix elements and parton showers, complementing each other in different regions of phase space. Fixed order matrix elements can account for a few widely separated jets, whereas parton showers give a good description of inter- and intrajet structure. To make more precise comparisons with data, it would be beneficial to make a combined description, but it is not trivial to do this consistently. In this thesis, an algorithm for doing this for the colour dipole cascade parton shower, as implemented in the program ARIADNE, is presented and tested.

The organization of this report is as follows: In the next section, section 2, some preliminaries concerning basic properties of quantum field theory, the Standard Model and quantum chromodynamics, essential for the rest of the report, are given. In the following section, section 3, the process of deep inelastic scattering is introduced, and the leading and next-to-leading order processes presented. The reader experienced in the field of high energy physics can safely skip these two sections.

In section 4 a brief introduction to the parton shower approach in general and the colour dipole cascade model in particular is given. Then section 5 discuss the algorithm to combine the matrix element approach with the colour dipole cascade in a consistent manner, and the results of this is presented in section 6. Finally, in section 7, the work is summarized and some possible future directions given.

2 Preliminaries

In the 1970s, the electromagnetic and weak interactions were unified into a single theory, the electroweak theory. Together with quantum chromodynamics (QCD) which was later added, these constitute the Standard Model (SM) of particle physics. This theory describes three of the four fundamental forces of nature; only gravity is missing, but at the energies available at present and in the foreseeable future gravitational effects are deemed insignificant.

Quantum field theory (QFT) is the basic framework within which the SM is formulated, so knowledge of this is obviously important for understanding the latter. However, this is not the place for a thorough exposition of the subject, so only some basic differences between QFT and nonrelativistic quantum mechanics (QM) necessary for the following discussion are noted. In the same spirit, the very useful technique of Feynman diagrams is explained.

Finally, a brief account of the SM is given, followed by some of the essential properties of QCD, and eventually concluded by a discussion of proton structure.

2.1 Quantum Field Theory

QFT combines quantum mechanics with special relativity. Specifically this means that since mass and energy are equivalent, it should be possible to create particles from energy and vice versa. This is indeed the case in QFT. Loosely speaking, adding quantum mechanical indeterminacy to this implies that the number of particles of a specific species is an observable with a corresponding operator — the number operator.

One of the fundamental differences between QFT and QM is then that, whereas the number of particles is something that is *put into* the problem in QM, it is part of the *solution* in QFT. The expected value of the number of particles for a certain state is then given by the expectation value of the number operator.

However, since the number of particles is now an observable, there is no fundamental reason that the number operator should commute with other basic operators, like eg. the Hamiltonian. This opens up the possibility to expand an arbitrary state in terms of the number operators, a so-called expansion in Fock space. This feature of QFT will be seen to be crucial for the understanding of proton structure.

2.2 Feynman Diagrams

Consider a general $A + B \to C + D$ process, where two incoming particles A and B with four-momentum p_A and p_B interact to produce the particles C and D with four-momentum p_C and p_D. The differential cross section for such a process can be written

$$\mathrm{d}\sigma = \frac{|\mathcal{M}|^2}{F}\mathrm{dLips}\,,$$

where

$$F = 4\sqrt{(p_A \cdot p_B)^2 - m_A^2 m_B^2},$$

is the incident flux, and

$$\mathrm{dLips} = (2\pi)^4 \delta^{(4)}(p_C + p_D - p_A - p_B)\frac{\mathrm{d}^3 p_C}{(2\pi)^3 2E_C}\frac{\mathrm{d}^3 p_D}{(2\pi)^3 2E_D},$$

the Lorentz invariant phase space. These two factors describe the kinematics of the process, whereas the matrix element (ME), $\mathcal{M}$, contains the actual physical theory. For brevity, only a $2 \to 2$ process was considered, but the result is readily generalized.

The standard way to calculate the ME is by a perturbative expansion

$$\mathcal{M} = C_0 + \alpha C_1 + \alpha^2 C_2 + \cdots\,,$$

where α denotes an arbitrary coupling constant. A very nice thing about such an expansion is that one can represent each term of a given order by topologically distinct and connected graphs, with the same number of vertices as the order. These graphs are the celebrated Feynman diagrams, and the rules by which they are constructed are called Feynman rules.

This way the procedure for calculating the ME to a given order for any process is clear. One simply draws the possible diagrams of the given number of vertices and can then immediately write down the amplitude according to the Feynman rules. Feynman diagrams are thus a very powerful tool, as long as the expansion converges fast enough so that a simple truncation at a fixed order can be justified.

2.3 The Standard Model

The basic building blocks of the SM, arranged into families, are the leptons

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}$$

and the quarks

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}.$$

All of them interact electroweakly via the exchange of the gauge bosons γ, Z and $W^\pm$, whereas only the quarks interact via the gauge bosons of QCD, the gluons. The strength of the interactions are characterized by coupling constants.

When renormalizing these theories, it turns out that the coupling constants are not constants at all, but rather depend on the momentum transfer of the process. This is called running of the coupling constants.

2.4 Quantum Chromodynamics

QCD is the theory of the strong interaction of the quarks, with corresponding charges called colours — red, green and blue. The quarks all carry colour charge, whereas the antiquarks carry anticolour charge. The gluons are also charged, existing as the eight non-singlet combinations of colour-anticolour.

That the gluons carry charge makes the running of α_s, the strong coupling constant, an especially big effect. At low momentum transfer α_s rises rapidly, so at a scale below around 1 GeV, the interaction is very strong. It has as a consequence that isolated quarks are never seen — they always appear in colour neutral states, hadrons, like eg. the proton. This important feature of QCD is called confinement.

With confinement in mind, it might seem strange that QCD also allows for the quarks inside a hadron to behave as if essentially free when probed,

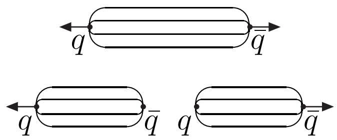

Figure 1: Schematic view of a $q\bar{q}$ pair flying off at high energies and producing two mesons.

a property known as asymptotic freedom. However, it too can be justified on the grounds of the the running of α_s, for at high momentum transfer α_s falls off. This means that if a quark inside a hadron is struck by a highly energetic probe, its coupling to the rest of the particles within the hadron is diminished.

A struck quark will thus just fly off on its own way from the hadron remnant, the latter continuing without disturbance. However, along the way the energy of the colour field between the hadron remnant and the struck quark will increase. In contrast to the electric field, the colour field is believed to form a tube, much like the magnetic field in a type II superconductor. This means that when the energy of the colour field is high enough a $q\bar{q}$ pair can be produced, dividing the colour field, as schematically shown in figure 1 for the simpler case of a $q\bar{q}$ pair separating.

This is repeated until all the original energy is converted to particles forming hadrons, a process called hadronization. This means that what emerges in a detector is always hadrons, and never free quarks. Several models for hadronization exists, one of the more commonly used being the Lund String Model. A feature of hadronization is that a parton of high energy will often emerge as a collimated stream of several hadrons, known as a jet.

2.5 Proton Structure

Knowing the quantum numbers of the quarks, a first simplified picture of the proton would be that of a bound state of two u quarks and one d quark. Such a constellation in its ground state shares the observed quantum numbers of the proton, making such an identification well-justified.

This picture also makes a great analogy to the atom as a bound state of a nucleus and a number of electrons, but whereas QM is a good approximation in atomic physics, it is *not* at the high energies of the strong interaction, and QFT has to be taken into account.

It turns out that in QCD the Hamiltonian, of which the proton is an eigenstate, does not commute with the number operators, so which particles actually make up the proton is not well-defined. However, expanding the proton eigenstate in Fock space, as discussed above, one can write it symbolically something like

$$|\text{proton}\rangle = a_1 \, |uud\rangle + a_2 \, |uudg\rangle + a_3 \, |uudgg\rangle + \cdots + b_1 \, |uudq\bar{q}\rangle + \cdots .$$

An interpretation of this is to view the proton as being composed of two u and one d quark, the so-called valence quarks. In order to stay bounded, these exchange gluons, which therefore also constitute part of the proton. At a given moment there is also a finite probability of finding a $q\bar{q}$ pair, so-called sea quarks, popping out of the vacuum for a short time in accordance with the Heisenberg uncertainty principle.

3 Deep Inelastic Scattering

In deep inelastic scattering (DIS) a lepton and a hadron are made to collide. Since leptons are (at least within the SM) assumed point-like and only interact electroweakly, they make an ideal candidate for probing a hadron. This is because whatever happens that is not accounted for by the electroweak interaction, which can determined by the lepton four-momentum with small radiative corrections due to γ radiation, must be an effect of hadronic structure.

In the remainder of this text only collision of an electron and proton will be considered, but it should be noted that an analogous experiment can be done, and indeed has been done, with other leptons and hadrons.

3.1 Dynamics of *ep* Scattering

To leading order in α_{EM}, the only possible Feynman diagram for ep scattering is the one photon (or Z boson) exchange of figure 2, where the QCD part of the interaction is represented by a blob for generality. The incoming proton four-momentum is denoted by P.

Writing the exchanged four-momentum $q = k - k'$, where k and k' is the four-momentum of the incoming and outgoing electron respectively, it is customary to introduce $Q^2 = -q^2$ as the virtuality of the exchanged boson. According to the de Broglie hypothesis, the energy of a particle is inversely proportional to its wavelength, so one can view Q^2 as a measure of how well the probing boson can resolve the proton structure.

For modest Q^2, the photon cannot "see" the internals of the proton. The proton behaves essentially like an elementary particle, and scatters as a whole. The ep scattering is *elastic*.

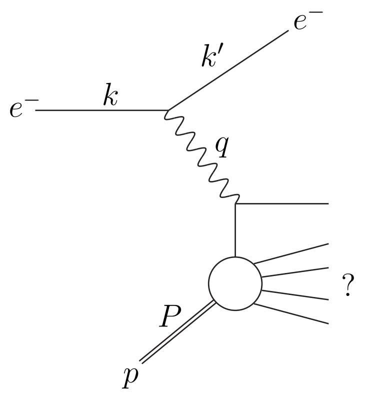

Figure 2: The scattering of an electron off a proton to first order in α_{EM}.

However, as Q^2 is increased the momentum transfer becomes large enough to excite the proton to a resonant baryon state, like eg. the Δ^+. The scattering becomes *inelastic*.

As Q^2 is increased even more, eventually the wavelength becomes small enough for the boson to resolve the internals of the proton, and find out what it is made up of. The scattering is *deeply inelastic*.

3.2 Parton Momentum Distributions

To describe DIS purely theoretically would mean somehow applying the Fock space expansion of subsection 2.5 in practice. However, no one has ever actually done this expansion explicitely. Indeed, finding solutions of QCD in the non-perturbative regime is at present beyond our abilities — which is why DIS experiments are so interesting in the first place.

To solve this, what one does is to introduce so-called parton momentum distribution functions $f_i(x)$, describing the probability that a parton of type i carries a momentum fraction x of the whole proton four-momentum P. Naturally, the sum of the four-momentum of all the constituents must add up to the four-momentum of the whole proton, so these functions must satisfy

$$\sum_i \int \mathrm{d}x\, x f_i(x) = 1.$$

3.3 The Naïve Quark Parton Model

In the naïve quark parton model (QPM), the scattering of the electron off the proton at high Q^2 is seen as the elastic scattering off one of the quarks. We can thus write the total differential cross section for DIS as the incoherent

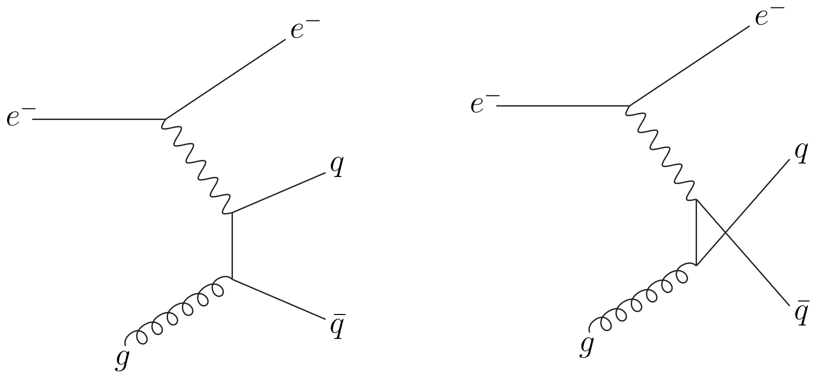

Figure 3: The BGF process.

sum of elastic scattering off the quarks

$$\mathrm{d}\sigma_{\,ep\to eX} = \sum_q \int \mathrm{d}x\, f_q(x) \mathrm{d}\sigma_{\,eq\to eq}, \tag{1}$$

where the summation runs over the quarks only. Such a simple model is justified by the property of asymptotic freedom.

The differential cross section of equation (1) depends only on one kinematical variable, x. The parton momentum distribution functions are said to satisfy *Bjorken scaling*, a fact that can be traced back to the incoherence assumption of the preceding paragraph. It was also well-established experimentally, until higher energies made it possible to investigate even higher Q^2, and scaling violations were found.

3.4 Next-to-Leading Order Processes

The QPM completely ignores all properties of QCD. This was originally its strength, since at the time of its formulation QCD was not firmly established as *the* theory of the strong interaction. However, at present one can use QCD to write down the next-to-leading order processes — the boson-gluon fusion (BGF) and QCD Compton (QCD-C) processes — the Feynman diagrams of which are shown in figures 3 and 4 respectively.

Obviously, the BGF process will make the differential cross section for $ep \to eX$ depend not only on the momentum distributions of the quarks, but also on the momentum distribution of the gluon.

In a sense, including these two processes like this is illogical. One might as well say that this should already be included in the parton momentum distributions, since these sum up the ignorance of the actual partonic struc-

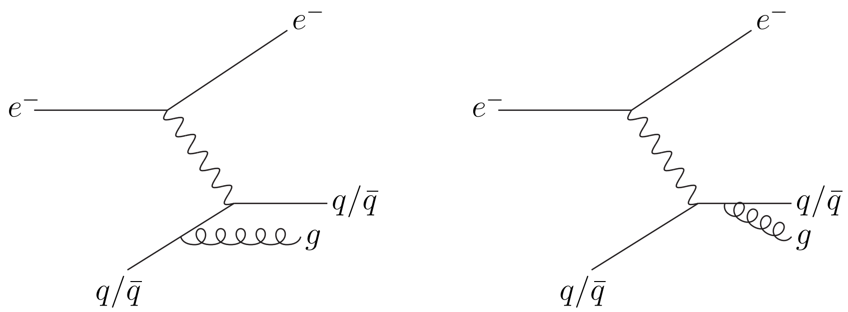

Figure 4: The QCD-C process.

ture of the proton. Indeed this can be done, making the f_i depend also on Q^2.

This dependence on Q^2 can be calculated approximately, the result being the famous Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equation. From measured values of $f_i(x, Q_0^2)$ at a given Q_0^2, one can then find an approximation for it at another Q^2.

The procedure is then to regard the $f_i(x, \mu^2)$ as something that is put into the problem. These signify the probability of finding parton i with momentum fraction x when probing the proton at virtuality μ^2. The rest of the interaction is handled by the Feynman diagrams. In this sense the scale μ^2 is a measure of the ignorance of the QCD interactions.

For the case of the QPM, the virtuality at which the proton is probed is $\mu^2 = Q^2$, and *all* the dependence on the details of the strong interactions are in the momentum distributions.

For the case of the QCD-C and BGF processes, the virtuality is *not* that of the exchanged boson, for it is not the actual parton being probed which is struck. In these cases, part of the details of the QCD interactions are handled by the Feynman diagrams, and another part by the momentum distributions.

Of course one can go to higher order processes as well, making the approximation better and better. However the Feynman diagrams and the MEs they result in, get more and more complicated, so this is not feasible to do by hand. Fortunately there exists computer programs to do this. However, already at the order of, say 5, the calculation time is getting so big that this is no longer feasible in practice.

Another disadvantage is that in the soft and collinear regions the α_s expansion breaks down, because of large logarithmic terms and a running α_s of the order of 1. Clearly a better method is needed in these regions.

4 The Colour Dipole Cascade Model

The idea behind the parton shower approach is the successive emission of gluons by a probability distribution, a so-called splitting function $\mathcal{P}$, approximating the correct ME. The approximation is such that only the leading logarithmic part of the ME is kept[1], but on the other hand they are summed to *all* orders.

One requires the emissions to be ordered in some kinematical variable, ρ, specific for different parton shower algorithms, such that splittings with higher ρ are performed before splittings with lower ρ. Such a procedure can be nicely implemented into a computer program, where the program can just keep doing emission after emission at smaller and smaller ρ, all the way down to a cut-off. This requires the introduction of a so-called Sudakov form factor of the form

$$\Delta(\rho_1, \rho_2) = \exp\left(-\int_{\rho_1}^{\rho_2} \mathrm{d}\mathcal{P}\right),$$

measuring the probability of no emission between the scale ρ_1 and ρ_2.

For example, say that the latest emission was done at the scale ρ_1. Then the emission of another gluon at scale ρ_2 will be associated with such a factor to take into account that there was no emission between these two scales. In other words, the probability for this emission would be

$$\mathcal{P}(\rho_2)\Delta(\rho_1, \rho_2).$$

A disadvantage of the parton shower approach is that no interference terms are kept, which means that this in some sense is a semi-classical description. This is an especially bad approximation for hard wide-angle emissions, where the ρ scales are not strongly ordered. However, comparing with experiments this procedure seems to work surprisingly well.

In the remainder of this text, only the colour dipole cascade will be considered.

4.1 Colour Dipoles in e^+e^--annihilation

Consider the annihilation of an electron and a positron into a $q\bar{q}$ pair in figure 5, where either the quark or the antiquark later emits a gluon. Two colour configurations which results in the same final state are shown; in the left one the gluon is sent out by an initially blue quark, whereas in the right

[1] actually, this is what is known as the leading log approximation (LLA) — there also exists approximations taking into account nonleading terms, so-called modified leading log approximation (MLLA)

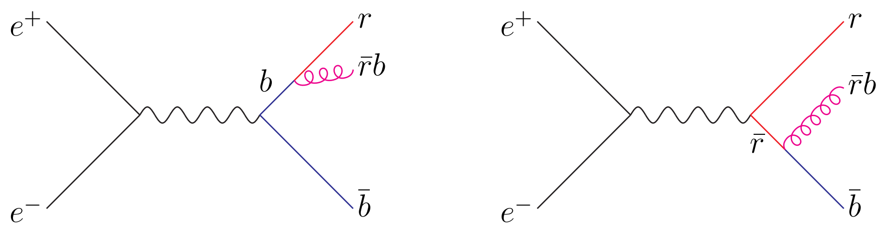

Figure 5: e^+e^--annihilation into a $q\bar{q}$-pair, later emitting a gluon.

one the gluon is sent out by an initially antired antiquark. Since the diagrams interfere, the question of whether the quark or the antiquark emitted the gluon is not a physical one, so it seems reasonable to make the interpretation that the gluon in some sense was emitted by both — a colour-dipole emission.

It can be shown [1] that the emission of a second softer gluon to a good approximation can be treated as being emitted from two independent dipoles, one between the quark and the first gluon, and one between the first gluon and the antiquark.

4.2 The Colour Dipole Cascade Model

The colour dipole cascade model (CDM) [2] generalizes the results above to any order; ie. a third even softer gluon is emitted by three independent dipoles, a fourth can then be emitted by four independent dipoles, and so on. In this way, by using splittings where two partons radiate a third rather than the $1 \to 2$ splittings found in most other parton shower algorithms, colour interference is automatically built into the model.

The three possible dipoles are the ones already presented in figure 5, together with the quark-gluon and the gluon-gluon dipoles. All of them are summarized in figure 6.

It can be shown that the CDM is a good approximation if one requires strongly ordered emissions in invariant $p_\perp$, defined by

$$p_\perp^2 = s_{123}(1 - x_1)(1 - x_3), \qquad (2)$$

where

$$x_i = \frac{2E_i}{\sqrt{s_{123}}},$$

is the final state energy fraction of particle i in the dipole center of mass system, and s_{123} is the total invariant mass of the dipoles squared. This ordering is attained by a Sudakov form factor as discussed above.

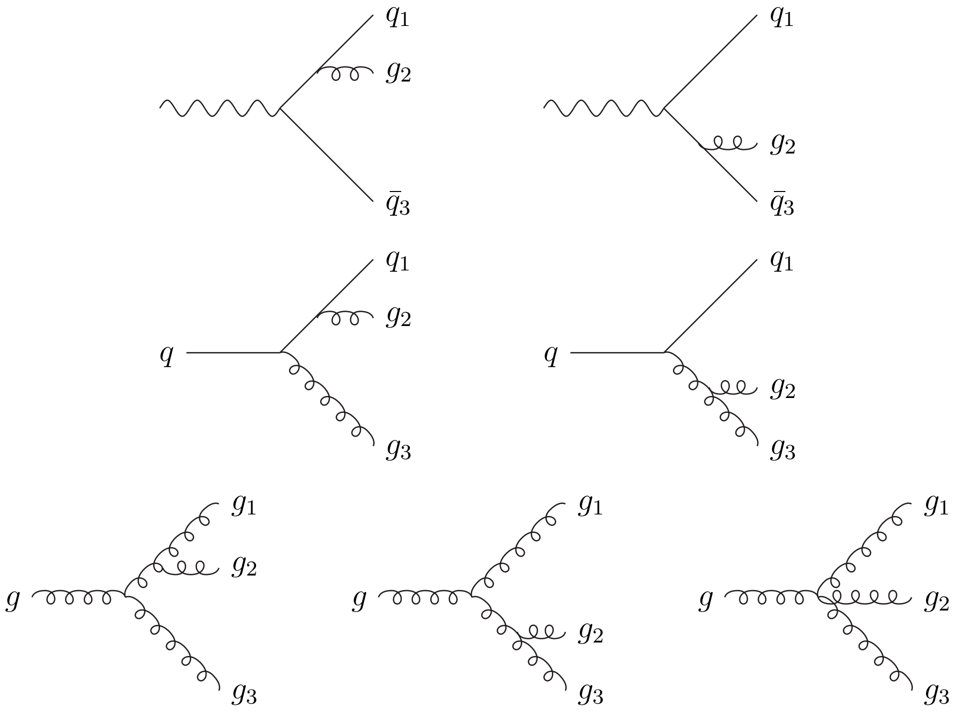

Figure 6: The possible dipoles.

In terms of the invariant $p_\perp$ and

$$y = \frac{1}{2} \ln \left(\frac{1 - x_1}{1 - x_3} \right),$$

roughly the rapidity of the emitted gluon in the dipole rest frame, the cross sections for all three dipoles can to a good approximation be written as

$$\mathrm{d}\sigma \propto \alpha_{\mathrm{s}}(p_\perp^2) \frac{\mathrm{d}p_\perp^2}{p_\perp^2} \mathrm{d}y,$$

with the approximate available phase space of figure 7. Note that the running coupling constant α_{s} is taken at scale $p_\perp^2$.

Actually, the cross sections above does not uniquely determine the emission from a dipole, since they only give the energy fractions of the particles. Neither the direction of the emitted gluon, nor the transverse recoil of the radiating particles due to the emission is specified. For details refer to [2].

It should be apparent that there is one process which simply cannot happen in terms of the dipoles presented above; that of a gluon splitting into a $q\bar{q}$ pair. Looking at e^+e^--annihilation to quark, antiquark and a gluon,

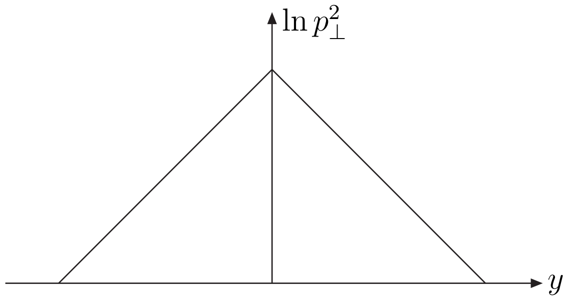

Figure 7: The available phase space for dipole emission.

the latter splitting into a $q\bar{q}$ pair, a suitable splitting function for a gluon splitting in a quark-gluon dipole can be found. Again, refer to [2] for details.

4.3 The Colour Dipole Cascade Model in Deep Inelastic Scattering

As one might guess, the CDM was first used to describe e^+e^--annihilation. In applying the model to DIS one assumes that it is the struck quark and the proton remnant which forms the initial dipole. For this case there are some changes that has to be made, due to the fact that whereas the quark and antiquark produced in e^+e^--annihilation are point-like, the proton remnant is certainly not.

In electrodynamics small wavelength emission from an extended antenna is suppressed. The prescription in ARIADNE[2] is to only allow a fraction a, defined as

$$a = \left(\frac{\mu}{p_\perp}\right)^\alpha,$$

of the proton remnant negative light-cone component, $P^- = E - P_z$, to take part in the emission. Here μ is the inverse size of the proton remnant, and α its dimension.

However, implementing this consistently, implies that also only part of the proton remnant should be allowed to recoil. This is achieved by yet another gluon, taking a fraction a of the energy of the proton remnant, which shares the recoil together with the other part of the emitting dipole. Again, for further details, refer to [2].

There is also the issue of how much of the proton remnant which should actually be a part of the initial dipole. In ARIADNE this is done differently

[2] actually, the *default* prescription in ARIADNE; others are available by choice

depending upon if the struck quark was a valence or a sea quark. In the former case the two remaining valence quarks are treated as a diquark, forming a dipole with the struck quark.

In the sea quark case, the antipartner of the struck quark is allowed to form a hadron with part of the valence quarks; the remaining valence quark(s) then forms a dipole with the struck quark. For a struck quark this means leaving a meson and letting the struck quark form a dipole with a diquark, whereas for a struck antiquark a baryon is left and the struck antiquark forms a dipole with the remaining valence quark.

The hadron is given a fraction of the proton four-momentum, which means that the phase space for emissions is reduced in the direction of the remnant. For details refer to [2]; the important thing here is to note that the model is sensitive to how the remnant is treated.

Another complication in DIS is that of the BGF process. This process is not described at all. Also, though the QCD-C process *is* described, it is not by the exact ME. The first emission from the initial dipole is therefore changed to better match these two processes. The procedure is to either let the initial dipole emit a gluon by a modified (to take into account the extendedness) form of the QCD-C ME, or to emit the antipartner of a struck sea quark according to the BGF ME. Which of the two is chosen is controlled by the Sudakov form factor.

In the latter case, the struck quark and its antipartner each form an independent dipole with the proton remnant. The remnant is split into a diquark and a quark forming dipoles with the quark and antiquark respectively. This is done in an analogous fashion to how the proton remnant is divided leaving a meson or baryon above.

Even if the initial emission was that of a gluon, BGF-like processes, ie. processes where the antipartner of a struck sea quark is emitted, can still be simulated in later stages, but now by the splitting function of initial state gluon splitting into a $q\bar{q}$ pair.

Correcting the CDM like this is fairly easy for the first emission, but for higher order ME corrections a different procedure is needed. This is the subject of the following section.

5 Combining Matrix Elements with the Colour Dipole Cascade Model

The objective is now to combine the generation of a handful quarks and gluons according to a fixed order ME, and then continue with a parton shower on this state. The ME generates some hard and widely separated jets, which the parton shower then fills with softer emissions, building up a more realistic picture with an internal structure of the jets, as well as a structure between them.

In this way, one uses the ME in that part of phase space where this is a good approximation, and the parton shower where this is a better description. The main problem is then how to separate these two areas so that no region of phase space is neglected, or populated twice.

In the following, the algorithm presented in [3], modified for DIS, will be used.

5.1 Combining Matrix Elements with the Colour Dipole Cascade Model

In order for the ME not to diverge one needs to set a cutoff scale Q_{min}. Denoting by R_n the fraction of $n \leq N$ jets, counted excluding the struck quark and proton remnant, these can be written to N:th order as

$$\begin{aligned} R_0(Q_{\mathrm{max}}, Q_{\mathrm{min}}) &= 1 + \alpha_{\mathrm{s}} C_{0,1}^{\mathrm{ME}}(Q_{\mathrm{max}}, Q_{\mathrm{min}}) + \cdots \\ R_1(Q_{\mathrm{max}}, Q_{\mathrm{min}}) &= \alpha_{\mathrm{s}} C_{1,1}^{\mathrm{ME}}(Q_{\mathrm{max}}, Q_{\mathrm{min}}) + \cdots \\ &\vdots \end{aligned}$$

where Q_{max} is the maximum scale set by the total center of mass energy. The coefficients $C_{n,m}$ corresponds to the emission of n jets at order α_{s}^m, so one can, eg., identify $C_{0,1}$ with the interference terms of the QPM process and its first order virtual corrections, and $C_{1,1}$ with the sum of the QCD-C and BGF processes.

In the CDM the scales are instead defined in terms of the invariant $p_\perp$ defined by equation (2). Denote the cutoff scale by $p_{\perp\mathrm{min}}$, and the maximum scale, again set by the total center of mass energy, by $p_{\perp\mathrm{max}}$. The fraction of 0 jets is then the probability of not having any emission between $p_{\perp\mathrm{max}}$ and $p_{\perp\mathrm{min}}$, ie. the Sudakov form factor

$$\begin{aligned} R_0(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) &= \Delta s_0(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) = \\ &= \exp\left(-\alpha_{\mathrm{s}} \int_{p_{\perp\mathrm{min}}}^{p_{\perp\mathrm{max}}} \mathrm{d}p_\perp \int \mathrm{d}z \int \mathrm{d}\phi \left(\mathcal{P}_{\mathrm{QCD\text{-}C}}(p_\perp, z, \phi) + \mathcal{P}_{\mathrm{BGF}}(p_\perp, z, \phi)\right)\right), \end{aligned}$$

where the integration over z and ϕ is over all kinematically possible values, and $\mathcal{P}_{\mathrm{QCD\text{-}C}}$ and $\mathcal{P}_{\mathrm{BGF}}$ are the two possible splitting functions from the initial state. Performing the integration and expanding in α_s one finds an expression similar to the one obtained from the ME

$$R_0(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) = 1 + \alpha_\mathrm{s} C^{\mathrm{PS}}_{0,1}(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) + \cdots .$$

Analogously, the fraction of 1 jets is given by the probability of having an emission at some scale, $p_{\perp 1}$, times the probability of no emission *before* this (ie. between $p_{\perp\mathrm{max}}$ and $p_{\perp 1}$) and the probability of no emission *after* this (ie. between $p_{\perp 1}$ and $p_{\perp\mathrm{min}}$)

$$\begin{aligned} R_1(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) = \alpha_\mathrm{s} \int_{p_{\perp\mathrm{min}}}^{p_{\perp\mathrm{max}}} \mathrm{d}p_{\perp 1} \int \mathrm{d}z \int \mathrm{d}\phi\, \Delta s_0(p_{\perp\mathrm{max}}, p_{\perp 1}) \times \\ \times \big(\mathcal{P}_{\text{QCD-C}}(p_{\perp 1}, z, \phi) \Delta s_{\text{QCD-C}}(p_{\perp 1}, p_{\perp\mathrm{min}}) + \\ + \mathcal{P}_{\mathrm{BGF}}(p_{\perp 1}, z, \phi) \Delta s_{\mathrm{BGF}}(p_{\perp 1}, p_{\perp\mathrm{min}})\big), \end{aligned}$$

where $\Delta s_{\text{QCD-C}}$ and Δs_{BGF} are the Sudakov form factors representing no further emissions from the QCD-C state and the BGF state respectively. These two can be written as exponentials and expanded in α_s, so again after integration and expansion an expression of the same form as the ME is obtained

$$R_1(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) = \alpha_\mathrm{s} C^{\mathrm{PS}}_{1,1}(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) + \cdots .$$

For the general $n \geq 2$ jet case, the fraction grows even more complicated, because of the factorial increase in final states. However, from above it should be clear that it too can be written in the same form as the ME.

In order to combine the MEs with the CDM, one then simply replaces the coefficients $C^{\mathrm{PS}}_{n,m}$ with the corresponding exact ones $C^{\mathrm{ME}}_{n,m}$. Here only the coefficients coming from the tree-level diagrams, ie. $C_{n,n}$, will be replaced.

Specifically, this means that the 0 jet fraction is the same as the one in the ordinary cascade. In general the fraction of n jets in the corrected version are given by

$$\begin{aligned} R_0(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) &= \Delta s_0(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) \\ R_{1,\text{QCD-C}}(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) &= \alpha_\mathrm{s} \int_{p_{\perp\mathrm{min}}}^{p_{\perp\mathrm{max}}} \mathrm{d}p_{\perp 1} \int \mathrm{d}z \int \mathrm{d}\phi\, c^{\mathrm{ME}}_{QCD-C} \times \\ &\times \Delta s_0(p_{\perp\mathrm{max}}, p_{\perp 1}) \Delta s_{\text{QCD-C}}(p_{\perp 1}, p_{\perp\mathrm{min}}) \\ R_{1,\mathrm{BGF}}(p_{\perp\mathrm{max}}, p_{\perp\mathrm{min}}) &= \alpha_\mathrm{s} \int_{p_{\perp\mathrm{min}}}^{p_{\perp\mathrm{max}}} \mathrm{d}p_{\perp 1} \int \mathrm{d}z \int \mathrm{d}\phi\, c^{\mathrm{ME}}_{BGF} \times \\ &\times \Delta s_0(p_{\perp\mathrm{max}}, p_{\perp 1}) \Delta s_{\mathrm{BGF}}(p_{\perp 1}, p_{\perp\mathrm{min}}), \\ &\vdots \end{aligned} \tag{3}$$

where c^{ME}_{QCD-C} and c^{ME}_{BGF} are the MEs for the QCD-C and the BGF process respectively, differential in $p_\perp$, z and ϕ.

When $n = N$ one wants to continue with the cascade directly from the ME generated state, ie. R_N should actually be N *or more* jets. Mathematically this means removing the last Sudakov of R_N in equation (3).

In practice this whole procedure is done by generating events according to the exact tree-level amplitudes. By then reweighting these with the Sudakov form factors of equation (3), and then adding a cascade, one generates events according to the corrected procedure.

However, there is one huge problem: the scales in the Sudakov form factors of equation (3) are unknown. From the point of view of ME generation, there is no such thing as a certain emission being performed at a certain scale by a certain dipole. Indeed, all diagrams with the same final state are treated together as a bunch.

Therefore one has to somehow calculate these scales from the generated event. The procedure is to take the final state given by the ME generator, and then find a succession of emissions that could have generated the same state according to the CDM. From such a reconstruction one gets a series of states $S_0, S_1, \ldots, S_{n-1}$ and scales $p_{\perp 1}, p_{\perp 2}, \ldots, p_{\perp n}$, such that S_i is the state from which the CDM would have done its emission at a scale $p_{\perp i+1}$ had it been responsible for creating the event. From these one can find the corresponding Sudakov form factors and reweight the state. In practice the number of jets an ME generator can generate is rather limited, so this is feasible to do.

An advantage of this approach is that not only the scales can be reconstructed, but the whole kinematical history as the CDM would have generated it. This means that the Sudakov form factors can be obtained exactly, ie. with the actual available phase space at the emission, and not just estimated.

One also has to take into account that the ME generator might have a different way of evaluating the parton momentum distribution functions. There is also a minor complication due to the running of α_{s}. In the CDM α_{s} is, as noted above, evaluated at the scale of the emission, whereas the ME generator uses a fixed factorization scale. The procedure in both cases is simply to divide with the factors the ME generator has introduced, and multiplied by what the CDM would have used instead.

5.2 The Event Generation

`MadGraph` [4] is a program which generates Fortran code to calculate the amplitude for a given process. It takes as input the desired incoming and outgoing particles and from this deduce all possible tree level Feynman diagrams. The routine given as output can then calculate the amplitude for such an event happening for a specified set of four-momenta of the particles involved.

These routines rely heavily on the HELicity Amplitude Subroutine package (HELAS) [5] to actually calculate the amplitude. This package uses so-called helicity amplitude techniques instead of the more conventional trace methods. A major advantage of this technique is that all diagrams are summed before being squared, resulting in that the complexity of the calculations increase linearly, instead of quadratically, with the number of diagrams.

These routines can then be used as the basis of a Monte-Carlo for generating events with weights according to the calculated amplitudes. However, for ease of usage, there exists also the event generator `MadEvent` [6] to do this. Then one simply calls `MadEvent` to obtain the desired number of events in a format conforming to the so-called "Les Houches" format [7].

An event in this format consists of a list of incoming and outgoing particles and their four-momentum, as well as the colour flow between particles. Each event also has a corresponding weight. These events can than be given as input to ARIADNE where all reconstruction, reweighting and cascading takes place.

In this thesis, events with up to 1 jet were generated. The energy of the incoming electron and proton were set to HERA energies, and some cuts on the rapidity

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right),$$

and the transverse energy

$$E_\perp = \sqrt{m^2 + p_\perp^2},$$

of all outgoing particles except the electron were introduced for events with at least 1 jet. For the 0 jet case no such cuts are needed, since there is no concept of wide angle emission when there is only a struck quark.

Also, a cut on d_{ij} for two particles i and j, defined as

$$d_{ij} = \min(E_{\perp i}, E_{\perp j})\big(\Delta\Psi_{ij}^2 + (y_i - y_j)^2\big), \tag{4}$$

where

$$\cos\Delta\Psi_{ij} = \frac{\mathbf{p}_{\perp i}\cdot\mathbf{p}_{\perp j}}{p_{\perp i}p_{\perp j}},$$

was made on all outgoing particle pairs, except those including the electron. Finally, a cut on Q^2 was chosen. To avoid confusion, note that $\mathbf{p}_\perp$ and $p_\perp$ here and in the preceding paragraph refer simply to (p_x, p_y) and $\sqrt{p_x^2 + p_y^2}$ respectively, and not the invariant $p_\perp$ of equation (2).

To simplify for the reweighting, α_s was used at a fixed scale. For the same reason, parton momentum distributions functions of the form $f_i(x, \mu^2) = C_i$, where C_i is a constant, were used. C_i were set to 1 for the quarks and 10 for the gluon.

5.3 Reconstructing the Events

In the reconstruction all combinations of three particles (corresponding to dipole emissions) and all $q\bar{q}$-pairs (corresponding to gluon splittings), which are possible according to the colour connections given by `MadEvent`, are clustered together to two particles and a gluon respectively.

In each of these reconstructions again all possible combinations of particles are clustered together, etc.. This way *all* possible histories of the state are considered. Then one particular history is chosen at random with the corresponding splitting functions as weights.

As in the dipole cascade, where the splitting functions does not uniquely determine the directions of the emitted gluons, there is also an ambiguity in the directions here. For details refer to [3].

A problem arises when a reconstructed history which is not ordered in $p_\perp$ is found. In the case where there are other ordered histories to choose from, such a history is simply discarded, but it may happen that no ordered histories exists. In that case an unordered history is chosen, but the scales are modified such that for two scales which are unordered, the larger of them is chosen for both emissions.

To the original reconstruction algorithm for final-state dipole emissions introduced in [3], reconstructions of emissions from dipoles involving hadron remnants was added in [8] for W production in hadron-hadron collisions. In this thesis the latter were generalized to also handle the reconstructions which are specific for DIS.

5.4 Reweighting of the Parton Momentum Distribution

In a DIS event generated by the standard cascade, the event is weighted by $f_q(x, Q^2)$, where x is the momentum fraction of the struck quark and Q^2 the virtuality of the exchanged boson. In addition to this an extra factor of $f_g(x', p_\perp^2)/f_q(x, p_\perp^2)$, where x' is the momentum fraction of the gluon, is included for a sea quark emission. To mimic this in the corrected cascade the reconstructed scales are used.

The procedure [8] then is to reweight each event from `MadGraph` by

$$\frac{f_q(x, Q^2)}{C_i},$$

where the struck quark q, the x and Q^2 are taken from the reconstructed initial state, while C_i is the density introduced in `MadGraph`. For the case of sea quark emission, an extra reweighting factor of

$$\frac{f_g(x', p_\perp^2)}{f_q(x, p_\perp^2)},$$

where x' and $p_\perp^2$ are calculated from the reconstructed first emission, is included.

For the case of gluon emission, a $1/z$ pole was found [8]. Adding the phase space suppression due to the extension of the proton remnant, the full reweighting factor becomes

$$\frac{\theta(p_\perp^2, y)}{z},$$

where θ is the function corresponding to the phase space suppression, and $p_\perp^2$ and y, again, are calculated from the reconstructed first emission.

5.5 The Full Algorithm

1. Generate a weighted event of $n \leq N$ jets with specific four-momenta, according to the ME amplitudes.

2. If the invariant $p_\perp$ of any of these particles is below the cut-off of the CDM cascade, then start over from step 1 again.

3. Choose a reconstructed history of states $S_0, \ldots, S_{n-1}$ with emissions at the scales $p_{\perp 1}, \ldots, p_{\perp n}$, as described above.

4. Reweight the event with
$$\prod_{i=1}^{n} \frac{\alpha_\mathrm{s}(p_{\perp i}^2)}{\alpha_\mathrm{s}^\mathrm{ME}}$$
to get rid of the α_s used in the ME generation and instead use what would have been used by the CDM.

5. Similarly, reweight for the parton momentum distribution functions and phase space suppression as described earlier.

6. Make a trial emission with a maximum scale of $p_{\perp\mathrm{max}}$ from the state S_0. If this emission occurs at a scale above $p_{\perp 1}$, then start over from step 1. The probability of retaining the event is then exactly $\Delta s_0(p_{\perp\mathrm{max}}, p_{\perp 1})$.

   If not, another trial emission is made from state S_1 with a maximum scale of $p_{\perp 1}$. The procedure is restarted from step 1 if this emission is above $p_{\perp 2}$, ie. the state is kept with a probability $\Delta s_1(p_{\perp 1}, p_{\perp 2})$.

If not, yet another trial emission is made from state S_2, etc. all the way down to S_{n-1}.

If $n = N$, then let ARIADNE continue the cascade as usual. If not, a trial emission is made from S_n with a maximum scale of $p_{\perp n}$. If all particles passes the same cut that was used in the ME generation, ie. if the emission is not within the phase space handled by the ME, then let ARIADNE continue the cascade. Otherwise restart from step 1.

This procedure is called the Sudakov veto algorithm, and corresponds exactly to the Sudakovs of equation (3), together with the discussion for the $n = N$ case in the paragraph following it.

6 Results

To test that this algorithm is working, normal ARIADNE was compared to ARIADNE corrected to first order. In this case the two algorithms should produce the same result, since ARIADNE is, as discussed at the end of subsection 4.3, already corrected to first order. The events fed to normal ARIADNE were the 0 jet events created by `MadGraph`, in order to ensure that things specific to the event generation were the same in both cases.

In the comparison, the $k_\perp$-clustering algorithm [9] was used. A brief review of this is therefore given, before discussing some complications to the correction algorithm above, and then finally concluded by the actual results.

6.1 The $k_\perp$-Clustering Algorithm

The point of a jet clustering algorithm is to cluster the final state partons of the event into jets in order to find information about the structure of the interaction. This can be used on both theoretical simulations as well as experimental data for comparison.

The $k_\perp$-clustering algorithm performs its clusterings in two steps, each of which have a number of possible choices. First it calculates so-called resolution variables between all particle pairs as well as between all particles and the beam. These are meant to be measures of how close the particles are in phase space, and thus how reasonable it is for them to be part of the same jet.

The minimum value of the resolution variable is then chosen, and the corresponding particles are clustered together to a pseudoparticle according some recombination scheme. This process is continued until all particles and pseudoparticles lie further apart than a certain cut-off on the resolution variables.

In this thesis the covariant E scheme was used, with all variables defined in the lab frame. The resolution variable between particles i and j is then chosen as d_{ij}, defined by equation (4), ie. the same variable that was used to define the cut on the ME. This enables one to clearly identify where the cut-off between the ME generated state and the cascade is.

Also, the resolution variable between a particle i and the beam is chosen to be $d_{iB} = p_{\perp i}^2$, and when two particles i and j are clustered together the resulting pseudoparticle is given a four-momentum $p_i + p_j$. No hadronization was used.

6.2 Complications

It was eventually revealed that there are some complications which is not accounted for in the correction algorithm presented above. In time these problems will have to be solved, but for the moment being they have been minimized, in order to first test the basic algorithm.

One such problem has to do with the different handling of the proton remnant between the two procedures. In normal ARIADNE the proton remnant is divided prior to the first emission of the initial dipole, whereas `MadGraph` does no such division at all, but rather this is simulated before feeding an event to ARIADNE. This means that the division in the corrected version is done *after* what gets reconstructed into an initial emission. To avoid this problem, an extra cut was set on the ME in the hadronic center of mass frame, excluding emissions close to the proton remnant. However, experiments at HERA would not be sensitive to emissions in this region.

Another issue is the masses of the quarks. `MadGraph` sets the mass of the u, d, s and c quarks to zero and ignores the t and b quarks. Therefore ARIADNE was told to use quark masses set to zero as well.

Also, recoil gluons were turned off, and the struck quark was set as a point-like object, contrary to how ARIADNE is normally run.

Finally, it turned out that there was a mistake in the BGF ME matching in normal ARIADNE. Naturally, ARIADNE was run with an option to correct for this.

6.3 Results

The result of the comparison can be seen in figure 8 with a cut-off at 4 GeV. What is plotted is the resolution variable at which the $k_\perp$-clustering algorithm considers the event to change from a 1 jet to a 0 jet event, d_2. Both normal ARIADNE and ARIADNE corrected to first order, as well as the 0 and 1 jet events parts of the latter are shown.

All events

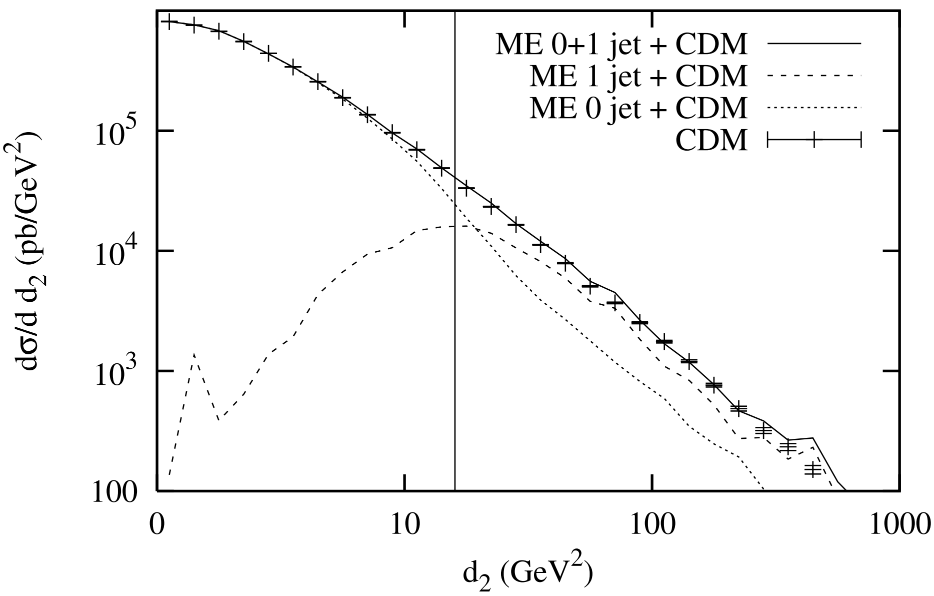

Figure 8: Comparison between the normal CDM and the combined approach.

As can be seen the transition around the cut-off is very smooth, which as very reassuring. Also, the two curves agree well, except at the very far end of the d_2 axis. This is not alarming though, since statistics is very low here.

In order to make a more detailed analysis, the events were separated into two cases: leading initial state gluon radiation (QCD-C) and initial state sea quark emission (BGF). The results of such a comparison can be seen in figures 9 and 10. It should perhaps be noted that there exists events which ARIADNE never cascades, so for the 0 jet case not all events fit into either of these two categories.

In both cases there is still a smooth transition around the cut-off, and again the curves show good agreement except for very high d_2. The BGF case should be contrasted to [8], where the corrected procedure had a tendency to lie a bit above normal ARIADNE. The reason for this was probably a bug in the Sudakov reweighting which has now been fixed.

Note also that the smooth transition around the cut-off is not trivial. Other approaches [10] typically has a noticeable discontinuity around the cut-off, probably due to that their method forces an analytic approximation of the Sudakov form factors, rather than the exact evaluation described above.

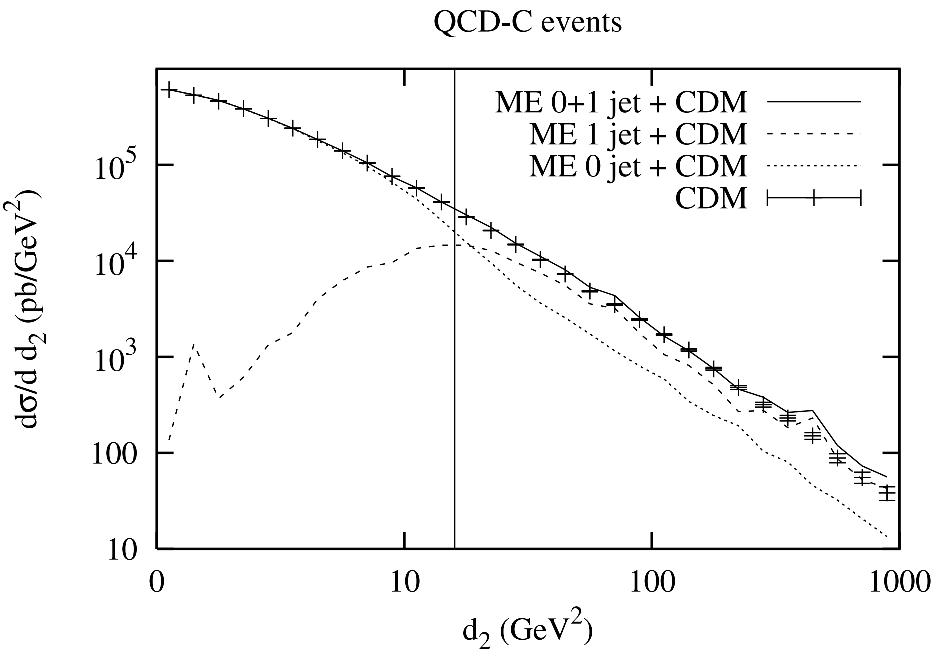

Figure 9: Comparison between the normal CDM and the combined approach, for initial state gluon emission.

7 Conclusion

The algorithm to correct the colour dipole cascade with fixed order matrix elements presented in [3] was modified for the case of deep inelastic scattering. Events were generated with the matrix element event generator `MadEvent` and then feed to ARIADNE to reconstruct and cascade upon. The modifications of the algorithm already carried out in [8] were generalized, and tested for the simplest case, with promising result.

Although not done in this thesis, the presented correction procedure is easy to extend to higher orders which is not possible to do for the leading order corrections already present in standard ARIADNE. The reconstruction and matching procedure for higher orders has already been implemented, but further tests are needed before confronting the program with data.

Acknowledgements

I wish to thank my supervisor Leif Lönnblad for the continuing support throughout the project, as well as the rest of the staff at the high energy

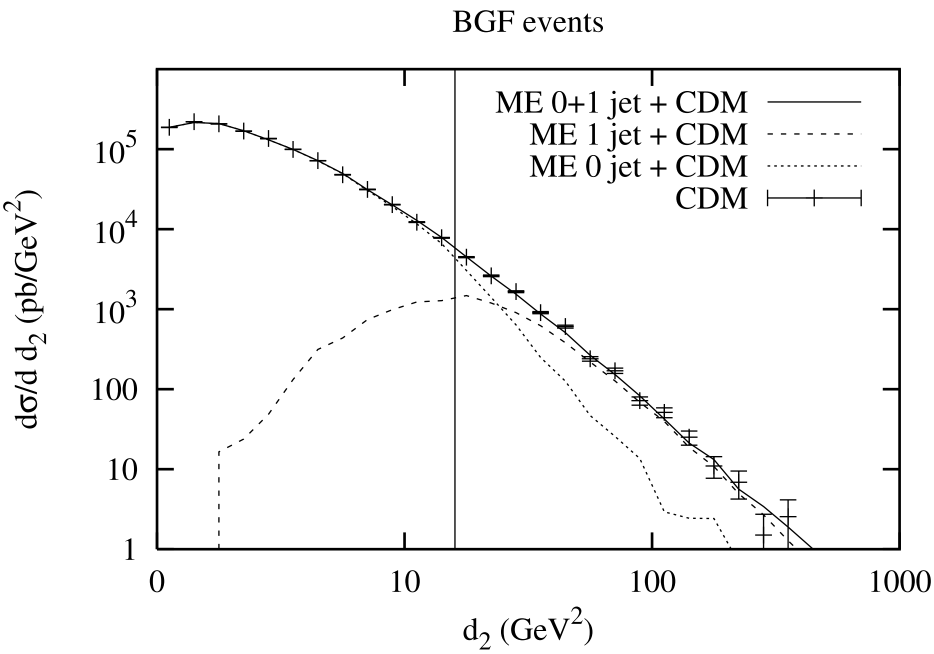

Figure 10: Comparison between the normal CDM and the combined approach, for initial state sea quark emission.

physics group for always gladly answering any question.

I am also in debt with the guys at "Vinterpalatset" for all the singing and joking, and my friends for moral support.

References

[1] G. Gustafson, "An Intuitive Semiclassical Picture of Proton Structure at Small x," Acta Phys. Polon. **B34** (2003) 2963-2988.

[2] L. Lönnblad, "ARIADNE version 4: A Program for Simulation of QCD-Cascades Implementing the Colour Dipole Model," Comput. Phys. Commun. **71** (1992) 15.

[3] L. Lönnblad, "Correcting the Colour-Dipole Cascade Model with Fixed Order Matrix Elements," JHEP **0205** (2002) 046.

[4] T. Stelzer and W. F. Long, "Automatic Generation of Tree Level Helicity Amplitudes," Nucl. Phys. Proc. Suppl. **37B** (1994) 158.

[5] H. Murayama, I. Watanabe and K. Hagiwara, "HELAS: HELicity Amplitude Subroutines for Feynman Diagram Evaluations," KEK Report 91-11, January 1992.

[6] F. Maltoni and T. Stelzer, "MadEvent: Automatic Event Generation with MadGraph," JHEP **0302** (2003) 027.

[7] E. Boos et. al, "Generic User Process Interface for Event Generators," Physics at TeV Colliders II Workshop, Les Houches, France, May 2001.

[8] N. Lavesson, "Correcting the Dipole Cascade with Matrix Elements in Hadron Collisions," Master of Science Thesis, Lund University, Sweden, March 2004.

[9] S. Catani, Yu.L. Dokshitzer, M.H. Seymour and B.R. Webber, "Longitudinally-invariant $k_\perp$-clustering algorithms for hadron-hadron collisions," Nucl. Phys. **B406** (1993) 187.

[10] S. Mrenna and P. Richardson, "Matching Matrix Elements and Parton Showers with HERWIG and PYTHIA."