Published by Institute of Physics Publishing for SISSA/ISAS
Received: *February 10, 2004*
Revised: *March 22, 2004*
Accepted: *March 23, 2004*

Multiple interactions and the structure of beam remnants

Torbjörn Sjöstrand and Peter Zeiler Skands

Department of Theoretical Physics, Lund University
Sölvegatan 14A, S-223 62 Lund, Sweden
E-mail: torbjorn@thep.lu.se, peter.skands@thep.lu.se

Abstract: Recent experimental data have established some of the basic features of multiple interactions in hadron-hadron collisions. The emphasis is therefore now shifting, to one of exploring more detailed aspects. Starting from a brief review of the current situation, a next-generation model is developed, wherein a detailed account is given of correlated flavour, colour, longitudinal and transverse momentum distributions, encompassing both the partons initiating perturbative interactions and the partons left in the beam remnants. Some of the main features are illustrated for the Tevatron and the LHC.

Keywords: QCD, Phenomenological Models, Hadronic Colliders.

JHEP03(2004)053

Contents

1. Introduction

The physics of high-energy hadron-hadron interactions has become a topic of increasing interest in recent years. With the Tevatron Run II well under way and with the startup of the LHC drawing closer, huge data samples are becoming available that will challenge our current understanding of this physics. From the point of view of QCD, many interesting questions remain to be answered, and we shall take up some of these in detail below. Moreover, for new physics searches and precision measurements, it is important that these questions can be given meaningful and trustworthy answers, since ever-present yet poorly-understood aspects of QCD can have a significant impact.

Much of the complexity involved in describing these phenomena — specifically the underlying event and minimum-bias collisions — derives from the composite nature of hadrons; we are dealing with objects which possess a rich internal structure that is not calculable from perturbation theory. This, however, does *not* imply that the physics of the underlying event as such has to be an inherently non-perturbative quagmire.

Viewing hadrons as 'bunches' of incoming partons, it is apparent that when two hadrons collide it is possible that several distinct pairs of partons collide with each other, as depicted in figure 1. Thus multiple interactions (also known as multiple scatterings) in hadronic collisions is a phenomenon which is a direct consequence of the composite nature of hadrons and which *must* exist, at some level. In fact, by extending simple perturbation theory to rather low $p_\perp$ values, though still some distance above Λ_{QCD}, most inelastic events in high-energy hadronic collisions are guaranteed to contain several *perturbatively calculable* interactions [1]. Furthermore, such interactions — even when soft — can be highly important, causing non-trivial changes to the colour topology of the colliding system as a whole, with potentially drastic consequences for the particle multiplicity in the final state.

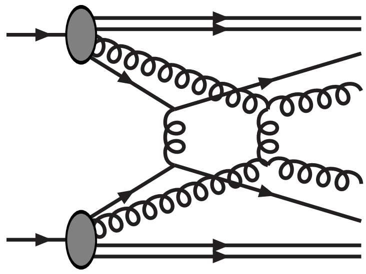


Figure 1: Schematic illustration of an event with two $2 \to 2$ perturbative interactions.

Nevertheless, traditionally the exploration of multiple interactions has not attracted much interest. For studies concentrating on high-$p_\perp$ jets, perturbative QCD emission is a more important source of multijets than separate multiple interactions. The underlying event, on the other hand, has in this context often been viewed as a mess of soft QCD interactions, that cannot be described from first principles but is better simply parametrized.

However, such parametrizations, even while reasonably successful in describing the average underlying activity, are not sophisticated enough to adequately describe correlations and fluctuations. This relates for instance to jet profiles and jet pedestals, and to systematic as well as random shifts in jet energies. The lack of sophistication implies that, even when tuned to describe a few distributions well, they could not be trusted for extrapolations beyond the fit region. Hence, a sound understanding of multiple interactions is prerequisite for precision physics involving jets and/or the underlying event.

It is interesting to note that this can also impact physics studies in areas well beyond the conventional QCD ones. As an example, consider the search for a Higgs particle in the $\mathrm{h}^0 \to \gamma\gamma$ channel at the LHC, where the Higgs mass resolution at high luminosity depends on picking the correct vertex between maybe 30 different pp events. If the event that contained the Higgs is special, by typically containing more charged particles (in general or in some regions of phase space), we would want to use that information to clean up the signal [2].

The crucial leap of imagination is to postulate that *all* particle production in inelastic hadronic collisions derives from the multiple-interactions mechanism. This is not to say that many nonperturbative and poorly known phenomena will not force their entrance on the stage, in going from the perturbative interactions to the observable hadrons, but the starting point is perturbative. If correct, this hypothesis implies that the typical Tevatron hadron-hadron collision may contain something like 2-6 interactions, and the LHC one 4-10.

A few models based on this picture were presented several years ago [1], and compared with the data then available. Though these models may still be tuned to give a reasonable description of the underlying event at various collider energies, several shortcuts had to be taken, particularly in the description of the nonperturbative aspects alluded to above. For instance, it was not possible to consider beam remnants with more than one valence quark kicked out.

The increased interest and the new data now prompts us to develop a more realistic framework for multiple interactions than the one in ref. [1], while making use of many of the same underlying ideas. This may not necessarily result in significant improvements for fits made to only a few distributions at a time, but we hope it will enhance our ability to simultaneously describe many different observations inside one framework, thereby improving the confidence with which we can make extrapolations from known measurements to new distributions and to higher energies.

One of the building blocks for the new model comes from our recent study of baryon-number-violating processes [3]. We then had to address the hadronization of colour topologies of the same kind as found in baryons. Specifically, as an extension of the standard Lund string fragmentation framework [4], we explored the concept of a junction in which three string pieces meet, with a quark at the end of each string. This also opens the way to a more realistic description of multiple interaction events.

The resulting improvements, relative to the framework in [1], are especially notable in the description of the structure of the incoming hadrons, i.e. how flavours, colours, transverse and longitudinal momenta are correlated between all the partons involved, both those that undergo interactions and those that are left behind in the remnants. (Brief descriptions of some of these aspects can also be found in [5].) To give one specific example, we introduce parton densities that are modified according to the flavours already affected by interactions.

Clearly, the model we present here is not the final word. For instance, we defer for the future any discussions of processes involving photons in the initial state and whether and how diffractive topologies could arise naturally from several interactions with a net colour singlet exchange. More generally, the whole issue of colour correlations will require further studies. The model also allows some options in a few places. A reasonable range of possibilities can then be explored, and (eventually) experimental tests should teach us more about the course taken by nature.

This article is organized as follows. In section 2 we give an introduction to multiple interactions in general, to the existing multiple interactions machinery, to other theoretical models, and to experimental data of relevance. Sections 3–5 then describe the improvements introduced by the current study: 3 a new option for impact-parameter dependence, 4 the main work on flavour and momentum space correlations, and 5 the very difficult topics of colour correlations and junction hadronization. Finally section 6 provides some further examples of the resulting physics distributions and important tests, while section 7 contains a summary and outlook.

2. Multiple interactions minireview

2.1 Basic concepts

The cross section for QCD hard $2 \to 2$ processes, as a function of the $p_\perp^2$ scale, is given by

$$\frac{\mathrm{d}\sigma_{\mathrm{int}}}{\mathrm{d}p_\perp^2} = \sum_{i,j,k} \int \mathrm{d}x_1 \int \mathrm{d}x_2 \int \mathrm{d}\hat{t} f_i(x_1, Q^2)\, f_j(x_2, Q^2) \frac{\mathrm{d}\hat{\sigma}_{ij \to kl}}{\mathrm{d}\hat{t}} \delta\left(p_\perp^2 - \frac{\hat{t}\hat{u}}{\hat{s}}\right), \tag{2.1}$$

where $\hat{s} = x_1 x_2 s$. The jet cross section is twice as large, $\sigma_{\mathrm{jet}} = 2\sigma_{\mathrm{int}}$, since each interaction gives rise to two jets, to first approximation. In the following, we will always refer to the interaction rather than the jet cross section, unless otherwise specified. We will also assume that the 'hardness' of processes is given by the $p_\perp$ scale of the interaction, i.e. $Q^2 = p_\perp^2$.

The cross section for QCD $2 \to 2$ processes, the sum of $\mathrm{qq}' \to \mathrm{qq}'$, $\mathrm{q\overline{q}} \to \mathrm{q'\overline{q}'}$, $\mathrm{q\overline{q}} \to \mathrm{gg}$, $\mathrm{qg} \to \mathrm{qg}$, $\mathrm{gg} \to \mathrm{gg}$ and $\mathrm{gg} \to \mathrm{q\overline{q}}$, is dominated by t-channel gluon exchange contributions. (This justifies the use of 'interaction' and 'scattering' as almost synonymous.) In the $|\hat{t}| \ll \hat{s}$ limit, where $p_\perp^2 = \hat{t}\hat{u}/\hat{s} \approx |\hat{t}|$, quark and gluon interactions just differ by the colour factors, so approximately we may write

$$\frac{\mathrm{d}\sigma_{\mathrm{int}}}{\mathrm{d}p_\perp^2} \approx \int\int \frac{\mathrm{d}x_1}{x_1}\frac{\mathrm{d}x_2}{x_2} F(x_1, p_\perp^2)\, F(x_2, p_\perp^2) \frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}p_\perp^2}, \tag{2.2}$$

where

$$\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}p_\perp^2} = \frac{8\pi\alpha_{\mathrm{s}}^2(p_\perp^2)}{9p_\perp^4}, \tag{2.3}$$

and

$$F(x, Q^2) = \sum_q \left(x\, q(x, Q^2) + x\, \overline{q}(x, Q^2)\right) + \frac{9}{4} x\, g(x, Q^2). \tag{2.4}$$

Thus, for constant α_{s} and neglecting the x integrals, the integrated cross section above some $p_{\perp\mathrm{min}}$ is divergent in the limit $p_\perp \to 0$:

$$\sigma_{\mathrm{int}}(p_{\perp\mathrm{min}}) = \int_{p_{\perp\mathrm{min}}}^{\sqrt{s}/2} \frac{\mathrm{d}\sigma}{\mathrm{d}p_\perp} \mathrm{d}p_\perp \propto \frac{1}{p_{\perp\mathrm{min}}^2}. \tag{2.5}$$

A numerical illustration of this divergence is given in figure 2. Note that the actual fall-off is everywhere steeper than $1/p_\perp^2$. We have here used full $2 \to 2$ matrix elements and the CTEQ 5L parton density parametrizations [6], which are valid for $Q > 1.1\,\mathrm{GeV}$ and $x > 10^{-6}$; therefore results at the lowest $p_\perp$ values are not to be taken too literally. For the studies in this article we are basing ourselves on leading-order cross sections and parton

JHEP03(2004)053

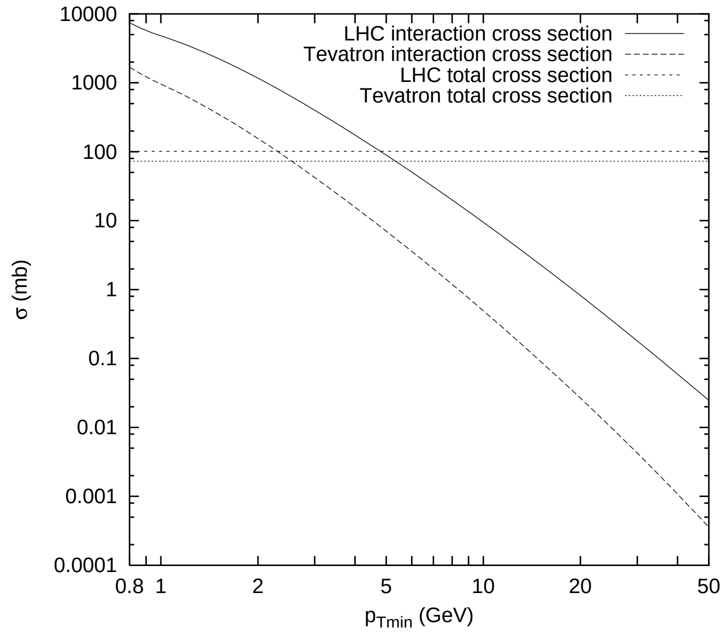


Figure 2: The integrated interaction cross section $\sigma_{\rm int}$ above $p_{\perp\rm min}$ for the Tevatron, with 1.8 TeV $\mathrm{p\overline{p}}$ collisions, and the LHC, with 14 TeV pp ones. For comparison, the flat lines represent the respective total cross section.

densities, with nontrivial higher-order corrections only approximately taken into account by the addition of parton showers. Nevertheless, the trend is quite clear, with an integrated cross section that exceeds the total $\mathrm{pp/\overline{p}p}$ cross section $\sigma_{\rm tot}$ (in the parametrization of ref. [7]) for $p_{\perp\rm min}$ of the order of a few GeV. As already mentioned in the introduction, this is well above $\Lambda_{\rm QCD}$, so one cannot postulate a breakdown of perturbation theory in the conventional sense.

The resolution of the $\sigma_{\rm int} > \sigma_{\rm tot}$ paradox probably comes in two steps.

Firstly, the interaction cross section is an inclusive number. Thus, if an event contains two interactions it counts twice in $\sigma_{\rm int}$ but only once in $\sigma_{\rm tot}$, and so on for higher multiplicities. Thereby we may identify $\langle n \rangle (p_{\perp\rm min}) = \sigma_{\rm int}(p_{\perp\rm min})/\sigma_{\rm tot}$ with the average number of interactions above $p_{\perp\rm min}$ per event, and that number may well be above unity.

One of the problems we will consider further in this article is that this simple calculation of $\langle n \rangle (p_{\perp\rm min})$ does not take into account energy-momentum conservation effects. Specifically, the average $\hat{s}$ of a scattering decreases slower with $p_{\perp\rm min}$ than the number of interactions increases, so naively the total amount of scattered partonic energy becomes infinite. Thus corrections reduce the $\langle n \rangle (p_{\perp\rm min})$ number, but not sufficiently strongly: one is lead to a picture with too little of the incoming energy remaining in the small-angle beam jet region [1].

Secondly, a more credible reason for taming the rise of $\langle n \rangle (p_{\perp\rm min})$ is that the incoming hadrons are colour singlet objects. Therefore, when the $p_\perp$ of an exchanged gluon is made small and the transverse wavelength correspondingly large, the gluon can no longer resolve the individual colour charges, and the effective coupling is decreased. Note that

JHEP03(2004)053

perturbative QCD calculations are always performed assuming free incoming and outgoing quark and gluon states, rather than partons inside hadrons, and thus do not address this kind of nonperturbative screening effects.

A naive estimate of an effective lower cutoff would be

$$p_{\perp\min} \simeq \frac{\hbar}{r_{\mathrm{p}}} \approx \frac{0.2\ \mathrm{GeV}\cdot\mathrm{fm}}{0.7\ \mathrm{fm}} \approx 0.3\,\mathrm{GeV} \simeq \Lambda_{\mathrm{QCD}}\,, \tag{2.6}$$

but this again appears too low. The proton radius r_{p} has to be replaced by the typical colour screening distance d, i.e. the average size of the region within which the net compensation of a given colour charge occurs. This number is not known from first principles, so effectively one is forced to introduce some kind of cutoff parameter, which can then just as well be put in transverse momentum space. The simplest choice is to introduce a step function $\theta(p_\perp - p_{\perp\min})$, such that the perturbative cross section completely vanishes below the $p_{\perp\min}$ scale. A more realistic alternative is to note that the jet cross section is divergent like $\alpha_{\mathrm{s}}^2(p_\perp^2)/p_\perp^4$, and that therefore a factor

$$\frac{\alpha_{\mathrm{s}}^2(p_{\perp0}^2 + p_\perp^2)}{\alpha_{\mathrm{s}}^2(p_\perp^2)} \frac{p_\perp^4}{(p_{\perp0}^2 + p_\perp^2)^2} \tag{2.7}$$

would smoothly regularize the divergences, now with $p_{\perp0}$ as the free parameter to be tuned to data. Empirically the two procedures give similar numbers, $p_{\perp\min} \approx p_{\perp0}$, and both of the order of 2 GeV.

The parameters do not have to be energy-independent, however. Higher energies imply that parton densities can be probed at smaller x values, where the number of partons rapidly increases. Partons then become closer packed and the colour screening distance d decreases. Just like the small-x rise goes like some power of x one could therefore expect the energy dependence of $p_{\perp\min}$ and $p_{\perp0}$ to go like some power of CM energy. Explicit toy simulations [8] lend some credence to such an ansatz, although with large uncertainties. Alternatively, one could let the cutoff increase with decreasing x; this would lead to a similar phenomenology since larger energies probe smaller x values.

2.2 Our existing models

The models developed in ref. [1] have been implemented and are available in the Pythia event generator. They form the starting point for the refinements we will discuss further on, so we here review some of the main features.

The approach is not intended to cover elastic or diffractive physics, so the $\sigma_{\mathrm{int}}(p_{\perp\min}, s)$ or $\sigma_{\mathrm{int}}(p_{\perp0}, s)$ interaction cross section is distributed among the $\sigma_{\mathrm{nd}}(s)$ nondiffractive inelastic one [7, 10]. The average number of interactions per such event is then the ratio $\langle n\rangle = \sigma_{\mathrm{int}}/\sigma_{\mathrm{nd}}$. As a starting point we will assume that all hadron collisions are equivalent, i.e. without an impact parameter dependence, and that the different parton-parton interactions take place independently of each other. The number of interactions per event is then distributed according to a Poisson distribution with mean $\langle n\rangle$, $\mathcal{P}_n = \langle n\rangle^n \exp(-\langle n\rangle)/n!$.

One (not used) approach would be, for each new event, to pick the actual number of interactions n according to the poissonian, and select the n $p_\perp$ values independently according to eq. (2.1). One disadvantage is that this does not take into account correlations, even such basic ones as energy-momentum conservation: the sum of interaction energies may well exceed the total CM energy.

In an event with several interactions, it is therefore convenient to impose an ordering. The logical choice is to arrange the scatterings in falling sequence of $p_\perp$ values. The 'first' scattering is thus the hardest one, with the 'subsequent' ('second', 'third', etc.) successively softer. This terminology is not primarily related to any picture in physical time although, by the uncertainty relation, large momentum transfers implies short timescales. When averaging over all configurations of soft partons, one should effectively obtain the standard QCD phenomenology for a hard scattering, e.g. in terms of parton distributions. Correlation effects, known or estimated, can be introduced in the choice of subsequent scatterings, given that the 'preceding' (harder) ones are already known. This will be developed further in section 4.

The generation of a sequence $\sqrt{s}/2 > p_{\perp 1} > p_{\perp 2} > \cdots > p_{\perp n} > p_{\perp \mathrm{min}}$ now becomes one of determining $p_{\perp i}$ from a known $p_{\perp i-1}$, according to the probability distribution

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}p_{\perp i}} = \frac{1}{\sigma_{\mathrm{nd}}} \frac{\mathrm{d}\sigma}{\mathrm{d}p_\perp} \exp\left[- \int_{p_\perp}^{p_{\perp i-1}} \frac{1}{\sigma_{\mathrm{nd}}} \frac{\mathrm{d}\sigma}{\mathrm{d}p'_\perp} \mathrm{d}p'_\perp \right] . \tag{2.8}$$

The exponential expression is the 'form factor' from the requirement that no interactions occur between $p_{\perp i-1}$ and $p_{\perp i}$, cf. radioactive decays or the Sudakov form factor [11] of parton showers.

When used with the standard differential cross section $\mathrm{d}\sigma/\mathrm{d}p_\perp$, eq. (2.8) gives the same Poisson distribution as above. This time n is not known beforehand, but is defined by the termination of the iterative procedure. Now, however, $\mathrm{d}\sigma/\mathrm{d}p_\perp$ can be modified to take into account the effects of the $i-1$ preceding interactions. Specifically, parton distributions are not evaluated at x_i for the i'th scattered parton from a hadron, but at the rescaled value

$$x'_i = \frac{x_i}{1 - \sum_{j=1}^{i-1} x_j} , \tag{2.9}$$

so that it becomes impossible to scatter more energy than initially available in the incoming beam. This also dynamically suppresses the high-multiplicity tail of the poissonian and thereby reduces the average number of interactions.

In a fraction of the events studied, there will be no hard scattering at all above $p_{\perp \mathrm{min}}$. Such events are associated with nonperturbative low-$p_\perp$ physics, and are simulated by exchanging a very soft gluon between the two colliding hadrons, making the hadron remnants colour-octet objects rather than colour-singlet ones. If only valence quarks are considered, the colour-octet state of a baryon can be decomposed into a colour triplet quark and an antitriplet diquark. In a baryon-baryon collision, one would then obtain a two-string picture, with each string stretched from the quark of one baryon to the diquark of the other. A baryon-antibaryon collision would give one string between a quark and an antiquark and another one between a diquark and an antidiquark.

JHEP03(2004)053

In a hard interaction, the number of possible string drawings are many more, and the overall situation can become quite complex. In the studies preceding this work, several simplifications were made. The hardest interaction was selected with full freedom of flavour choice and colour topology, but for the subsequent ones only three simple recipes were available:

- Interactions of the gg $\to$ gg type, with the two gluons in a colour-singlet state, such that a double string is stretched directly between the two outgoing gluons, decoupled from the rest of the system.

- Interactions gg $\to$ gg, but colour correlations assumed to be such that each of the gluons is connected to one of the strings 'already' present. Among the different possibilities of connecting the colours of the gluons, the one which minimizes the total increase in string length is chosen. This is in contrast to the previous alternative, which roughly corresponds to a maximization (within reason) of the extra string length.

- Interactions gg $\to$ q$\overline{\mathrm{q}}$, with the final pair again in a colour-singlet state, such that a single string is stretched between the outgoing q and $\overline{\mathrm{q}}$.

The three possibilities can be combined in suitable fractions.

Many further approximations were also required in the old framework, e.g. the addition of initial- and final-state parton showers was feasible only for the hardest interaction, and we address several of those in the following.

The model also includes several options for the impact-parameter dependence. This offers an additional element of variability: central collisions on the average will contain more interactions than peripheral ones. Even if a Poisson distribution in the number of interactions would be assumed for each impact parameter separately, the net result would be a broader-than-Poisson distribution. The amount of broadening can be 'tuned' by the choice of impact-parameter profile. We discuss this further in section 3, where a new set of profiles is studied.

The above framework was originally formulated for p$\overline{\mathrm{p}}$/pp collisions, but has also been extended to γp and $\gamma\gamma$ interactions [12]. In these latter processes, however, the nature of the photon needs to be modelled in detail, and this introduces many further uncertainties. A study of such aspects is beyond the scope of the current article.

2.3 Other models

While the models of ref. [1] may well be the ones most frequently studied, owing to their implementation in Pythia [9], a number of other models also exist. Many of the basic concepts have also been studied separately. We here give a few examples, without any claim of completeness.

In Dual Topological Unitarization (DTU) language [13], and the Dual Parton Model based on it [14], or other similar techniques [15], inelastic events are understood in terms of cut pomerons [16]. Translated into modern terminology, each cut pomeron corresponds to the exchange of a soft gluon, which results in two 'strings' being drawn between the two beam remnants. Uncut pomerons give virtual corrections that help preserve unitarity. A variable number of cut pomerons are allowed. This approach has been the basis for the simulation of underlying events in ISAJET [17], and was the starting point for DTUJET [18]. However, note that cut pomerons were originally viewed as purely soft objects, and so did not generate any transverse momentum, unlike the multiple interactions considered in this article. In DTUJET and its PHOJET [19] and DPMJET [20] relatives, however, also hard interactions have been included, so that the picture now is one of both hard and soft pomerons, ideally with a smooth transition between the two. Since the three related programs make use of the PYTHIA hadronization description, the differences relative to our scenarios is more a matter of details (but "the devil is in the details") than of any basic incompatibility.

The possibility of observing two separate hard interactions has been proposed since long [21], and from that has also developed a line of studies on the physics framework for having several hard interactions [22], also involving e.g. electroweak processes [23]. Again this is similar to what we do, except that lower $p_\perp$ values and the transition to nonperturbative physics are not normally emphasized.

The possibility of multiple interactions has also been implicit [24] or explicit [25] in many calculations of total cross sections for hadron-hadron, hadron-γ and $\gamma\gamma$ events. The increase of σ_{int} with CM energy is here directly driving an increase also of σ_{tot}; that the latter is rising slower than the former comes out of an eikonalization procedure that implies also an increasing $\langle n_{\mathrm{int}} \rangle$.

Multiple interactions require an ansatz for the structure of the incoming beams, i.e. correlations between the constituent partons. Some of these issues have been studied, e.g. with respect to longitudinal momenta [26, 14], colours [27] or impact parameter [28], but very little of this has been tested experimentally. Dense-packing of partons could become an issue [15], of unknown importance, but not necessarily a major one [29].

The HERWIG [30] event generator does not contain any physics simulation of multiple interactions. Instead a parametrization procedure originally suggested by UA5 [31] is used, without any underlying physics scenario. It thus does parametrize multiplicity and rapidity distributions, but does not contain any minijet activity in the underlying event. The add-on JIMMY package [32] offers a multiple-interaction component (both for pp, γp and $\gamma\gamma$ events), which has recently been extended to include also a model of soft interactions [33].

The introduction of unintegrated parton densities, as used in the BFKL/CCFM/LDC approaches to initial-state radiation [34, 35, 36], allows the possibility to replace our $p_{\perp\mathrm{min}}/p_{\perp 0}$ cutoff by parton densities that explicitly vanish in the $p_\perp \to 0$ limit [37]. This opens up the possibility of an alternative implementation of multiple interactions [38].

In heavy-ion collisions the multiple interactions rate can become huge [39]. For small impact parameters, a major fraction of the energy of the two incoming nuclei is carried by partons undergoing perturbative interactions, and which therefore define a 'resolved' partonic content. This suggests a mechanism for the construction of an 'initial state' for the continued formation and thermalization (or not) of a quark-gluon plasma.

JHEP03(2004)053

2.4 Experimental tests

Experimental input to the understanding of multiple interactions comes in essentially three categories: direct observation of double parton scattering, event properties that directly and strongly correlate with multiple interactions, and event properties that do not point to multiple interactions by themselves but still constrain multiple interactions models.

If an event contains two uncorrelated $2 \to 2$ interactions, we expect to find four jets, grouped in two pairs that each internally have roughly opposite and compensating transverse momenta, and where the relative azimuthal angle between the scattering planes is isotropic. Neither of these properties are expected in a $2 \to 4$ event, where two of the partons can be thought of as bremsstrahlung off a basic $2 \to 2$ process. The problem is that $2 \to 4$ processes win out at large $p_{\perp}$, so there is a delicate balance between having large enough jet $p_{\perp}$ that the jets can be well measured and still not so large that the signal drowns.

When the $p_{\perp\mathrm{min}}$ of the jets is sufficiently large that $\exp(-\langle n \rangle) \approx 1$, Poisson statistics implies that $\mathcal{P}_2 = \mathcal{P}_1^2/2$, where $\mathcal{P}_i$ is the probability to have i interactions. Traditionally this is rewritten as

$$\sigma_2 = \sigma_{\mathrm{nd}} \left(\frac{\sigma_1}{\sigma_{\mathrm{nd}}} \right)^2 \frac{1}{2} \frac{\sigma_{\mathrm{nd}}}{\sigma_{\mathrm{eff}}} = \frac{1}{2} \frac{\sigma_1^2}{\sigma_{\mathrm{eff}}} \,, \tag{2.10}$$

where the ratio $\sigma_{\mathrm{nd}}/\sigma_{\mathrm{eff}}$ gauges deviations from the poissonian ansatz. Values above unity, i.e. $\sigma_{\mathrm{eff}} < \sigma_{\mathrm{nd}}$, arise naturally in models with variable impact parameter.

The first observation of double parton scattering is by AFS [40]. The subsequent UA2 study [41] ends up quoting an upper limit, but has a best fit that requires them. A CDF study [42] also found them. These experiments all had to contend with limited statistics and uncertain background estimates. The strongest signal has been obtained with a CDF study involving three jets and a hard photon [43]; here $\sigma_2 = \sigma_A \sigma_B / \sigma_{\mathrm{eff}}$, without a factor 1/2, since the two $2 \to 2$ processes A and B are inequivalent. In all cases, including the UA2 best fit, σ_{eff} comes out smaller than σ_{nd}; typically the double parton scattering cross section is a factor of three to four larger than the poissonian prediction. For instance, the CDF number is $\sigma_{\mathrm{eff}} = 14.5 \pm 1.7^{+1.7}_{-2.3}$ mb. More recently, ZEUS has observed a signal in γp events [44]. The D0 four-jet study shows the need to include multiple interactions, but this is not quantified [45].

So far, no direct tests of triple or more parton scattering exist. However, the UA1 minijet rates [46], going up to 5 jets, are difficult to understand without such events.

Tests involving jets at reasonably large $p_{\perp}$ values do not probe the total rate of multiple interactions, since the bulk of the interactions occur at so small $p_{\perp}$ values that they cannot be identified as separate jets. By the way colours are drawn across the event, soft partons can drive the production of particles quite out of proportion to the $p_{\perp}$ values involved, however. The multiplicity distribution of multiple interactions thereby strongly influences the multiplicity distribution of charged hadrons, n_{ch}.

A notable aspect here is that the measured n_{ch} distribution, when expressed in the KNO variable $z = n_{\mathrm{ch}}/\langle n_{\mathrm{ch}} \rangle$ [47], is getting broader with increasing CM energy [48, 49]. This is contrary to the essentially poissonian hadronization mechanism of the string model, where the KNO distribution becomes narrower. As an example, consider the UA5 measurements at 900 GeV [48], where $\langle n_{\mathrm{ch}} \rangle = 35.6$ and $\sigma(n_{\mathrm{ch}}) = 19.5$, while the poissonian prediction would be $\sigma(n_{\mathrm{ch}}) = \sqrt{\langle n_{\mathrm{ch}} \rangle} = 6.0$. It is possible to derive approximate KNO scaling in $\mathrm{e}^+\mathrm{e}^-$ annihilation [50] , but this then rests on having a perturbative shower that involves the whole CM energy. However, allowing for at most one interaction in $\mathrm{p\overline{p}}$ events and assuming that hadronization is universal (so it can be tuned to $\mathrm{e}^+\mathrm{e}^-$ data), there is no (known) way to accommodate the experimental multiplicity distributions, neither the rapid increase with energy of the average nor the large width. Either hadronization is *very* different in hadronic events from $\mathrm{e}^+\mathrm{e}^-$ ones, or one must accept multiple interactions as a reality. (Unfortunately it is difficult to test the 'hadronization universality' hypothesis completely separated from the multiple interactions and other assumptions. To give two examples, the relative particle flavour composition appears to be almost but not quite universal [51], and low-mass diffractive events display 'string-like' flavour correlations [52].)

Further support is provided by the study of forward-backward multiplicity correlations. For instance, UA5 and E735 define a 'forward' n_F and a 'backward' n_B multiplicity in pseudorapidity windows of one unit each, separated by a $\Delta\eta$ variable-width gap in the middle [53]. A forward-backward correlation strength is now defined by

$$b = \frac{\langle (n_F - \langle n_F \rangle)\,(n_B - \langle n_B \rangle) \rangle}{\sigma(n_F)\,\sigma(n_B)} = \frac{\langle n_F n_B \rangle - \langle n_F \rangle^2}{\langle n_F^2 \rangle - \langle n_F \rangle^2}\,, \tag{2.11}$$

where the last equality holds for a symmetric η distribution, i.e. for $\mathrm{pp}/\mathrm{p\overline{p}}$ but not for $\gamma\mathrm{p}$. Measurements give a positive and surprisingly large b, also for $\Delta\eta$ of several units of rapidity. So it appears that there is some global quantity, different for each event, that strongly influences particle production in the full phase space. Again known fragmentation mechanisms are too local, and effects of a single hard interaction not strong enough. But the number of multiple interactions is indeed a global quantity of the desired kind, and multiple-interaction models can describe the data quite well.

It is a matter of taste which evidence is valued highest. The direct observation of double parton scattering is easily recognized as evidence for the multiple-interactions concept, but only affects a tiny fraction of the cross section. By comparison, the broadening multiplicity distribution and the strong forward-backward correlations offer more indirect evidence, but ones strongly suggesting that the bulk of events have several interactions. We are not aware of any realistic alternative explanations for either of the observables.

Another interesting phenomenon is the pedestal effect: events with high-$p_\perp$ jets on the average contain more underlying activity than minimum-bias ones, also well away from jets themselves. It has been observed by several collaborations, like UA1 [54], CDF [55, 56] and H1 [57]. When the jet energy is varied from next-to-nothing to the highest possible, the underlying activity initially increases, but then flattens out around $p_{\perp\mathrm{jet}} = 10\,\mathrm{GeV}$ (details depending on the jet algorithm used and the CM energy). This fits very nicely with an impact-parameter-dependent multiple-interactions scenario: the presence of a higher $p_\perp$ scale biases events towards a smaller impact parameter and thereby a higher additional activity, but once $\sigma_{\mathrm{int}}(p_{\perp\mathrm{jet}}) \ll \sigma_{\mathrm{nd}}$ the bias saturates [1]. The height of the pedestal depends on the form of the overlap function $\mathcal{O}(b)$, and can thus be adjusted, while the $p_{\perp\mathrm{jet}}$ at which saturation occurs is rather model-insensitive, and in good agreement with the data.

The presence of pedestals also affects all measurements of jet profiles [59, 55]. It can lead to seemingly broader jets, when the full underlying event cannot be subtracted, and enrich the jet substructure, when a multiple-interactions jet is mistaken for radiation off the hard subprocess. It can also affect (anti)correlations inside a jet and with respect to the rest of the event [57].

Many further observables influence the modeling and understanding of multiple interactions, without having an immediate interpretation in those terms. A notable example here is the $\langle p_\perp \rangle (n_{\mathrm{ch}})$ distribution, i.e. how the average transverse momentum of charged particles varies as a function of the total charged multiplicity. The observed increasing trend [58] is consistent with multiple interactions: large multiplicity implies many interactions and therefore more perturbatively generated $p_\perp$ to be shared between the hadrons. For it to work, however, each new interaction should add proportionately less to the total n_{ch} than to the total $p_\perp$. Whether this is the case strongly depends on the colour connections between the interactions, i.e. whether strings tend to connect nearest neighbours in momentum space or run criss-cross in the event. A rising trend can easily be obtained, but it is a major challenge to get the quantitative behaviour right, as we shall see.

Finally, one should mention that global fits to hadron collider data [60]–[64] clearly point to the importance of a correct understanding of multiple interactions, and constrains models down to rather fine details. This brings together many of the aspects raised above, plus some further ones. A convenient reference for our continued discussion is Tune A, produced by R.D. Field, which is known to describe a large set of CDF minimum bias and jet data [61]. Relative to the defaults of the old scenario, it assumes $p_{\perp 0} = 2.0\,\mathrm{GeV}$ (`PARP(82)=2.0`) at the reference energy 1.8 TeV (`PARP(89)=1800.0`), with an energy rescaling proportional to $E_{\mathrm{cm}}^{1/4}$ (`PARP(90)=0.25`). It is based on a double gaussian matter distribution (`MSTP(82)=4`), with half of the matter (`PARP(83)=0.5`) in a central core of radius 40% of the rest (`PARP(84)=0.4`). Almost all of the subsequent interactions are assumed to be of the type $\mathrm{gg} \to \mathrm{gg}$ with minimal string length (`PARP(85)=0.9`, `PARP(86)=0.95`). Finally the matching of the initial-state showers to the hard scattering is done at a scale $Q^2_{\mathrm{shower}} = 4p^2_{\perp\mathrm{hard}}$ (`PARP(67)=4.0`).

The above parameter set is sensible, within the framework of the model, although by no means obvious. The matter distribution is intermediate between the extremes already considered in [1], while the string drawing is more biased towards small string lengths than foreseen there. The $p_{\perp 0}$ energy dependence is steeper than previously used, but in a sensible range, as follows. In reggeon theory, a Pomeron intercept of $1 + \epsilon$ implies a total cross section rising like s^ϵ, and a small-x gluon density like $xg(x) \propto x^{-\epsilon}$ (at small Q^2). A $p_{\perp 0}$ rising (at most) like s^ϵ would then be acceptable, while one rising significantly faster would imply a decreasing interaction cross section $\sigma_{\mathrm{int}}(p_{\perp 0})$, and by implication a decreasing σ_{tot}, in contradiction with data. The DL fit to σ_{tot} [7] gives $\epsilon \approx 0.08$, which would imply (at most) a $p_{\perp 0}$ dependence like $s^{0.08} = E_{\mathrm{cm}}^{0.16}$. However, σ_{tot} already represents the unitarization of multiple-pomeron exchanges, and the 'bare' pomeron intercept should be larger than this, exactly by how much being a matter of some debate [65]. A value like

$\epsilon_{\text{bare}} \approx 0.12$ is here near the lower end of the sensible range; the $xg(x)$ shape is consistent with a rather larger value. Since it is actually the bare pomeron that corresponds to a single interaction, an $E_{\text{cm}}^{0.25}$ behaviour is thereby acceptable.

3. Impact-parameter dependence

In the simplest multiple-interactions scenarios, it is assumed that the initial state is the same for all hadron collisions. More realistically, one should include the possibility that each collision also could be characterized by a varying impact parameter b [1]. Within the classical framework we use here, b is to be thought of as a distance of closest approach, not as the Fourier transform of the momentum transfer. A small b value corresponds to a large overlap between the two colliding hadrons, and hence an enhanced probability for multiple interactions. A large b, on the other hand, corresponds to a grazing collision, with a large probability that no parton-parton interactions at all take place.

In order to quantify the concept of hadronic matter overlap, one may assume a spherically symmetric distribution of matter inside a hadron at rest, $\rho(\mathbf{x})\,\mathrm{d}^3x = \rho(r)\,\mathrm{d}^3x$. For simplicity, the same spatial distribution is taken to apply for all parton species and momenta. Several different matter distributions have been tried. A gaussian ansatz makes the subsequent calculations especially transparent, but there is no reason why this should be the correct form. Indeed, it appears to lead to a somewhat too narrow multiplicity distribution and too little of a pedestal effect. The next simplest choice, that does provide more fluctuations, is a double gaussian

$$\rho(r) \propto \frac{1-\beta}{a_1^3} \exp\left\{-\frac{r^2}{a_1^2}\right\} + \frac{\beta}{a_2^3} \exp\left\{-\frac{r^2}{a_2^2}\right\}. \tag{3.1}$$

This corresponds to a distribution with a small core region, of radius a_2 and containing a fraction β of the total hadronic matter, embedded in a larger hadron of radius a_1. If we want to give a deeper meaning to this ansatz, beyond it containing two more adjustable parameters, we could imagine it as an intermediate step towards a hadron with three disjoint core regions ('hot spots'), reflecting the presence of three valence quarks, together carrying the fraction β of the proton momentum. One could alternatively imagine a hard hadronic core surrounded by a pion cloud. Such details would affect e.g. the predictions for the t distribution in elastic scattering, but are not of any consequence for the current topics.

For a collision with impact parameter b, the time-integrated overlap $\mathcal{O}(b)$ between the matter distributions of the colliding hadrons is given by

$$\begin{aligned}\mathcal{O}(b) &\propto \int \mathrm{d}t \int \mathrm{d}^3x \rho(x,y,z)\,\rho(x+b,y,z+t) \\ &\propto \frac{(1-\beta)^2}{2a_1^2} \exp\left\{-\frac{b^2}{2a_1^2}\right\} + \frac{2\beta(1-\beta)}{a_1^2+a_2^2} \exp\left\{-\frac{b^2}{a_1^2+a_2^2}\right\} + \frac{\beta^2}{2a_2^2} \exp\left\{-\frac{b^2}{2a_2^2}\right\}. \end{aligned} \tag{3.2}$$

The necessity to use boosted $\rho(\mathbf{x})$ distributions has been circumvented by a suitable scale transformation of the z and t coordinates. The overlap function $\mathcal{O}(b)$ is closely related to the $\Omega(b)$ of eikonal models [24], but is somewhat simpler in spirit.

JHEP03(2004)053

The larger the overlap $\mathcal{O}(b)$ is, the more likely it is to have interactions between partons in the two colliding hadrons. In fact, to first approximation, there should be a linear relationship

$$\langle \tilde{n}(b) \rangle = k\mathcal{O}(b) \,, \tag{3.3}$$

where $\tilde{n} = 0, 1, 2, \ldots$ counts the number of interactions when two hadrons pass each other with an impact parameter b. At this stage k is an undefined constant of proportionality, to be specified below.

For each given impact parameter, the number of interactions $\tilde{n}$ is assumed to be distributed according to a poissonian, $\mathcal{P}_{\tilde{n}} = \langle \tilde{n} \rangle^{\tilde{n}} \exp(-\langle \tilde{n} \rangle)/\tilde{n}!$, before energy-momentum and other constraints are included. If the matter distribution has a tail to infinity (as the gaussians do), events may be obtained with arbitrarily large b values. In order to obtain finite total cross sections, it is necessary to assume that each event contains at least one semi-hard interaction. (Unlike the simpler, impact-parameter-independent approach above, where $p_\perp = 0$ no-interaction events are allowed as a separate class.) The probability that two hadrons, passing each other with an impact parameter b, will produce a real event is then given by

$$\mathcal{P}_{\mathrm{int}}(b) = \sum_{\tilde{n}=1}^{\infty} \mathcal{P}_{\tilde{n}}(b) = 1 - \mathcal{P}_0(b) = 1 - \exp(-\langle \tilde{n}(b) \rangle) = 1 - \exp(-k\mathcal{O}(b)) \,, \tag{3.4}$$

according to Poisson statistics. The average number of interactions per event at impact parameter b is now $\langle n(b) \rangle = \langle \tilde{n}(b) \rangle / \mathcal{P}_{\mathrm{int}}(b)$, where the denominator comes from the removal of hadron pairs that pass without interaction, i.e. which do not produce any events. While the removal of $\tilde{n} = 0$ from the potential event sample gives a narrower-than-Poisson interaction distribution at each fixed b, the variation of $\langle n(b) \rangle$ with b gives a b-integrated broader-than-Poisson interaction multiplicity distribution.

Averaged over all b the relationship $\langle n \rangle = \sigma_{\mathrm{int}}/\sigma_{\mathrm{nd}}$ should still hold. Here, as before, σ_{int} is the integrated interaction cross section for a given regularization prescription at small $p_\perp$, while the inelastic nondiffractive cross section σ_{nd} is taken from parametrizations [7, 10]. This relation can be used to solve for the proportionality factor k in eq. (3.3). Note that, since now each event has to have at least one interaction, $\langle n \rangle > 1$, one must ensure that $\sigma_{\mathrm{int}} > \sigma_{\mathrm{nd}}$. The $p_{\perp 0}$ parameter has to be chosen accordingly small — since now the concept of no-interaction low-$p_\perp$ events is gone, aesthetically it is more appealing to use the smooth $p_{\perp 0}$ turnoff than the sharp $p_{\perp\mathrm{min}}$ cutoff, and thereby populate interactions continuously all the way down to $p_\perp = 0$. The whole approach can be questioned at low energies, since then very small $p_{\perp 0}$ values would be required, so that many of the interactions would end up in the truly nonperturbative $p_\perp$ region.

Technically, the combined selection of b and a set of scattering $\mathrm{p}_{\perp i}$ values now becomes more complicated [1, 9]. It can be reduced to a combined choice of b and $p_{\perp 1}$, according to a generalization of eq. (2.8)

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}p_{\perp 1}\,\mathrm{d}^2 b} = \frac{\mathcal{O}(b)}{\langle \mathcal{O} \rangle} \frac{1}{\sigma_{\mathrm{nd}}} \frac{\mathrm{d}\sigma}{\mathrm{d}p_\perp} \exp\left[-\frac{\mathcal{O}(b)}{\langle \mathcal{O} \rangle} \int_{p_\perp}^{\sqrt{s}/2} \frac{1}{\sigma_{\mathrm{nd}}} \frac{\mathrm{d}\sigma}{\mathrm{d}p'_\perp} \mathrm{d}p'_\perp \right] . \tag{3.5}$$

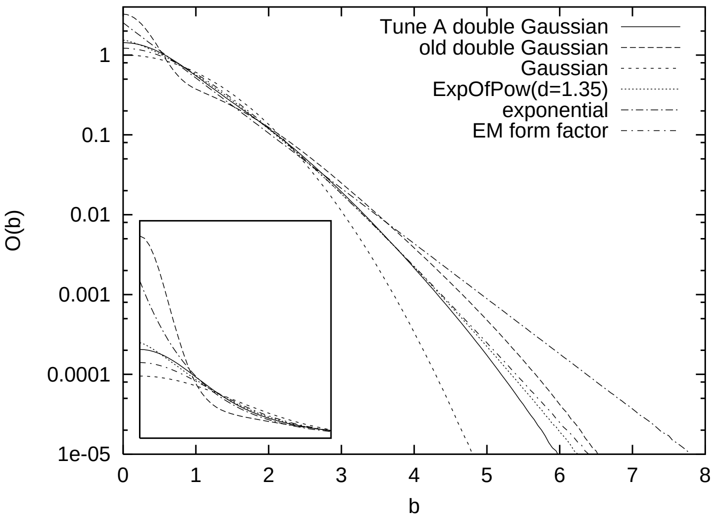


Figure 3: Overlap profile $\mathcal{O}(b)$ for a few different choices. Somewhat arbitrarily the different parametrizations have been normalized to the same area and average b, i.e. same $\int \mathcal{O}(b)\,\mathrm{d}^2 b$ and $\int b\mathcal{O}(b)\,\mathrm{d}^2 b$. (Recall that we have not specified b in terms of any absolute units, so both a vertical and a horizontal scale factor have to be fixed for each distribution separately.) Insert shows the region $b < 2$ on a linear scale.

The removal of the $\tilde{n} = 0$ non-events leads to a somewhat special definition of the average [1]

$$\langle \mathcal{O} \rangle = \frac{\int \mathcal{O}(b)\,\mathrm{d}^2 b}{\int \mathcal{P}_{\mathrm{int}}(b)\,\mathrm{d}^2 b} = \frac{1}{k}\frac{\sigma_{\mathrm{int}}}{\sigma_{\mathrm{nd}}}\,. \tag{3.6}$$

The subsequent interactions can be generated sequentially in falling $p_\perp$ as before, with the only difference that $\mathrm{d}\sigma/\mathrm{d}p_\perp^2$ now is multiplied by $\mathcal{O}(b)/\langle \mathcal{O} \rangle$, where b is fixed at the value chosen above.

Note that this lengthy procedure, via $\rho(r)$ and $\mathcal{O}(b)$, is not strictly necessary: the probability $\mathcal{P}_n$ for having n interactions could be chosen according to any desired distribution. However, with only $\mathcal{P}_n$ known and an n selected from this distribution, there is no obvious way to order the interactions in $p_\perp$ during the generation stage, with lower-$p_\perp$ interactions modified by the flavours, energies and momenta of higher-$p_\perp$ ones. (This problem is partly addressed in ref. [38], by a post-facto ordering of interactions and a subsequent rejection of some of the generated interactions, but flavour issues are not easily solved that way.)

There is also another issue, the parton-level pedestal effect, related to the transition from hard events to soft ones. To first approximation, the likelihood that an event contains a very hard interaction is proportional to $n\,\mathcal{P}_n$, since n interactions in an event means n chances for a hard one. If the requested hardest $p_\perp$ is gradually reduced, the bias towards large-N dies away and turns into its opposite: for events with the hardest $p_\perp \to 0$ the likelihood of further interactions vanishes. The interpolation between these two extremes can be covered if an impact parameter is chosen, and thereby an $\mathcal{O}(b)$, such that one can calculate the probability of *not* having an interaction harder than the requested hardest one, i.e. the exponential in eq. (3.5).

JHEP03(2004)053

If the gaussian matter distribution is the simplest possible choice, the double gaussian in some respects is the next-simplest one. It does introduce two new parameters, however, where we might have preferred to start with only one, to see how far that goes. As an alternative, we will here explore an exponential of a power of b

$$\mathcal{O}(b) \propto \exp\left\{-b^d\right\} \tag{3.7}$$

where $d \neq 2$ gives deviations from a gaussian behaviour. We will use the shorthand ExpOfPow($d = \ldots$) for such distributions. Note that we do not present an ansatz for $\rho(r)$ from which the $\mathcal{O}(b)$ is derived: in the general case the convolution of two ρ is nontrivial. A peaking of $\mathcal{O}(b)$ at $b = 0$ is related to one of $\rho(r)$ at $r = 0$, however.

A lower d corresponds to an overlap distribution more spiked at $b = 0$ and with a higher tail at large b, figure 3, i.e. leads to larger fluctuations. Specifically, the height of the $b = 0$ peak is related to the possibility of having fluctuations out to high multiplicities. To give some feeling, an exponential, ExpOfPow($d = 1$), is not too dissimilar to the old Pythia double gaussian, with $\beta = 0.5$ and $a_2/a_1 = 0.2$. Conveniently, the Tune A double gaussian, still with $\beta = 0.5$ but now $a_2/a_1 = 0.4$, is well approximated in shape by an ExpOfPow($d = 1.35$). Another alternative, commonly used, is to assume the matter distribution to coincide with the charge distribution, as gauged by the electric form factor $G_E(p_\perp^2) = (1 + p_\perp^2/\mu^2)^{-2}$, with $\mu^2 = 0.71\,\mathrm{GeV}^2$. This gives an $\mathcal{O}(b) \propto (\mu b)^3 K_3(\mu b)$, which also is close in form to Tune A, although somewhat less peaked at small b.

As indicated above, there are two key consequences of a an overlap profile choice. One is the interaction multiplicity distribution and the other the parton-level pedestal effect. These two are illustrated in figures 4 and 5, respectively, for $\mathrm{p\overline{p}}$ at 1.8 TeV, with $p_{\perp 0} = 2.0\,\mathrm{GeV}$ as in Tune A. The three frames of each figure illustrate how momentum conservation effects suppress the probability to have an event with large multiplicity. This effect is even stronger now that each interaction is allowed to undergo full shower evolution, so that it carries away more of the available energy. In the figures, the default lower shower cut-off of 1 GeV has been used; obviously a larger cut-off would give results intermediate to the two lower frames. Further, the possibility of two hard-scattering partons being part of the same shower is not included. Note that the suppression of the high-multiplicity tail implies that a distribution with large fluctuations in reality will have fewer interactions on the average than a less-fluctuating one, if they (as here) start with the same assumed average before the momentum conservation effects are considered. This means that the choice of $p_{\perp 0}$ is somewhat dependent on the one of overlap profile.

Let us now study the hadron-level multiplicity distribution, and begin with UA5 data at 200 and 900 GeV [48]. Tune A then does impressively well, figure 6, in spite of primarily having been tuned to pedestal effects rather than multiplicity distributions. In this comparison, we do not put too much emphasis on the low-multiplicity end, which is largely probing diffractive physics. Here the Pythia description is known to be too simple, with one or two strings stretched at low $p_\perp$ and no hard interactions at all. More relevant is the mismatch in peak position, which mainly is related to the multiplicity in events with only one interaction. Assuming that most hadronization parameters are fixed by $\mathrm{e^+e^-}$ data, it is not simple to tune this position. The beam remnant structure does offer some leeway, but

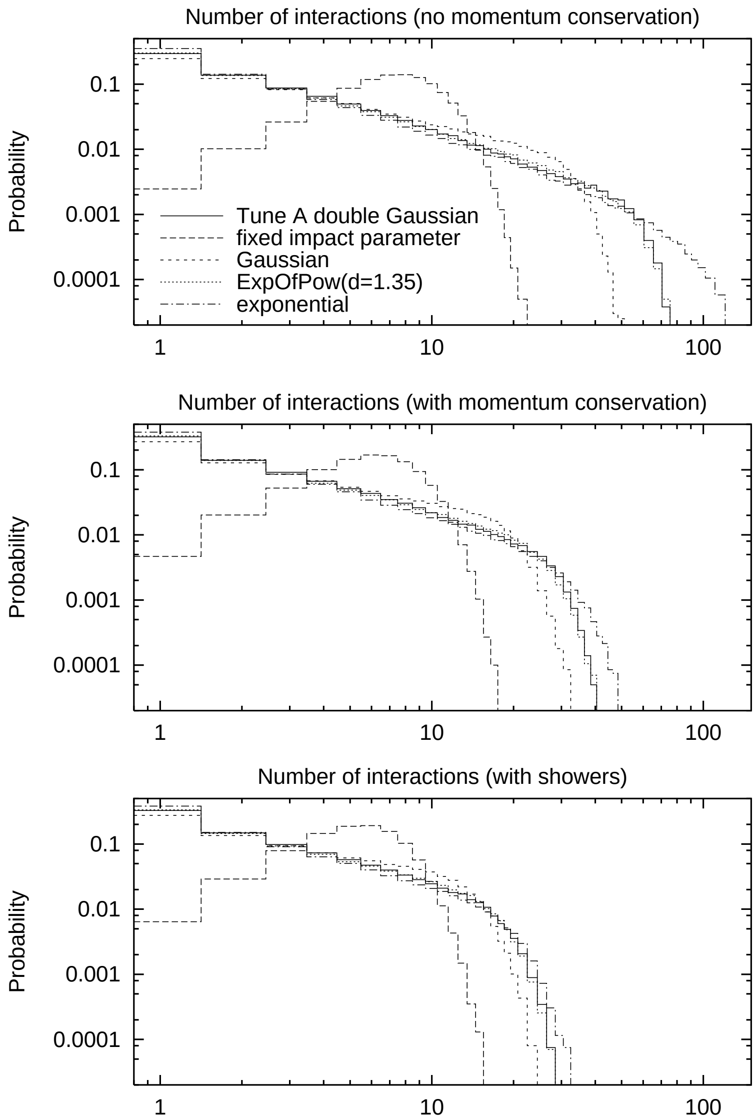


Figure 4: Distribution of the number of interactions for different overlap profiles $\mathcal{O}(b)$, for $\mathrm{p\overline{p}}$ at 1.8 TeV, top without momentum conservation constraints, middle with such constraints included but without (initial-state) showers, and bottom also with shower effects included.

actually the defaults are already set towards the end of the sensible range that produces the lower peak position, and still it comes out on the high side.

However, the main impression is of a very good description of the fluctuations to higher multiplicities, better than obtained with the old parameters explored in [1]. Of course, many aspects have changed significantly since then, such as the shape of parton densities at small x. One main difference is that Tune A 90% of the time picks subsequent interactions to be of the $\mathrm{gg} \to \mathrm{gg}$ type with colour flow chosen to minimize the string length. Since each further interaction thereby contributes less additional multiplicity, the mean number of interactions can be increased, and this obviates the need for the more extreme double-gaussian default parameters.

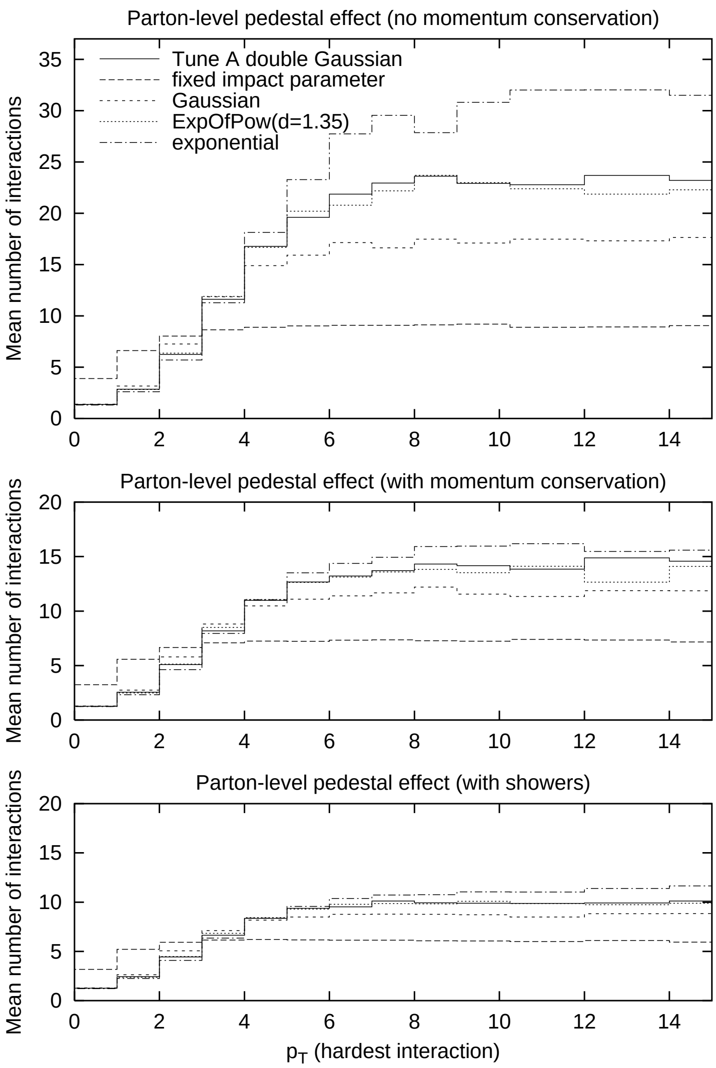


Figure 5: Average number of interactions as a function of the $p_\perp$ of the hardest interaction, for $\mathrm{p\overline{p}}$ at 1.8 TeV, top without momentum conservation constraints, middle with such constraints included but without (initial-state) showers, and bottom also with shower effects included.

If, nevertheless, one should attempt to modify the Tune A parameters, deviating from its ExpOfPow($d = 1.35$) near equivalent, it would be towards a smaller d, i.e. a slight enhancement of the tail towards high multiplicities. An example is shown in figure 6, with ExpOfPow($d = 1.2$) and $p_{\perp 0} = 1.9$ GeV (at 1800 GeV, with the Tune A energy rescaling).

However, the nice picture is shattered if one instead considers the E735 data at 1800 GeV [49], figure 7. Tune A gives a way too small tail out to large multiplicities, and also the ExpOfPow($d = 1.2$) falls below the data. One would need something like an exponential with a rather low $p_{\perp 0} = 1.6$ GeV to come near the E735 data, and that then disagrees with the lower-energy UA5 data, figure 6. The agreement could be improved, but not to the level of Tune A, by playing with the energy dependence of $p_{\perp 0}$. However, the E735 collaboration itself notes that results from the two collaborations are incompatible

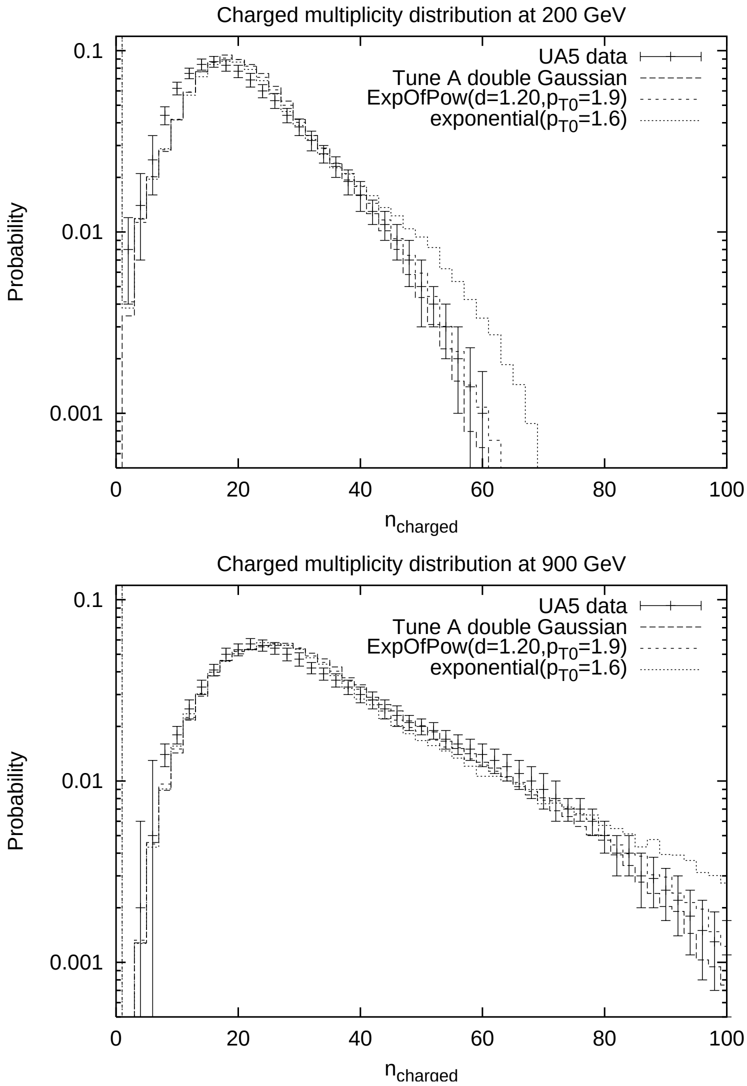


Figure 6: Charged multiplicity distribution at 200 and 900 GeV; different overlap profiles compared with UA5 data [48].

over the whole UA5 energy range and especially at 546 GeV, where both have data [49]. Furthermore, we do not have the expertise to fully simulate E735 selection criteria, nor to assess the impact of the large acceptance corrections. E735 only covered the pseudorapidity range $|\eta| < 3.25$, so about half of the multiplicity is obtained by extrapolation from the measured region for the 1800 GeV data. UA5 extended further and observed 70%-80%, depending on energy, of its multiplicity.

Obviously new experimental studies would be required to resolve the UA5–E735 ambiguity. As it stands, presumably a tune adjusted to fit E735 would give disagreement with the CDF data that went into Tune A. In this particular case we suspect the differences to be of an experimental origin, but in other cases it could well be that the Pythia model is incapable of fitting different (correct) distributions simultaneously, the model not being perfect. Indeed, speaking in general terms, that is a main reason why we try to improve the

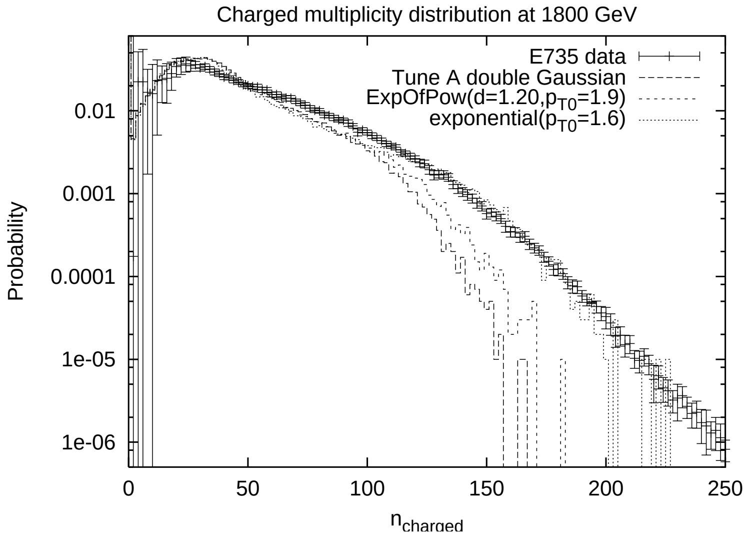


Figure 7: Charged multiplicity distribution at 1800 GeV; different overlap profiles compared with E735 data [49].

model in this article. In this particular case and for the moment being, however, we choose to use the UA5-compatible Tune A as a convenient reference for a realistic multiplicity distribution at Tevatron energies.

4. Correlations in momentum and flavour

Consider a hadron undergoing multiple interactions in a collision. Such an object should be described by multi-parton densities, giving the joint probability of simultaneously finding n partons with flavours $f_1, \ldots, f_n$, carrying momentum fractions $x_1, \ldots, x_n$ inside the hadron, when probed by interactions at scales $Q_1^2, \ldots, Q_n^2$. However, just like the standard one-particle-inclusive parton densities, such distributions would involve nonperturbative initial conditions that ultimately would have to be pinned down by experiment. We are nowhere near such a situation: the experimental information on double parton scattering, $n = 2$, boils down to one single number, the σ_{eff} of eq. (2.10), and for $n \geq 3$ there is no information whatsoever. Wishing to make maximal use of the existing ($n = 1$) information, we thus propose the following strategy.

As described above, the interactions may be generated in an ordered sequence of falling $p_\perp$. For the hardest interaction, all smaller $p_\perp$ scales may be effectively integrated out of the (unknown) fully correlated distributions, leaving an object described by the standard one-parton distributions, by definition. For the second and subsequent interactions, again all lower-$p_\perp$ scales can be integrated out, but the correlations with the first cannot, and so on.

The general situation is depicted in figure 8. This illustrates how, for the i'th interaction, only the correlations with the $i-1$ previous interactions need be taken into account, with all lower $p_\perp$ scales integrated out. Note, however, that this is only strictly true for the hard scatterings themselves. The initial-state shower evolution of, say, the first interaction, should exhibit correlations with the i'th at scales smaller than $p_{\perp i}$. Thus, the $p_\perp$ order-

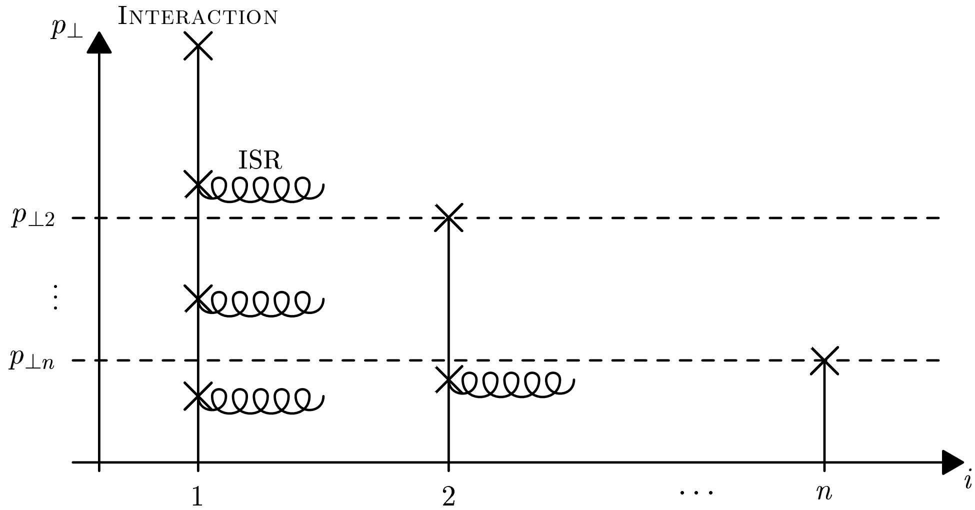


Figure 8: Schematic representation of the evolution of parton shower initiators in a hadron collision with n interactions (see text).

ing (or equivalently, a virtuality ordering) is in some sense equivalent to a time ordering, with the harder physics being able to influence the softer physics, but not vice versa. For two interactions of comparable $p_\perp$ this order may appear quite arbitrary, and also should not matter much, but consider the case of one very hard and one very soft interaction. The soft one will then correspond to a long formation time (field regeneration time) [66], $\sim p/p_\perp^2 \sim 1/p_\perp$, and indeed it is to be expected that the hard one can pre-empt or at least modify the soft one, whereas the influence in the other direction would be minor. This gives additional motivation to the choice of a $p_\perp$ ordering of interactions.

The possibility of intertwined shower evolution is not (yet) addressed. Rather, we introduce modified parton densities, that correlate the i'th interaction and its shower evolution to what happened in the $i-1$ previous ones, but we do not let the previous showers be affected by what happens in subsequent interactions. As partons are successively removed from the hadron by hard scatterings at smaller and smaller $p_\perp$ scales, the flavour, momentum and colour structure of the remaining object changes. The colour structure in particular is a thorny issue and will be discussed separately, in the next section. Here, we focus on deriving a set of parton distributions for a hadron after an arbitrary number of interactions have occurred, on the need for assigning a primordial transverse momentum to shower initiators, and on the kinematics of the partons residing in the final beam remnants.

Our general strategy is thus to pick a succession of hard interactions and to associate each interaction with initial- and final-state shower activity, using the parton densities introduced below. The initial-state shower is constructed by backwards evolution [67], back to the shower initiators at some low Q_0 scale, the parton shower cutoff scale. Thus, even if the hard scattering does not involve a valence quark, say, the possibility exists that the shower will reconstruct back to one. This necessitates dealing with quite complicated beam remnant structures. For instance, if two valence quarks have been knocked out of the same baryon in different directions, there will be three quarks, widely separated in momentum space, of which no two may naturally be collapsed to form a diquark system.

In the old model, technical limitations in the way the fragmentation was handled made it impossible to address such remnant systems. Consequently, it was not possible to associate initial-state radiation with the interactions after the first, i.e. the one with the highest $p_\perp$ scale, and only a very limited set of $\mathrm{q\overline{q}}$ and gg scatterings were allowed.

In a recent article [3], the Lund string model was augmented to include string systems carrying non-zero baryon number, by the introduction of 'junction fragmentation'. In the context of multiple interactions, this improvement means that almost arbitrarily complicated beam remnants may now be dealt with. Thus, a number of the restrictions that were present in the old model may now be lifted.

4.1 Parton densities

As mentioned above, we take the standard parton density functions as our starting point in constructing parton distributions for the remnant hadronic object after one or several interactions have occurred. Based on considerations of momentum and flavour conservation we then introduce successive modifications to these distributions.

The first and most trivial observation is that each interaction i removes a momentum fraction x_i from the hadron remnant. This is the fraction carried by the initiator of the initial-state shower, at the shower cutoff scale Q_0, so that the two initiators of an interaction together carry all the energy and momentum eventually carried by the hard scattering and initial-state shower products combined. To take into account that the total momentum of the remaining object is thereby reduced, already in the old model the parton densities were assumed to scale such that the point $x = 1$ would correspond to the remaining momentum in the beam hadron, rather than the total original beam momentum, cf. eq. (2.9). In addition to this simple x scaling ansatz we now introduce the possibility of genuine and non-trivial changes in both shape and normalization of the distributions.

4.1.1 Valence quarks

Whenever a valence quark is knocked out of an incoming hadron, the number of remaining valence quarks of that species should be reduced accordingly. Thus, for a proton, the valence d distribution is completely removed if the valence d quark has been kicked out, whereas the valence u distribution is halved when one of the two is kicked out. In cases where the valence and sea u and d quark distributions are not separately provided from the PDF libraries, we assume that the sea is flavour-antiflavour symmetric, so that one can write e.g.

$$u(x, Q^2) = u_{\mathrm{v}}(x, Q^2) + u_{\mathrm{s}}(x, Q^2) = u_{\mathrm{v}}(x, Q^2) + \overline{u}(x, Q^2)\,. \tag{4.1}$$

Here and in the following, q_{v} (q_{s}) denotes the q valence (sea) distribution. The parametrized u and $\overline{u}$ distributions should then be used to find the relative probability for a kicked-out u quark to be either valence or sea. Explicitly, the quark valence distribution of flavour f after n interactions, $q_{f\mathrm{v}n}(x, Q^2)$, is given in terms of the initial distribution, $q_{f\mathrm{v}0}(x, Q^2)$, and the ratio of remaining to original q_f valence quarks, $N_{f\mathrm{v}n}/N_{f\mathrm{v}0}$, as:

$$q_{f\mathrm{v}n}(x, Q^2) = \frac{N_{f\mathrm{v}n}}{N_{f\mathrm{v}0}} \frac{1}{X} q_{f\mathrm{v}0}\left(\frac{x}{X}, Q^2\right); \qquad X = 1 - \sum_{i=1}^{n} x_i\,, \tag{4.2}$$

where $N_{u\mathrm{v}0} = 2$ and $N_{d\mathrm{v}0} = 1$ for the proton, and $x \in [0, X]$ is the fraction of the original beam momentum ($\sum_{i=1}^{n} x_i$ is the total momentum fraction already taken out of the incoming hadrons by the preceding parton-shower initiators). The Q^2 dependence of $q_{f\mathrm{v}n}$ is inherited from the standard parton densities $q_{f\mathrm{v}0}$, and this dependence is reflected both in the choice of a hard scattering and in the backwards evolution. The factor $1/X$ arises since we squeeze the distribution in x while maintaining its area equal to the number of q_f valence quarks originally in the hadron, $N_{f\mathrm{v}0}$, thereby ensuring that the sum rule,

$$\int_0^X q_{f\mathrm{v}n}(x, Q^2)\ \mathrm{d}x = N_{f\mathrm{v}n}\,, \tag{4.3}$$

is respected. There is also the total momentum sum rule,

$$\int_0^X \left(\sum_f q_{fn}(x, Q^2) + g_n(x, Q^2) \right) x\,\mathrm{d}x = X\,. \tag{4.4}$$

Without any further change, this sum rule would not be respected since, by removing a valence quark from the parton distributions in the above manner, we also remove a total amount of momentum corresponding to $\langle x_{f\mathrm{v}} \rangle$, the average momentum fraction carried by a valence quark of flavour f:

$$\langle x_{f\mathrm{v}n}(Q^2) \rangle \equiv \frac{\int_0^X q_{f\mathrm{v}n}(x, Q^2)\ x\ \mathrm{d}x}{\int_0^X q_{f\mathrm{v}n}(x, Q^2)\ \mathrm{d}x} = X\,\langle x_{f\mathrm{v}0}(Q^2) \rangle\,. \tag{4.5}$$

The removal of $\sum_i x_i$, the total momentum carried by the previously struck partons, has already been taken into account by the 'squeezing' in x of the parton distributions (and expressed in eq. (4.4) by the r.h.s. being equal to X rather than 1). By scaling down the $\mathrm{q_v}$ distribution, we are removing an *additional* fraction, $\langle x_{f\mathrm{v}n} \rangle$, which must be put back somewhere, in order to maintain the validity of eq. (4.4).

Strictly speaking, $\langle x_{f\mathrm{v}0} \rangle$ of course depends on which specific PDF set is used. Nevertheless, for the purpose at hand this variation is negligible between most modern PDF sets. Hence we make the arbitrary choice of restricting our attention to the values obtained with the CTEQ5L PDF set [6].

More importantly, all the above parton densities depend on the factorization scale Q^2. This dependence of course carries over to $\langle x_{f\mathrm{v}0} \rangle$, for which we assume the functional form

$$\langle x_{f\mathrm{v}0}(Q^2) \rangle = \frac{A_f}{1 + B_f \log(\log(\max(Q^2, 1\ \mathrm{GeV}^2)/\Lambda^2_{\mathrm{QCD}}))}\,, \tag{4.6}$$

inspired by the $\mathrm{d}s = \mathrm{d}\log(\log Q^2/\Lambda^2) \propto \mathrm{d}Q^2/Q^2\,\alpha_{\mathrm{s}}(Q^2)$ pace of evolution, where $\mathrm{d}\langle x \rangle/\mathrm{d}s \approx -B\langle x \rangle$ suggests a solution of the form $\langle x \rangle \propto \exp(-Bs) \approx 1/(1+Bs)$. Reasonable fits to the CTEQ5L valence quark distributions in the proton are obtained for $A_{\mathrm{d}} = 0.385$, $B_{\mathrm{d}} = 1.60$, $A_{\mathrm{u}} = 0.48$ and $B_{\mathrm{u}} = 1.56$, with the isospin conjugate for neutrons.

JHEP03(2004)053

Essentially nothing is known about parton densities for other baryons, such as the reasonably long-lived hyperons Λ^0, $\Sigma^{+,-}$, $\Xi^{0,-}$ and Ω^-, which can undergo secondary interactions that one may wish to study. We here use essentially the same parton densities and parameters as for protons. Thus the influence of the larger strange quark mass is neglected, which ought to lead to harder x spectra for s quarks and softer for everything else. The fact that the two proton valence u quarks have a harder distribution than the single d one is carried over to other baryons with two equal quarks, while the average (sum) of the u and d distributions are used for baryons with three unequal (equal) quarks. For mesons and the Vector Meson Dominance part of photons one could use a similar strategy, with π^+ measurements as a starting point instead of protons, while the anomalous part of the photon densities is perturbatively calculable. There are still many further assumptions that would have to go into a complete model of multiple interactions in γp and $\gamma\gamma$ events [12], however, and so far we did not pursue this further.

We now know how much momentum is 'missing' in eq. (4.4). It is not possible to put this momentum back onto valence quarks without changing the shape of the distributions (beyond the mere 'x squeezing') or invalidating eq. (4.3). Rather, we here assume that the missing momentum is taken up by the sea+gluon distributions, which thus are scaled up slightly when a valence quark is kicked out. This enhancement of the sea+gluon momentum fraction may over- or undercompensate the 'x squeezing' reduction, depending on whether the kicked-out valence quark had a small or large x. However, before the procedure can be discussed in more detail, we must consider another effect which affects the normalization of the sea: changes in the content of the sea itself.

4.1.2 Sea quarks and their companions

When a sea quark is kicked out of a hadron, it must leave behind a corresponding antisea parton in the beam remnant, by flavour conservation. We call this a companion quark. In the perturbative approximation the sea quark $\mathrm{q_s}$ and its companion $\mathrm{q_c}$ come from a gluon branching $\mathrm{g} \to \mathrm{q_s} + \mathrm{q_c}$, where it is implicitly understood that if $\mathrm{q_s}$ is a quark, $\mathrm{q_c}$ is its antiquark, and vice versa. This branching often would not be in the perturbative regime, but we choose to make a perturbative ansatz, and also to neglect subsequent perturbative evolution of the q_c distribution. Even if approximate, this procedure should catch the key feature that a sea quark and its companion should not be expected too far apart in x (or, better, in $\ln x$).

With this approximation, we obtain the q_c distribution from the probability that a sea quark $\mathrm{q_s}$, carrying a momentum fraction x_s, is produced by the branching of a gluon with momentum fraction y, so that the companion has a momentum fraction $x = y - x_\mathrm{s}$,

$$\begin{aligned} q_\mathrm{c}(x; x_\mathrm{s}) &= C \int_0^1 g(y)\, P_{\mathrm{g}\to\mathrm{q_s q_c}}(z)\, \delta(x_\mathrm{s} - zy)\ \mathrm{d}z \\ &= C g(y)\, P_{\mathrm{g}\to\mathrm{q_s q_c}}\left(\frac{x_\mathrm{s}}{y}\right)\frac{1}{y} \\ &= C \frac{g(x_\mathrm{s}+x)}{x_\mathrm{s}+x} P_{\mathrm{g}\to\mathrm{q_s q_c}}\left(\frac{x_\mathrm{s}}{x_\mathrm{s}+x}\right), \end{aligned} \tag{4.7}$$

with C a normalization constant to be determined below, and $P_{\mathrm{g}\to\mathrm{q_s q_c}}$ the DGLAP splitting kernel

$$P_{\mathrm{g}\to\mathrm{q_s q_c}}(z) = \frac{1}{2}\left(z^2 + (1-z)^2\right). \tag{4.8}$$

In view of the approximate nature of the procedure, allowing a generic $g(x)$ shape would give disproportionately complex expressions. Instead, the following simple ansatz for the gluon distribution at low Q^2 is used:

$$g(x) \propto \frac{(1-x)^p}{x}, \tag{4.9}$$

with the integer choices $p = 0, 1, 2, 3, 4$ giving a range of variability for the large-x behaviour of the distribution and the $1/x$ controlling the small-x behaviour. Note that all the above equations are defined assuming no previous energy loss and, as for the valence quarks, should be 'squeezed' by a factor X, to ensure momentum conservation.

The overall normalization of a companion quark distribution is obtained by imposing the sum rule:

$$\int_0^{1-x_\mathrm{s}} q_\mathrm{c}(x; x_\mathrm{s})\mathrm{d}x = 1. \tag{4.10}$$

Inserting eqs. (4.7)–(4.9) and inverting, one obtains the normalization constants C_p:

$$\begin{aligned}
C_0 &= \frac{3x_\mathrm{s}}{2 - x_\mathrm{s}(3 - 3x_\mathrm{s} + 2x_\mathrm{s})}, &(4.11)\\
C_1 &= \frac{3x_\mathrm{s}}{2 - x_\mathrm{s}^2(3 - x_\mathrm{s}) + 3x_\mathrm{s}\log(x_\mathrm{s})}, &(4.12)\\
C_2 &= \frac{3x_\mathrm{s}}{2(1 - x_\mathrm{s})(1 + 4x_\mathrm{s} + 4x_\mathrm{s}^2) + 6x_\mathrm{s}(1 + x_\mathrm{s})\log(x_\mathrm{s})}, &(4.13)\\
C_3 &= \frac{6x_\mathrm{s}}{4 + x_\mathrm{s}(27 - 31x_\mathrm{s}^2) + 6x_\mathrm{s}(3 + 2x_\mathrm{s}(3 + x_\mathrm{s}))\log(x_\mathrm{s})}, &(4.14)\\
C_4 &= \frac{3x_\mathrm{s}}{2(1 + 2x_\mathrm{s})((1 - x_\mathrm{s})(1 + x_\mathrm{s}(10 + x_\mathrm{s})) + 6x_\mathrm{s}(1 + x_\mathrm{s})\log(x_\mathrm{s}))}. &(4.15)
\end{aligned}$$

To illustrate, figure 9 shows properly normalized companion momentum distributions for $p = 4$ and $p = 0$, each for two different values of x_s. There are two noteworthy aspects about these distributions. Firstly, for $p = 0$ the discontinuity at the point $x = 1 - x_\mathrm{s}$ is merely an artifact of our parametrization of the gluon density, i.e. that $g(x)$ does not vanish at $x = 1$. That problem is absent for more realistic p values; in our continued discussions we will use $p = 4$ as default, this being the closest to the CTEQ5L small-Q^2 gluon distribution. Secondly, the falling gluon distribution convoluted with the almost flat $\mathrm{g} \to \mathrm{q\overline{q}}$ splitting kernel give distributions that roughly tend to a constant $q_\mathrm{c}(x; x_s) \sim C_p/2x_\mathrm{s}^2$ below x_s and exhibit power-like fall-offs $q_\mathrm{c}(x; x_s) \propto 1/x^2$ above it, with some modulation of the latter depending on p. In order to display the probability per $\log x$ interval, figure 9 gives $x\,q_\mathrm{c}$ rather than q_c itself, and then a peaking occurs around $x \approx x_\mathrm{s}$, as should be expected from the symmetric splitting kernel in eq. (4.8).

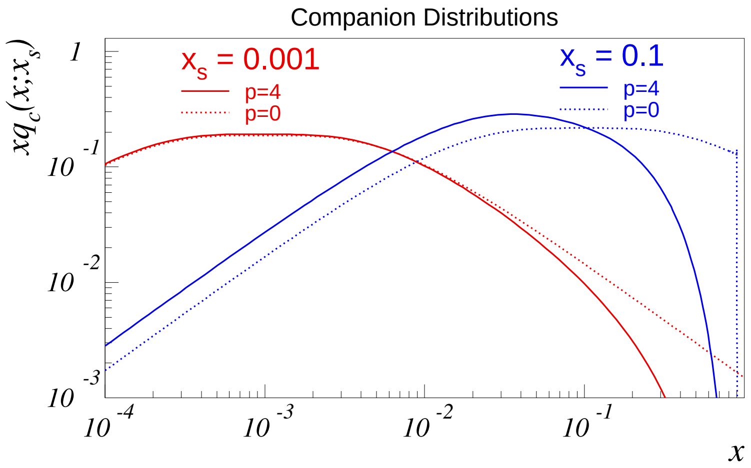


Figure 9: Companion distributions for $p = 4$ (solid lines) and $p = 0$ (dotted lines), for two different values of x_s.

Also here, the question arises of ensuring that the total momentum sum rule, eq. (4.4), is respected, but now the difference has the opposite sign; by adding a companion quark distribution, we are in effect bookkeeping a part of the flavour and momentum content of the sea separately. One possibility is that this momentum comes *only* from the sea+gluons and that the valence quarks are not affected, i.e. that the rest of the sea+gluons fluctuate down, in order to compensate.

The amount of momentum that will have to be compensated for each companion quark, $\langle x_{cn} \rangle = X \langle x_{c0} \rangle$, with $\langle x_{cn} \rangle$ defined in analogy with eq. (4.5), is straightforward to compute using the distribution in eq. (4.7) and the normalizations given by eqs. (4.11)–(4.15):

$$\langle x_{c0} \rangle_{p=0} = x_s \frac{5 - 9x_s + 6x_s^2 - 2x_s^3 + 3\log x_s}{(x_s - 1)(2 - x_s + 2x_s^2)} , \tag{4.16}$$

$$\langle x_{c0} \rangle_{p=1} = -1 - 3x_s + \frac{2(1 - x_s)^2(1 + x_s + x_s^2)}{2 - 3x_s^2 + x_s^3 + 3x_s \log x_s} , \tag{4.17}$$

$$\langle x_{c0} \rangle_{p=2} = \frac{x_s}{4} \frac{19 + 24x_s - 39x_s^2 - 4x_s^3 + 6(1 + 6x_s + 4x_s^2)\log x_s}{-1 - 3x_s + 3x_s^2 + x_s^3 - 3x_s(1 + x_s)\log x_s} , \tag{4.18}$$

$$\langle x_{c0} \rangle_{p=3} = 3x_s \frac{-7 - 21x_s + 15x_s^2 + 13x_s^3 - 2(1 + 9x_s + 12x_s^2 + 2x_s^3)\log x_s}{4 + 27x_s - 31x_s^3 + 6x_s(3 + 6x_s + 2x_s^2)\log x_s} , \tag{4.19}$$

$$\langle x_{c0} \rangle_{p=4} = 3x_s \frac{3(5 + 24x_s - 4x_s^2 - 24x_s^3 - x^4) + 4(1 + 12x_s + 24x_s^2 + 8x_s^3)\log x_s}{8(1 + 2x_s)(-1 - 9x_s + 9x_s^2 + x_s^3 - 6x_s(1 + x_s)\log x_s)} . \tag{4.20}$$

4.1.3 Sea quark and gluon density normalizations

As described above, the normalization of valence and companion distributions is fixed by the respective number of quarks, i.e. the sum rules (for each flavour, f)

$$\int_0^X q_{f\mathrm{v}n}(x, Q^2)\, \mathrm{d}x = N_{f\mathrm{v}n} , \tag{4.21}$$

$$\int_0^X q_{fc_jn}(x; x_{s_j})\, \mathrm{d}x = 1 \qquad \text{(for each } j) , \tag{4.22}$$

where X is still the longitudinal momentum fraction left after the n previous interactions and $N_{f\mathrm{v}n}$ is the number of q_f valence quarks remaining. The index j on the companion distribution, $q_{f\mathrm{c}_j n}$, counts different companion quarks of the same flavour, f.

On the other hand, the sea+gluon distributions do not have fixed multiplicities, hence no corresponding sum rules exist for their normalizations. We use this freedom to fulfill the last remaining sum rule, eq. (4.4), letting the sea+gluon normalizations fluctuate up when we reduce a valence distribution and down when we add a companion distribution. In addition, the requirement of a physical x range is of course maintained by still 'squeezing' all distributions into the interval $x \in [0, X]$.

For simplicity, and since eq. (4.4) only furnishes us with one equation of constraint, we assume the same scale factor for all sea flavours as well as for the gluon, i.e.

$$q_{f\mathrm{s}}(x,Q^2) \to a q_{f\mathrm{s}}(x,Q^2)\,, \tag{4.23}$$

$$g(x,Q^2) \to a g(x,Q^2)\,. \tag{4.24}$$

The momentum sum rule now reads:

$$\begin{aligned}
1 &= \frac{1}{X}\int_0^X \left(\sum_f \left[q_{f\mathrm{v}n}(x,Q^2) + \sum_j q_{f\mathrm{c}_j n}(x;x_j) + a q_{f\mathrm{s}}(x,Q^2)\right] + a g_n(x,Q^2)\right) x \,\mathrm{d}x \\
&= \int_0^1 \left(\sum_f \left[\frac{N_{f\mathrm{v}n}}{N_{f\mathrm{v}0}} q_{f\mathrm{v}0}(x,Q^2) + \sum_j q_{f\mathrm{c}_j 0}(x;x_j) + a q_{f\mathrm{s}0}(x,Q^2)\right] + a g_0(x,Q^2)\right) x \,\mathrm{d}x \\
&= a + \sum_f \int_0^1 \left[\left(\frac{N_{f\mathrm{v}n}}{N_{f\mathrm{v}0}} - a\right) q_{f\mathrm{v}0}(x,Q^2) + \sum_j q_{f\mathrm{c}_j 0}(x;x_j)\right] x \,\mathrm{d}x \\
&= a\left(1 - \sum_f N_{f\mathrm{v}0}\langle x_{f\mathrm{v}0}\rangle\right) + \sum_f N_{f\mathrm{v}n}\langle x_{f\mathrm{v}0}\rangle + \sum_{f,j}\langle x_{f\mathrm{c}_j 0}\rangle\,,
\end{aligned} \tag{4.25}$$

and hence,

$$a = \frac{1 - \sum_f N_{f\mathrm{v}n}\langle x_{f\mathrm{v}0}\rangle - \sum_{f,j}\langle x_{f\mathrm{c}_j 0}\rangle}{1 - \sum_f N_{f\mathrm{v}0}\langle x_{f\mathrm{v}0}\rangle}\,. \tag{4.26}$$

One easily checks that $a = 1$ before the first interaction, as it should be, and that a is driven larger by $N_{f\mathrm{v}n} < N_{f\mathrm{v}0}$, while introducing companion quarks drives it the opposite way, also as expected.

4.2 Beam remnants

The longitudinal momenta and flavours of the initiator partons are defined by the sequence of $p_\perp$-ordered hard scatterings and their associated initial-state showers, as described above. What is left in the beam remnant is then a number of partons, with flavours given by the remaining valence content plus the number of sea quarks required for overall flavour conservation. That is, gluons in the remnant are not explicitly accounted for, but are implicit as confinement clouds around the quarks and as unresolved originators of sea quark pairs.

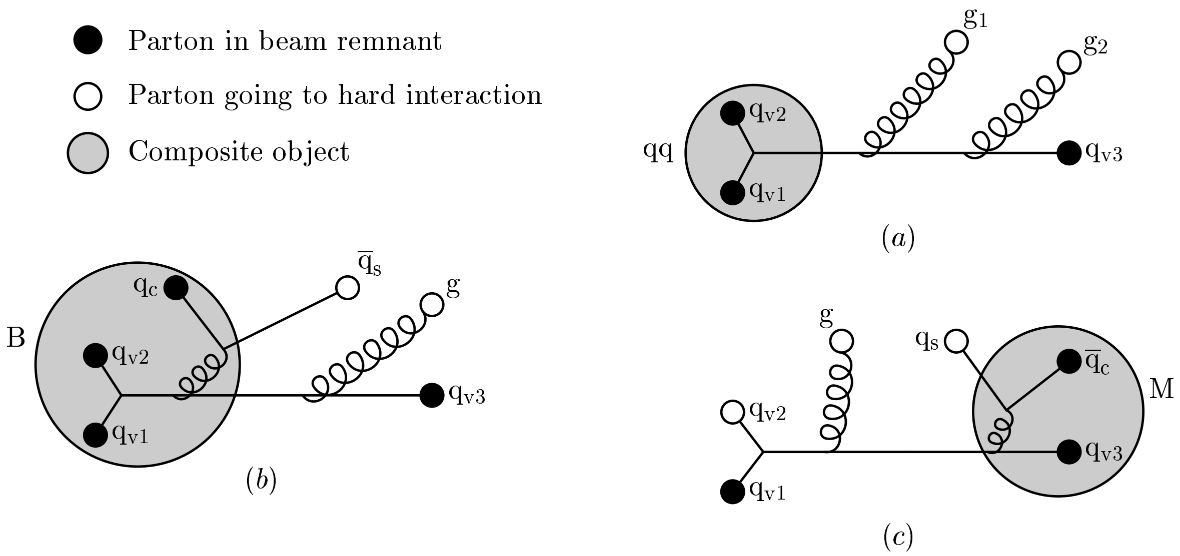


Figure 10: Examples of the formation of composite objects in a baryon beam remnant: (a) diquark, (b) baryon and (c) meson.

A remnant may thus contain several objects but, when the colour configuration is studied, simplifications can occur. A colour antitriplet qq pair in the remnant can be associated with a diquark, a colour singlet qqq triplet with a baryon, and a colour singlet $q\overline{q}$ pair with a meson, see figure 10. When hadrons are formed, the standard string fragmentation relative probabilities are used to select spin and other quantum numbers, i.e. whether π or ρ, etc.

It is here assumed that the respective pair/triplet has a sufficiently small invariant mass that it can reasonably be projected onto a single composite state. Thus a $q\overline{q}$ system with large invariant mass would define a string that could fragment into several mesons, rather than collapse to a single meson. In principle this could be modeled dynamically, but it would require the introduction of some nonperturbative parameters, to describe the partitioning of the proton into arbitrary-mass subsystems. At this point, we consider it meaningful only to study a few specific scenarios for which partons to allow in the formation of composite objects. We have chosen four such:

1. No composite objects are formed *ab initio*. All partons act as single units, either as endpoints (quarks) or kinks (gluons) on strings that fragment in the normal way.

2. Composite objects may be formed, but only when all partons involved in the formation are valence quarks.

3. The formation of diquarks may involve both valence and sea quarks, but the formation of colour singlet subsystems (i.e. hadrons) is still restricted to involve valence quarks only.

4. Sea quarks may also be used for colour singlet formation.

The idea is thus that (spectator) valence quarks tend to have comparable velocities, while sea quarks can be more spread out and therefore are less likely to form low-mass systems.

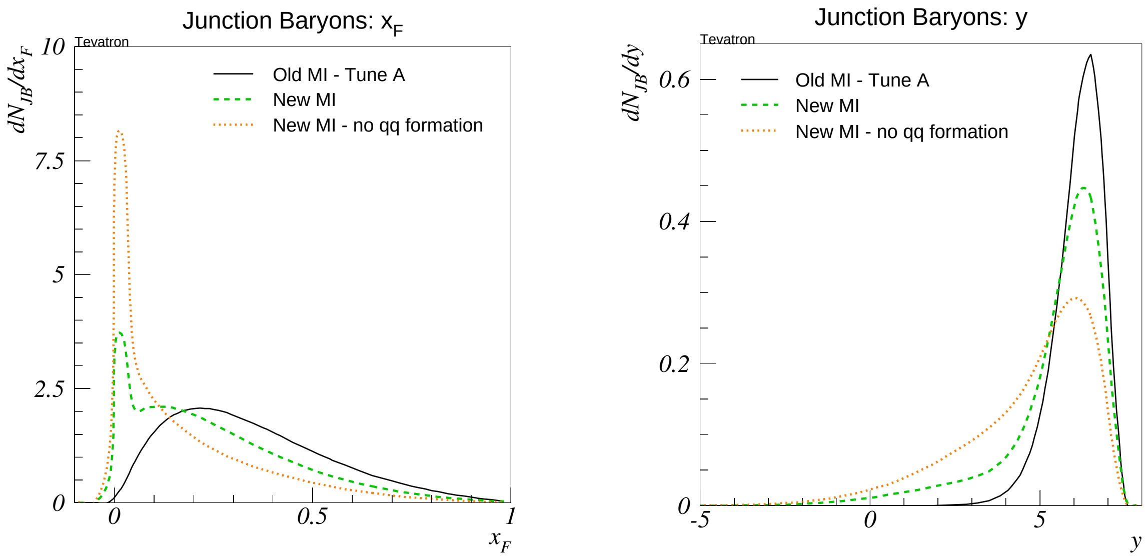


Figure 11: Feynman x (left) and rapidity (right) distributions for junction baryons: distributions are shown for Tune A of the old MI scenario (solid lines), and for the new model with diquark formation in the beam remnant switched on (dashed) and off (dotted).

Whether composite systems in the beam remnant are formed or not has important consequences for the baryon number flow. For $\mathrm{p\bar{p}}$ collisions at 1.8 TeV CM energy, we show in figure 11 the Feynman x (left plot) and rapidity (right plot) distributions for the baryon which 'inherits' the beam baryon number. We denote this baryon the 'junction baryon'. To better illustrate what happens to each of the two initial beam baryon numbers separately, only distributions for the junction baryon, not *anti*baryon, are shown. Possibilities 1 and 2 above are compared with the old multiple interactions model (Tune A). One immediately observes that the beam baryon number migrates in a radically different way when diquark formation is allowed or not (compare the dashed and dotted sets of curves). In fact, in the new model it is not possible to reproduce the old distribution (compare the solid curve). This comes about since, even when all possible diquark formation is allowed in the new model, it is not certain that the beam remnant actually contains the necessary quark content, hence in some fraction of the events the formation of a beam remnant diquark is simply not possible. Here is thus an example where the introduction of more physics into the model has given rise to a qualitatively different expectation: the beam baryon number appears to be stopped to a larger extent than would previously have been expected.

One should note that, also at later stages, a small-mass string piece can collapse to a single hadron, as part of the normal string fragmentation procedure. There, however, it is intended only to cover a rare low-mass tail of systems mainly defined by hard processes and perturbative shower evolution, while the simplifications considered for the above kinematical configurations are quite common and in a nonperturbative context. Beam remnant quarks that were not collapsed in the nonperturbative first stage could in the later stage be collapsed with other partons, e.g. from the showers. Also collapses not allowed in the more restrictive scenarios above could occur at this stage. Whether that happens or not depends on the transverse and longitudinal momenta that will be defined below.

JHEP03(2004)053

4.3 Primordial $k_\perp$

Until now, we have considered only the longitudinal part of parton momenta. In reality, partons are also expected to have some non-zero $k_\perp$ values caused by Fermi motion inside the incoming hadrons. This kind of $k_\perp$ is denoted 'primordial $k_\perp$', since it is not generated by the (DGLAP) shower evolution nor from hard interactions, but rather represents an input to the perturbative stages of the event. Based on Fermi motion alone, one would expect values of the order of a few hundred MeV, just like in eq. (2.6). But to reproduce e.g. the $p_\perp$ distributions of Z bosons produced in hadron-hadron collisions, one notes a need for a significantly larger nonperturbative input, either in parton showers or in resummation descriptions [68]. This problem is still awaiting a satisfactory explanation, although some ideas have been explored that might alleviate it [69]. Until such an explanation has been found, we therefore have reason to consider an effective 'primordial $k_\perp$', at the level of the initiators, larger than the one above. For simplicity, a parametrized Q-dependent width

$$\sigma(Q) = \max\left(\sigma_{\mathrm{min}}, \sigma_\infty \frac{1}{1 + Q_{\frac{1}{2}}/Q}\right) \tag{4.27}$$

is introduced, where σ is the width of the two-dimensional gaussian distribution of the initiator primordial $k_\perp$ (so that $\langle k_\perp^2 \rangle = \sigma^2$), Q is the scale of the hard interaction, σ_∞ is the value asymptotically approached as $Q \to \infty$, and $Q_{\frac{1}{2}}$ is the Q scale where $\sigma = \frac{1}{2}\sigma_\infty$. A reasonable fit to the few available experimental 'data points',

$$\begin{aligned}
\sigma(Q \sim 1\,\mathrm{GeV}) \approx \sigma_{\mathrm{min}} &\approx 0.36\,\mathrm{GeV} &\quad& \text{(fragmentation)} \\
\sigma(Q \sim 5-10\,\mathrm{GeV}) &\approx 0.9\,\mathrm{GeV} && \text{(EMC) [70]} \\
\sigma(Q \sim M_Z) &\approx 2\,\mathrm{GeV} && \text{(Tevatron) [68]}\,,
\end{aligned} \tag{4.28}$$

is obtained with the values $\sigma_\infty = 2.1\,\mathrm{GeV}$ and $Q_{\frac{1}{2}} = 7\,\mathrm{GeV}$. The σ_{min} in eq. (4.27) represents a minimum broadening, at the level of Fermi motion, which we take to be the standard fragmentation $p_\perp$ width. In addition to partons participating in relatively soft interactions, this minimum broadening is also applicable to the remnant partons, which by definition do not participate in hard interactions and hence are not naturally associated with a particular Q scale.

Apart from the selection of each individual $k_\perp$, there is also the requirement that the total $k_\perp$ of the beam adds up to zero. The question of how partons recoil off one another in transverse momentum space inside hadrons is so far largely unaddressed in the literature. We imagine a few different possibilities here:

1. The primordial transverse momenta are generated at a stage where the partons have low virtualities and hence large wavefunctions. Moreover, at least for that part which is due to Fermi motion, the dynamics responsible for the generation of primordial $k_\perp$ is that of a Fermi gas of partons in equilibrium, where each parton has received its total primordial $k_\perp$ through a sequence of many collisions with many different partons. Therefore, one possibility is to let the recoil of one parton be shared uniformly among all other initiator and remnant partons.

2. Since Fermi motion alone appears unable to account for the bulk of primordial $k_\perp$ in large-Q interactions, there may exist a mechanism, involving presently unknown dynamics, which ties the generation of this $k_\perp$ to the presence of a large virtuality in the interaction, for instance by unresolved/unresummed bremsstrahlung radiation off the initiator parton and/or by the $k_\perp$ compensations in the shower being of a not strictly local nature (as happens e.g. in the dipole description of parton showers [36]). Regardless of the exact nature of the mechanism, recoils should in this picture primarily be taken up by initiators and beam remnant partons which are close in colour space.

   (a) The extreme variant is here to let the recoil of a particular parton be taken up by its nearest colour neighbours only.

   (b) A perhaps more realistic possibility is to let the compensation happen along a parton chain in colour space, with successive dampening of the compensation along the chain.

These three possibilities are included in the present study. However, the possibilities involving colour chains are complicated by the fact that the colour connections between initiator and remnant partons are very poorly known, since no perturbative information is available. These problems will be discussed in more detail in section 5 below.

Irrespective of which particular method is used to ensure $\sum_i \vec{k}_{\perp i} = 0$, the question now arises how the kinematics of the initially collinear partons should be reinterpreted to include non-zero $k_\perp$ assignments. This may be done either by associating the generation of $k_\perp$ with the building up of space-like virtualities among the partons, or by keeping the partons massless while allowing (non-perturbatively small) longitudinal momentum transfers between the beam remnants. (In the latter approximation also heavy quarks are kept massless rather than assigned a spacelike virtuality; when initial-state showers are included no heavy quarks need be assigned to the beam remnants, however.) In the first case, the invariant mass of initiator and beam remnant partons combined in each hadron is maintained equal to the original hadron mass, while in the second the mass can be significantly larger. The difference, however, should be considered mostly technical, since the momentum transfers involved are quantitatively small. For this study, the second option is chosen, since this avoids potential technical problems in dealing with string systems having negative mass squares.

For a specific interaction, consider a pair of massless initiator partons in their rest frame, before $k_\perp$ is added:

$$p_{1,2} = \frac{\sqrt{\hat{s}}}{2}(1,0,0,\pm 1)\,; \qquad \hat{s} = x_1 x_2 s\,. \tag{4.29}$$

With primordial $k_\perp$ included, these momentum vectors should now be recast as

$$p_{1,2} = \left(\sqrt{p_z^2 + p_{\perp 1,2}^2},\, \vec{p}_{\perp 1,2},\, \pm p_z\right), \tag{4.30}$$

if the system should still be at longitudinal rest. Since we are merely reinterpreting the kinematics of the initial-state partons, the centre-of-mass energy, $\sqrt{\hat{s}}$, of the interaction should be left unchanged. To ensure this, it is simple to solve

$$\hat{s} = \left(\sqrt{p_z^2 + p_{\perp 1}^2} + \sqrt{p_z^2 + p_{\perp 1}^2} \right)^2 - (\vec{p}_{\perp 1} + \vec{p}_{\perp 2})^2 \,, \tag{4.31}$$

to obtain the required p_z as a function of $\vec{p}_{\perp 1}$, $\vec{p}_{\perp 2}$, and $\hat{s}$:

$$p_z^2 = \frac{\lambda(\hat{s}_\perp, p_{\perp 1}^2, p_{\perp 2}^2)}{4\hat{s}_\perp} \,; \qquad \hat{s}_\perp \equiv \hat{s} + (\vec{p}_{\perp 1} + \vec{p}_{\perp 2})^2 \,, \tag{4.32}$$

with λ the standard Källén function,

$$\lambda(a, b, c) = a^2 + b^2 + c^2 - 2ab - 2bc - 2ac \,. \tag{4.33}$$

Naturally, only $k_\perp$ assignments which result in $p_z^2 > 0$ are acceptable.

4.4 Beam remnant longitudinal momenta

In addition to flavours and transverse momenta, the beam remnants must also together carry the remaining fraction, approximately X (as defined by eq. (4.2)) of longitudinal momentum. The sharing is based on the character of the remnant constituents. First a fraction x is defined for each constituent, and then these x fractions are rescaled for overall energy and momentum conservation.

Thus a valence quark receives an x picked at random according to a small-Q^2 valence-like parton density, proportional to $(1-x)^a/\sqrt{x}$, where $a = 2$ for a u quark in a proton and $a = 3.5$ for a d quark. A sea quark must be the companion of one of the initiator quarks, and can have an x picked according to the $q_{\mathrm{c}}(x; x_{\mathrm{s}})$ distribution introduced above. In the rare case that no valence quarks remain and no sea quarks need be added for flavour conservation, the beam remnant is represented by a gluon, carrying all of the beam remnant longitudinal momentum.

Among composite objects, a diquark would naïvely obtain an x given by the sum of its constituent quarks, while baryons and mesons would receive an x equated with the z value obtainable from a fragmentation function, in this study the Lund symmetric fragmentation function. However, earlier studies on quark-diquark remnants [1] have shown that, within the multiple interactions formalism, it is very difficult to accommodate observed remnant multiplicity distributions if the composite system (the diquark) does not take a much larger fraction than implied by the naïve estimate above. Physically, this could correspond to the momentum carried by a surrounding pion/gluon cloud being larger for a composite object than for a single parton. The possibility of enhancing the x values picked for composite objects is therefore retained in the present study.

Finally, once x values (and primordial $k_\perp$) have been picked for each of the remnants, an overall rescaling is performed such that the remnants together carry the desired longitudinal momentum. The simplest way to accomplish this would be to fix the normalization of the beam remnant x values on each side separately, by requiring conservation of longitudinal momentum, $\sum_i x_{i_{\rm BR}} = X$ on each side. Unfortunately, the introduction of non-zero $k_\perp$ values with massless partons and on-shell hadrons rules out such a simple approach, since energy would then not be conserved. Instead, small non-zero lightcone momentum fractions in the direction opposite to the parent hadron direction must be allowed. As already noted in section 4.3, this procedure should be thought of merely as a technical trick, necessitated by insisting on a description in terms of on-shell partons.

The amount of light-cone momentum removed from the remnant system by each pair of initiators, i, is in the overall cm frame of the event (omitting subscript i to avoid cluttering the notation),

$$\begin{aligned} w^+ &= E^{\rm cm} + p_z^{\rm cm} = \gamma(1+\beta_z)(E'_1 + E'_2)\,, \\ w^- &= E^{\rm cm} - p_z^{\rm cm} = \gamma(1-\beta_z)(E'_1 + E'_2)\,, \end{aligned} \tag{4.34}$$

where subscripts 1 and 2 denote the initiator partons of the hard scattering on each side respectively, and the primed frame is chosen as the longitudinal rest frame, defined by eq. (4.30). Then the boost is only along the z direction,

$$\beta_z = \beta = \frac{x_1 - x_2}{x_1 + x_2}\,, \tag{4.35}$$

and from equations (4.30)–(4.32)

$$E'_1 + E'_2 = \sqrt{\hat{s}_\perp}\,. \tag{4.36}$$

Inserting these results in eq. (4.34) above yields

$$\begin{aligned} w^+ &= \sqrt{\frac{1+\beta}{1-\beta}}\sqrt{\hat{s}_\perp} = \sqrt{\frac{x_1}{x_2}}\sqrt{\hat{s}_\perp}\,, \\ w^- &= \sqrt{\frac{1-\beta}{1+\beta}}\sqrt{\hat{s}_\perp} = \sqrt{\frac{x_2}{x_1}}\sqrt{\hat{s}_\perp}\,. \end{aligned} \tag{4.37}$$

For vanishing $p_\perp$ this simplifies to the familiar $\hat{s}_\perp = \hat{s} = x_1 x_2 s$, $w^+ = x_1\sqrt{s}$ and $w^- = x_2\sqrt{s}$.

The light-cone momenta remaining for the combined beam remnant system are thus:

$$W^+_{\rm rem} = \sqrt{s} - \sum_i w_i^+ = \sqrt{s} - \sum_i \sqrt{\frac{x_{i1}}{x_{i2}}}\sqrt{\hat{s}_{\perp i}} \tag{4.38}$$

$$W^-_{\rm rem} = \sqrt{s} - \sum_i w_i^- = \sqrt{s} - \sum_i \sqrt{\frac{x_{i2}}{x_{i1}}}\sqrt{\hat{s}_{\perp i}} \tag{4.39}$$

$$W^2_{\rm rem} = W^+_{\rm rem} W^-_{\rm rem}\,. \tag{4.40}$$

In extreme cases, it may happen that the hard interactions have removed so much energy and momentum from the beam remnants that the remnant system nominally becomes space-like, $W^2_{\rm rem} < 0$, if large $k_\perp$ values have been assigned. Though not strictly speaking unphysical, such a situation could lead to problems at the fragmentation stage. The requirement $W_{\rm rem}^2 > 0$ is therefore imposed as an additional constraint when primordial $k_\perp$ values are assigned.

The x values picked for the beam remnant partons are now interpreted as fractions of the light-cone momenta, $W_{\rm rem}^+$ and $W_{\rm rem}^-$, of the beam remnant system, modulo an overall rescaling on each side, to leave room for overall momentum conservation. Using index j to refer to beam remnant partons on side 1 and index k for the ones on side 2, we thus make the identification

$$\begin{aligned} p_j^+ &= \alpha x_j W_{\rm rem}^+ \implies p_j^- = \frac{m_{\perp j}^2}{p_j^+} \\ p_k^- &= \beta x_k W_{\rm rem}^- \implies p_k^+ = \frac{m_{\perp k}^2}{p_k^-} \,, \end{aligned} \tag{4.41}$$

where $m_\perp^2 = m^2 + p_\perp^2$ and α and β are global normalizations, to be determined from overall energy and momentum conservation in the beam remnant system:

$$W_{\rm rem}^+ = \sum_j p_j^+ + \sum_k p_k^+ = \alpha W_{\rm rem}^+ \sum_j x_j + \frac{1}{\beta W_{\rm rem}^-} \sum_k \frac{m_{\perp k}^2}{x_k} \tag{4.42}$$

$$W_{\rm rem}^- = \sum_j p_j^- + \sum_k p_k^- = \beta W_{\rm rem}^- \sum_k x_k + \frac{1}{\alpha W_{\rm rem}^+} \sum_j \frac{m_{\perp j}^2}{x_j} \,. \tag{4.43}$$

Equating these expressions with eqs. (4.38) and (4.39), one obtains for α and β

$$\alpha = \frac{W_{\rm rem}^2 + W_1^2 - W_2^2 + \lambda^{1/2}(W_{\rm rem}^2, W_1^2, W_2^2)}{2W_{\rm rem}^2 \sum_j x_j} \,, \tag{4.44}$$

$$\beta = \frac{W_{\rm rem}^2 + W_2^2 - W_1^2 + \lambda^{1/2}(W_{\rm rem}^2, W_1^2, W_2^2)}{2W_{\rm rem}^2 \sum_k x_k} \,, \tag{4.45}$$

with the Källén function given by eq. (4.33) and W_1 and W_2 the total transverse masses of each of the two beam remnant systems,

$$W_1^2 = \left(\sum_j x_j\right)\left(\sum_j \frac{m_{\perp j}^2}{x_j}\right), \qquad W_2^2 = \left(\sum_k x_k\right)\left(\sum_k \frac{m_{\perp k}^2}{x_k}\right). \tag{4.46}$$

Finally, for physical choices of $x_{j,k}$ and primordial $k_\perp$, the sum of the individual remnant transverse masses must be smaller than that of the total remnant system, $W_1 + W_2 < W_{\rm rem}$. If this is not the case, new $k_\perp$ sets and/or new $x_{j,k}$ values are tried, until a physical set of values is found.

5. Colour correlations and string topologies

The formalism described in the previous sections may be used to obtain a sequence of hard scatterings with associated initial- and final-state showers and ordered by the $p_\perp$ of the hard interactions they contain. Kinematics is completely specified for all partons involved in the scatterings, in the associated showers as well as in the left-behind beam

JHEP03(2004)053

remnants. However, at some point, the time evolution of this system results in inter-parton distance scales larger than about a femtometer, where the perturbative QCD description in terms of partons breaks down and must be supplanted by one of hadrons. Now, if each parton hadronized independently of the rest of the event, the information on kinematics and flavours alone would suffice to pass from the language of partons to that of hadrons, but confinement is precisely *not* a strictly local effect. Rather, it is a statement about a (colour singlet) *system* of partons, and hadronization is one concerning the evolution when colour charges inside such a system have been imparted with large momenta *relative to each other*. It is therefore not meaningful to study the hadronization of a single parton in isolation. Instead, it becomes necessary to consider the interplay between colour charges and to take correlations into account when modeling the hadronization process.

In this section, we begin by considering the hadronization of systems containing non-zero baryon numbers, but where the colours of all participating partons are known. We will use the planar approximation of QCD as a starting point, with a junction picture for the baryons.

Next, since the hard-scattering and parton-shower histories discussed above do not provide sufficient colour information — specifically information about how the different scattering initiators are correlated in colour is completely absent — we consider several possibilities for the assignment of correlated colours to the parton-shower initiators of the scatterings. Combined with the hadronization model, this allows us to study what is obtained under 'minimal' assumptions.

Finally, further issues are whether the original colour arrangement survives all the way to the long-distance hadronization era, and whether nonlinear effects arise in the hadronization process itself. That is, with several partons and string pieces moving out from the collision process, these partons and pieces will largely overlap in space and time. We do not know whether such overlaps can lead to colour rearrangements or nontrivial hadronization effects, e.g. of the Bose-Einstein kind. In principle $e^+e^- \to W^+W^- \to q_1\overline{q}_2q_3\overline{q}_4$ offers a clean environment to study such crosstalk, but experimental results are inconclusive [79]. In hadronic collisions, Bose-Einstein studies by UA1 and E735 also give a splintered image [80]: the strength parameter of BE effects drops with increasing particle density, consistent with a picture where a higher multiplicity comes from having several independently hadronizing strings, but the BE radius also increases, which suggests correlations between the strings.

5.1 Hadronization

Taking colour interference into account when modeling the hadronization process could easily become an unmanageable task. One simplification (disregarding baryons for the moment) is to go to the limit with infinitely many colours, $N_C \to \infty$ [71]. In this limit the confinement force acting on a gluon is twice that on a quark, i.e. the gluon colour and anticolour charges decouple. Further, colour diagrams are planar, so that final-state colour-anticolour pairs are always uniquely matched, via an unbroken colour line through the diagram.

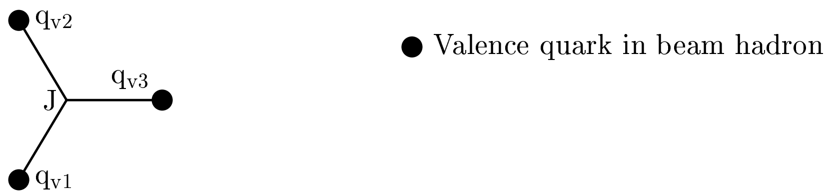


Figure 12: The initial state of a baryon, consisting of 3 valence quarks connected antisymmetrically in colour via a central 'string junction', J.

In the Lund string model [4], the two ends of such a colour line define a string piece. The string piece can be viewed as a Lorentz covariant and causal implementation of a linear confinement potential between the two partons. Transverse degrees of freedom here play no dynamical role, but one can visualize the colour field lines as compressed into a tube of a typical hadronic width ($\sim 1\,\mathrm{fm}$). As the partons move apart and a string piece is stretched out, it can break by the production of new $\mathrm{q\overline{q}}$ pairs that screen the endpoint colour charges. The q of one such break and the $\overline{\mathrm{q}}$ of an adjacent break together define a meson, which may be unstable and decay further.

The classical example is $\mathrm{e^+e^-} \to \mathrm{q\overline{q}g}$, where one may assign a red colour to the q, antired+green to the g and antigreen to the $\overline{\mathrm{q}}$, so that the string consists of two pieces, one q-g and one g-$\overline{\mathrm{q}}$. There is no piece directly between q and $\overline{\mathrm{q}}$, with observable consequences in the particle flow [72].

Turning now to baryon beams as the more interesting and difficult example, we picture the initial state of a baryon as consisting of three valence quarks connected in colour via a central 'junction', cf. figure 12. At the most basic level, such a picture finds its motivation by considering the simplest locally gauge invariant operator in SU(3) which carries non-zero baryon number [73]:

$$B_{f_1 f_2 f_3} = \epsilon^{\alpha_1\alpha_2\alpha_3} \prod_{i=1}^{3} P \left[\exp\left(ig \int_{\mathcal{P}(x,x_i)} G_\mu \, \mathrm{d}x^\mu \right) q_{f_i}(x_i) \right]_{\alpha_i} . \tag{5.1}$$

Physically, this operator assigns the space-time coordinates x_i to three valence quarks (with flavours f_i and colours α_i) and connects each of them via the gluon field G_μ along the path $\mathcal{P}$ to the point x (with P representing the path-ordering operation), which we may identify as the locus of the string junction. Such ideas were already introduced in the early string model of hadrons [73, 74, 75], and have been used to construct baryon wavefunctions in confinement studies [76]. In a recent article [3], we argued that this picture also arises naturally from string energy minimization considerations.

In a collision, the fact that a gluon carries colour implies that the junction will, in general, be separated in colour space from the original valence quarks. As a simple example, consider a valence $\mathrm{qq} \to \mathrm{qq}$ scattering in a pp collision. The exchanged gluon will flip colours, so that each junction becomes attached to a q from the other proton.

Hence the junction may well end up colour-connected to partons — or chains of partons — which are widely separated in momentum space and of which no two may be naturally considered to form a diquark system. To describe the hadronization of such systems, a model capable of addressing colour topologies containing explicit non-zero baryon numbers,

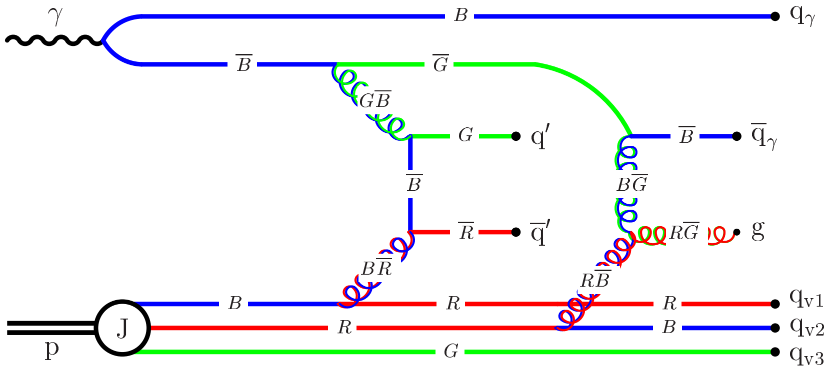


Figure 13: Example of colour assignments in a γp collision with two interactions. Explicit colour labels are shown on each propagator line. In this example, the string system containing the junction is spanned by J connected to q_γ, to $\mathrm{q}_{\mathrm{v}3}$, and via g to q′, as can be seen by tracing each of the three colour lines to the junction.

here in the form of junctions, becomes necessary. Such a model was first developed in [3], for dealing with the colour topologies that arise in baryon number violating supersymmetric scenarios. In the following, we show how this approach may be applied in a multiple interactions context to describe the physics of beam remnants.

We begin by considering a simplified situation where only one of the initial beam particles is a baryon. Leaving the ambiguities in assigning correlated colours aside for the moment, figure 13 gives an example of how the colour structure of a γp collision might look. In this example, the final-state colour-singlet system containing the junction consists of the three string pieces J—q_γ, J—$\mathrm{q}_{\mathrm{v}3}$, and J–g–q′. The two other string systems in the event, $\mathrm{q}_{\mathrm{v}1}$–$\overline{\mathrm{q}}'$ and $\mathrm{q}_{\mathrm{v}2}$–$\overline{\mathrm{q}}_\gamma$ are standard and do not concern us here.

To understand how the junction system hadronizes, the motion of the junction must first be established. This can be inferred from noting that the opening angle between any pair of the connected string pieces is 120° in the rest frame of the junction, i.e. in that frame the system consisting of the junction and its nearest colour-connected neighbours looks like a perfect Mercedes topology. This is derivable [3] from the action of the classical string [74] (which has a linear potential and thus exerts a constant force), but follows more directly from symmetry arguments.

Note that the junction motion need not be uniform. In the example above, one of the string pieces goes from the junction, via g to q′. At early times, the junction only experiences the pull of its immediate neighbour, g, and the direction of q′ is irrelevant. However, as the gluon moves out from the origin, it loses energy to the string traced out behind it. From the point when *all* its energy has been converted to potential energy of the string and this information has propagated back to the junction, it will be the direction of q′ which determines the direction of the 'pull' exerted by this string piece on the junction, and not that of the gluon. In the general case, with arbitrarily many gluons, the junction will thus be 'jittering around', being pulled in different directions at different times.

However, rather than trying to trace this jitter in detail — which at any rate is at or below what it is quantum mechanically meaningful to speak about — we choose to define an effective pull of each string on the junction, as if from a single parton with four-momentum given by [3]:

$$p_{\text{pull}} = \sum_{i=1}^{n} p_i \exp\left(-\sum\nolimits_{j=1}^{i-1} \frac{E_j}{E_{\text{norm}}}\right), \tag{5.2}$$

where the outermost sum runs over the parton chain which defines the string piece, from the junction outwards (in colour space), and where the sum inside the exponent runs over all gluons closer to the junction than the one considered (meaning it vanishes for $i = 1$). The energy normalization parameter E_{norm} is by default associated with the characteristic energy stored in the string at the time of breaking, $E_{\text{norm}} \simeq 1.5\,\text{GeV}$. Naturally, the energies E_j should be evaluated in the junction rest frame, yet since this is not known to begin with, we use an iterative sequence of successively improved guesses.

With the motion of the junction determined, the fragmentation of the system as a whole may now be addressed. Since the string junction represents a localized topological feature of the gluon/string field, we would not expect the presence of the junction in the string topology to significantly affect the fragmentation in the regions close to the endpoint quarks. Specifically, in an event where each of the three endpoint quarks have large energies in the junction rest frame, the energies of the leading and hence hardest particles of each jet should agree, on average, with that of an equivalent jet in an ordinary two-jet event.

The hadronization model developed in [3] ensures this by fragmenting each of the string pieces outwards-in, as for a normal $q\overline{q}$ string (in both cases opposite to the physical time ordering of the process). The leading quark of a string piece is combined with a newly created quark-antiquark pair to form a meson plus a new leftover quark, and so on. Parton flavours and hadron spins are selected in a manner identical to that of the ordinary string, as are fragmentation functions and the handling of gluon kinks on the string pieces.

However, junctions were not included in the original string model, so here a new procedure needs to be introduced. If all three string pieces were fragmented in the above way until little energy was left in each to form more hadrons, then it would be extremely unlikely that the resulting leftover system of three unpaired quarks would just happen to have an invariant mass equal to that of any on-shell baryon. While one could in principle amend this by shuffling momentum and energy to other hadrons in the vicinity of the junction, such a procedure would be arbitrary and result in an undesirable and large systematic distortion of the junction baryon spectrum. The way such systematic biases are avoided for ordinary $q\overline{q}$ strings is to alternate between fragmenting the system from the q end and from the $\overline{q}$ end in a random way, so that the hadron pair that is used to ensure overall energy-momentum conservation does not always sit at the same location. Thus, while the distortion is still local in each event, it is smeared out when considering a statistical sample of events.

In the case of a junction system, such a procedure is not immediately applicable. Instead, we first fragment two of the three string pieces, from their respective endpoint quarks inwards. At the point where more energy has been used up for the fragmentation than is

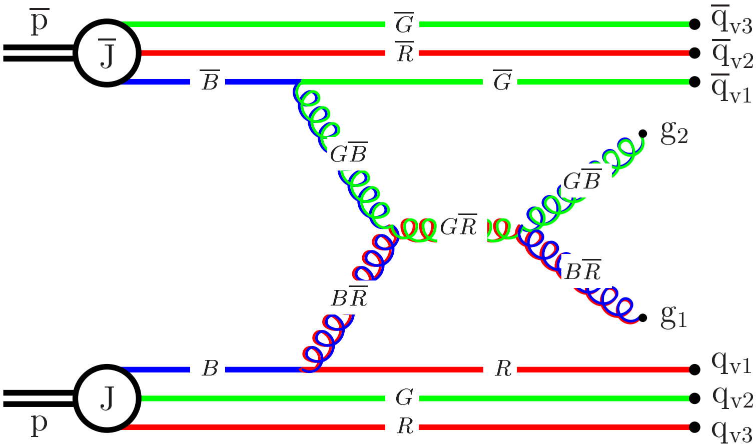


Figure 14: Example of colour assignments in a $p\overline{p}$ collision, with explicit colour labels shown on each propagator line. Note that the blue colour line starting on the junction J is connected via the colour flow of the hard scattering to the antiblue colour line of $\overline{J}$.

available in the piece, the last quark-antiquark pair formed is rejected and the fragmentation is stopped. The two resulting unpaired quarks, one from each fragmented string piece, are then combined into a single diquark, which replaces the junction as the endpoint of the third string piece. Subsequently, this last string piece is fragmented in the normal way, with overall energy and momentum conservation ensured exactly as described for ordinary strings above. In order to minimize the systematics of the distortion and ensure that it is at all possible to produce at least two hadrons from this final string system, we choose to always select the highest energy string piece as the last to be fragmented. It was shown in [3, 77] that this asymmetry in the description does not lead to large systematic effects.

In proton-proton collisions, two junction systems will be present, but it is physically impossible for these to be connected in colour. Hence, the hadronization of each system again proceeds exactly as described above. However, in $p\overline{p}$ collisions a new possibility arises, as depicted in figure 14. This simple example goes to illustrate that a junction and an 'antijunction' may become colour-connected by the colour exchanges taking place in a given process. In such cases, the fragmentation of each of the junctions is no longer disconnected from what happens to the other one; instead the fragmentation of the system as a whole must be considered. The necessary generalization of the principles outlined above to the case of connected junction-junction systems [3] is not very complicated.

As before, two of the three strings from a junction are fragmented first, outwards-in towards the junction, but in this case we always choose these two string pieces to be the ones not connected to the other junction. Diquarks are then formed around each junction exactly as before. What remains is a single string piece, spanned between a diquark at one end and an antidiquark at the other, which can be fragmented in the normal way. In fact, the only truly new question that arises at this point is how to generalize eq. (5.2) to describe the pull of one junction on another. Here gluons on the string between the two junctions are considered as normally, i.e. their momenta are added, with a suppression factor related to the energy of the intermediate gluons. The partons on the far side of the

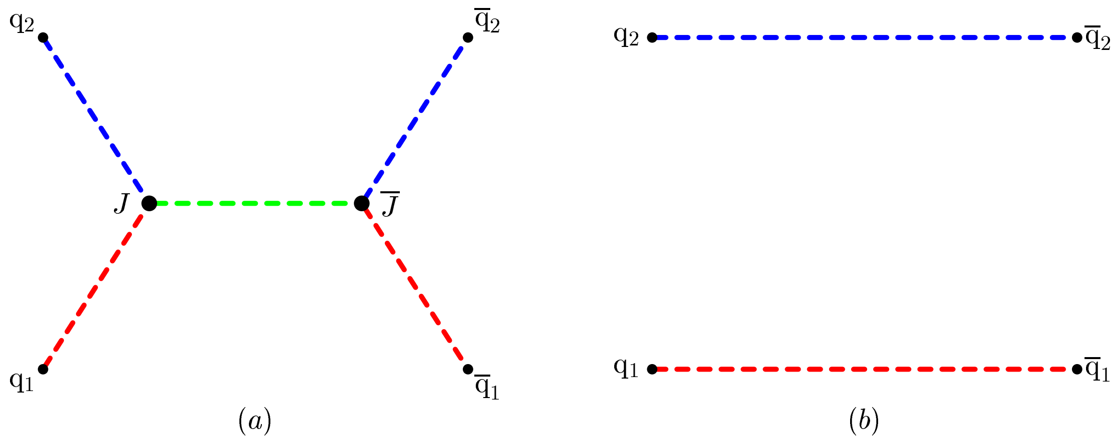


Figure 15: (*a*) A string system (dashed lines) spanned between four quarks and containing a junction and an antijunction. (*b*) The same parton configuration in colour space but with an alternative string topology. In (*a*) the beam baryon numbers will still be present in the final state, while in (*b*) they will have disappeared through annihilation.

other junction also contribute their momenta, separately for each of the two strings, with an energy sum suppression now given by the intervening gluons on that particular string, plus the gluons on the junction-junction string.

However, an alternative topology is also possible, where the junction and the antijunction annihilate to produce two separate $q\overline{q}$ systems [3], as illustrated in figure 15*b*. While it is not clear from basic principles how often this should happen, it seems likely that, for a given event, the topology which has the minimal string length is the one selected dynamically. In this case, the string topology depicted in figure 15*a* would result when the q_1q_2 and $\overline{q}_1\overline{q}_2$ opening angles are small, while the topology in figure 15*b* would result if the $q_1\overline{q}_1$ and $q_2\overline{q}_2$ opening angles are the small ones. Since, in the context we discuss here, the quarks colour-connected to the junction will more often than not reside in the beam remnants, we do not expect annihilation between the incoming baryon numbers to be a large effect. Indeed, for the range of more realistic models that are investigated in section 6 below, junction-junction annihilation is a feature of less than 1% of the events at Tevatron energies.

An ugly situation occurs in the rare events when the two junctions are connected by *two* colour lines. If these lines contain intermediate gluons, it would be possible but difficult to fragment the system, in particular when the energy of these gluons becomes small. Without intermediate gluons, a first guess would be that the junction and antijunction annihilate to give a simple string spanned between a quark and an antiquark endpoint, so that the original baryon numbers are lost. However, this assumes that the system starts out from a point in space and time, a commonly used approximation in the string language. Viewed in the transverse plane of the collision, the original positions of the junctions and of the hard scatterings involved could well be separated by distances up to a fm, i.e. the intervening strings could have energies up to a GeV. It may then be that the strings can break before the junctions annilate, so that a baryon-antibaryon pair nevertheless is produced. A detailed modeling would be required, beyond the scope of the current study, and possibly beyond the validity of the string framework, so for now we choose to reject these rare events.

5.2 Initial-state colour correlations

In the planar approximation, $N_C \to \infty$, a $2 \to 2$ process, such as gg $\to$ gg, can receive contributions from several possible colour flows, but the cross section for a colour flow is uniquely defined, so that each flow can be selected according to its appropriate weight [78]. Furthermore, in our leading-order parton showers, the colour flow is well-defined in each branching. Within each separate interaction and its associated shower activity, the colour flow can thus be selected unambiguously. In events with only one hard interaction, the colour of the shower initiator can also uniquely be hooked up to that of the beam remnant. Thus a knocked-out quark leaves behind a colour antitriplet beam remnant, a gluon leaves a colour octet beam remnant, and so forth.

Unfortunately, once several interactions are allowed, there is no longer a unique answer how to hook up the different shower initiators with the beam-remnant partons. To illustrate, consider an incident meson out of which n gluons are kicked out. These gluons may be ordered in colour sequence such that the colour of the quark matches the anticolour of one gluon, which then has another colour that matches the anticolour of another gluon, and so on till the antiquark. Obviously there are $n!$ such possible arrangement of the n gluons, each leading to a unique colour topology for the hadronizing partonic system. Perturbation theory has nothing to say about the relative probability for each of these configurations; the colour correlations we now consider arise at scales below the parton-shower cutoff $Q_0 \sim 1\,\mathrm{GeV}$. Further arrangements would exist if we allowed some of the above gluons to form a separate colour singlet, disconnected from the sequence between the quark and antiquark ends; some of these could contribute to diffractive topologies.

So far we only considered the planar colour topologies of the $N_C \to \infty$ limit. The real-world $N_C = 3$ offers further complications. First, interference terms of order $1/N_C^2$ — modulated by kinematical factors — arise between different possible colour flows in hard processes. The more partons are involved, the more assignments need be made, and the larger the total uncertainty. Even at the perturbative level it is thus no longer possible to speak of a unique colour arrangement. Second, the situation is more complicated for baryon beams. As described above, the initial state of a baryon, before any scatterings occur, is represented by three valence quarks, connected in a Y-shaped topology via a central junction which acts as a switchyard for the colour flow and which carries the net baryon number. This situation is illustrated in figure 12. Each of the gluons considered in the meson-beam example above may now be arranged in colour on either of the three string pieces, leading to a further multiplication of possibilities.

We choose to address this question by determining a sequence of fictitious gluon emissions by which this configuration evolves (in colour space) to give rise to the parton shower initiators and beam remnant partons actually present in a given event. We here assume that only the minimal number of emissions required to obtain the given set of initiators and remnants is dynamically relevant. Further, since sea quarks together with their companion partners can pairwise be associated with a gluon branching below the parton shower cutoff, only gluon emissions remain to be considered. (This also means that a sea quark, in our model, can never form a colour singlet system together with its own companion.)

5.2.1 Random colour correlations

The simplest solution would be to assume that Nature arranges these correlations randomly, i.e. that gluons should be attached to the initial quark lines in a random order, see figure 16. In this case, the junction (and hence the baryon number) would rarely be colour connected directly to two valence quarks in the beam remnant, even in the quite common case that two such quarks are actually present (multiple valence quark interactions are rare). It should be clear that the migration of the baryon number depends sensitively upon which partons in the final state the junction ends up being connected to (see further section 5.1). The conclusion is that if the connections are *purely* random, as above, then the baryon number will in general be disconnected from the beam remnant valence quarks. Hence the formation of a diquark in the beam remnant would be rare and the baryon number of the initial state should quite often be able to migrate to small x_F values, as previously illustrated in figure 11. One could expect this longitudinal migration to be accompanied by a migration in the transverse plane, such that the junction baryon should generally migrate to larger $p_\perp$ values when the junction is allowed to 'float' more. However, as figure 17*a* illustrates, no large differences in the total $p_\perp$ spectrum are apparent when comparing the new model (thick dashed) with the old Tune A (solid).

The reason that no large transverse migration effect is visible, relative to the old model, is that the latter also has a broader leading-baryon $p_\perp$-spectrum than for normal baryons, as follows. In the old model, the beam remnant diquark (around which the junction baryon forms in the fragmentation) always receives the full primordial $k_\perp$ kick from the hard interaction initiator, by default a gaussian distribution with a width of 1 GeV. In the fragmentation process, when the baryon acquires a fraction of the diquark longitudinal momentum, it obtains the same transverse momentum fraction. Additional $p_\perp$ will be imparted to the junction baryon from the newly created quark, at the level of 0.36 GeV, but with some dependence on the momentum-space location of its nearest neighbour in colour space. Essentially, these two effects combine to yield the solid curve in figure 17*a*.

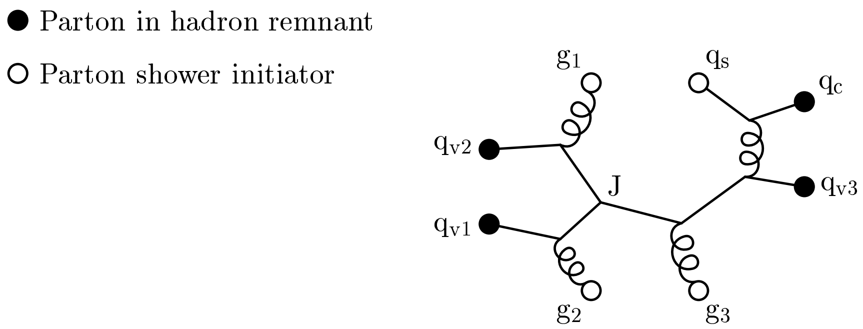
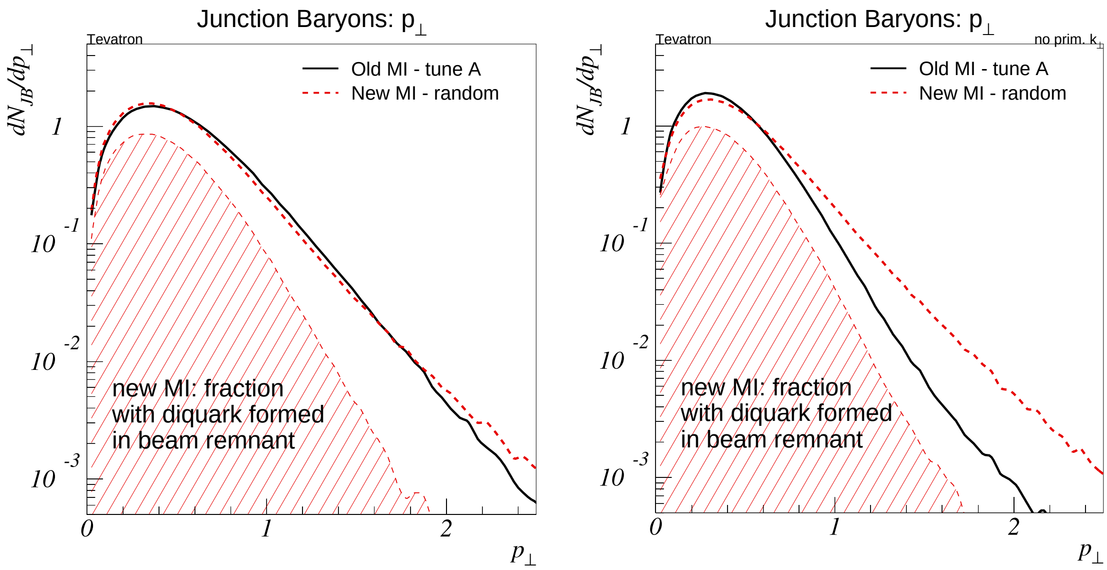


Figure 16: Example of how a given set of parton shower initiators could have been radiated off the initial baryon valence configuration, in the case of the 'purely random' correlations discussed in the text. In this example, the baryon number is disconnected from the beam remnant. Instead, it is the final-state partons connected to the colour lines of g_1, g_2, and g_3 which determine how the junction moves and hence how the baryon number flows in the event.



(a) (b)

Figure 17: $p_\perp$ spectra for junction baryons, (a) with primordial $k_\perp$ switched on, and (b) with primordial $k_\perp$ switched off. The shaded area represents the distribution in the new model of those junction baryons which arose by first forming a diquark in the beam remnant, cf. section 4.2.

In the new model, we must distinguish between the old-model-like case when a diquark is formed in the beam remnant, on one hand, and those cases where the junction is 'free' to migrate, on the other.

If a diquark is formed, then it consists of two undisturbed beam remnant quarks which are colour connected to the junction, and the situation is indeed very similar to the old model. The baryon forming around this diquark receives $p_\perp$ from three sources. Firstly, the diquark will have some intrinsic primordial $k_\perp$, distributed according to eq. (4.27). Since the diquark resides entirely within the beam remnant, this $k_\perp$ will always be at the level of the fragmentation $p_\perp$. Secondly, primordial $k_\perp$ will be imparted to the diquark by recoil effects from other beam remnant and initiator partons. In the case that primordial $k_\perp$ kicks are compensated for uniformly by all initiator and remnant partons, it is normally impossible for the diquark to acquire more than a fraction of the hardest interaction initiator's primordial $k_\perp$. Even when $k_\perp$ compensation is more local, by straightforward combinatorics, the more initiators present in the event, the smaller the chances that the initiator parton(s) closest in colour to the diquark is associated with a scattering at large Q^2. Hence, again according to eq. (4.27), it is apparent that the diquark usually will not receive a very hard primordial $k_\perp$ kick. Thus, such a diquark will in general have a smaller total primordial $k_\perp$ than a diquark in the old model. As before, the baryon will keep a large fraction of this diquark $p_\perp$ in the fragmentation process, as well as obtaining extra fragmentation $p_\perp$. The net result is a softer junction baryon transverse momentum spectrum than in the old model, as can be verified by comparing the asymptotic slope of the shaded area in figure 17a with that of the solid curve. This conclusion is further established by the observation that, when primordial $k_\perp$ effects are not included, see figure 17b, indeed

JHEP03(2004)053

the spectrum of the old model becomes almost identical to that of the shaded region. In addition, it can already here be recognized that the junction baryon must have larger $p_\perp$ in those events where a diquark is *not* formed, by comparing the slopes of the full junction baryon spectrum (dashed curves) with those of the shaded regions in either figure. We now study this further.

If a diquark is not formed, then the junction may *a priori* be colour connected to partons going in widely different directions in the transverse plane. Nonetheless, as was described in section 5.1, the fragmentation occurs in such a way that the junction baryon is always the last, i.e. normally slowest, hadron to be formed in either of the three directions. Hence, while the colour neighbours of the junction may themselves have large transverse momenta, this momentum will in general be taken by the leading hadrons formed in the fragmentation and not by the junction baryon. Unless two of those partons are going in roughly the same direction in φ, the junction baryon itself will still only obtain a fairly small $p_\perp$. The end result is a rather small $p_\perp$ enhancement, that is masked by the decreased primordial $k_\perp$, figure 17.

Thus, the main difference in the new model is that the beam baryon number can migrate *longitudinally* to a much larger extent than in the old model. Empirically, it may be desirable to be able to limit the degree to which this baryon number stopping occurs, and furthermore both perturbative and impact-parameter arguments allow much of the activity to be correlated in 'hot spot' regions that leave the rest of the proton largely unaffected. Therefore a free suppression parameter is introduced, such that further gluons more frequently connect to a string piece that has already been disturbed. In this way, gluons would preferentially be found on one of the three colour lines to the junction. This will reduce the amount of baryon number stopping and is an important first modification, but most likely it is not the *only* relevant ordering principle.

5.2.2 Ordered colour correlations

With the gluons connected preferentially along one of the three colour lines to the junction, we now address the question of their relative order along that line. If this order is random, then strings will in general be stretched criss-cross in the event. This is illustrated in figure 18*a* for a very simplified situation. However, it is unlikely that such a scenario catches all the relevant physics. More plausible is that, among all the possible final-state colour topologies, those that correspond to the smaller total string length are favoured, all other aspects being the same. One possible way of introducing such correlations is illustrated in figure 18*b*.

In case (a) of figure 18 there are four string pieces criss-crossing the rapidity range between the systems I_2 and I_3, while in case (b) there are only two string pieces spanning this range. Hence the total string length should on average be smallest for the latter type of correlations. However, the rapidity distance is not the only variable determining the string lengths, also the transverse separations play a role. Moreover, when the interactions exchange colour between the colliding objects, there is no longer a unique correspondence between the colour flow of the initial state and that of the final state.

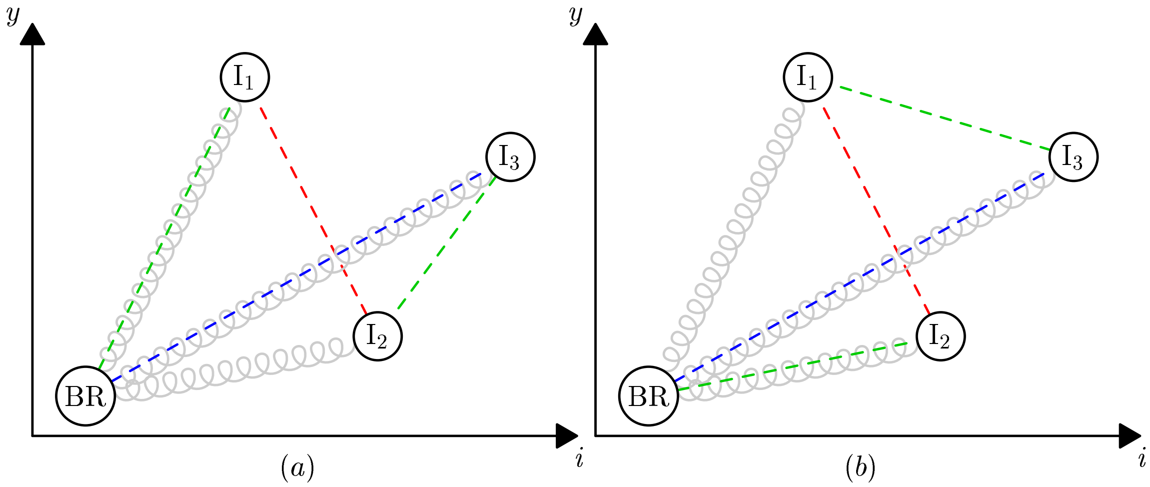


Figure 18: Example of initial-state colour correlations in an imagined event in which three gluons have been knocked out of an incoming hadron by colourless objects (for simplification), with no parton showers. In case (*a*) the gluons have been randomly attached in colour to each other and to the beam remnant, as indicated by the dashed lines, whereas in case (*b*) the connections follow the rapidities of the hard scattering systems.

To investigate these effects quantitatively, we consider three different possibilities for initial-state colour correlations (with the suppression of attachments breaking up the beam remnant applicable to all cases):

1. Random correlations, as in figure 18*a*.

2. Initiator gluons are attached preferentially in those places that order the hard scattering systems in rapidity, as in figure 18*b*. The rapidities are calculated at a stage before primordial $k_\perp$ is included. Hence, $y = \frac{1}{2} \ln \frac{x_1}{x_2}$.

   For beam remnant partons, the rapidities are not yet known at the stage discussed here, since the initial-state colour connections are in our framework made *before* primordial $k_\perp$ and beam remnant x values are assigned. However, beam remnant partons are almost by definition characterized by having large longitudinal and small transverse momenta. Thus, we assign a fixed, but otherwise arbitrary, large rapidity to each of the beam remnant partons, in the direction of its parent hadron. Finally, gluons are attached sequentially to the initial valence topology, with the attachments ordered by minimization of the measure

$$\Delta y = |y_g - y_1| + |y_g - y_2| \,, \tag{5.3}$$

   where y_g is the rapidity associated with the attached gluon and $y_{1,2}$ are the rapidities associated with the partons it is inserted between. For those gluons which appear only as parents of sea quark pairs, the rapidity of the most central of the daughters is used.

   Note that, since the same hard-scattering rapidities are used for both beam remnants, the ordering in the two remnants will be closely correlated in this scenario, at least as long as only gluon-gluon interactions are considered.

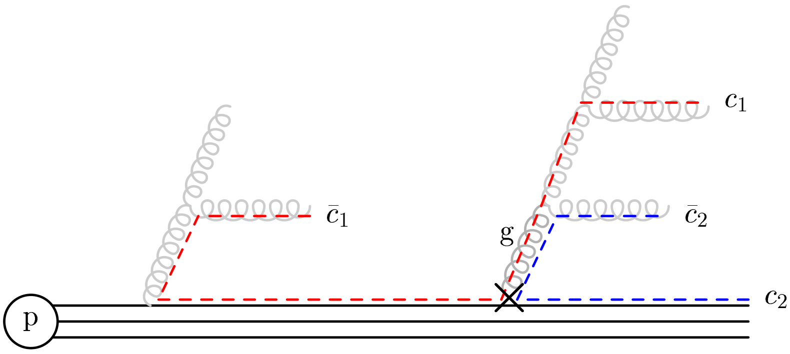


Figure 19: Example showing the colour flow produced by attaching the gluon g at the place indicated by the cross.

3. Initiator gluons are attached preferentially in those places that will give rise to the smaller string lengths in the final state. This is the most aggressive possibility, where the actual momentum separations of final-state partons, together with the full colour flow between the two sides of a hadronic collision, is used to determine which gluon attachment will result in the smallest increase in potential energy (string length) of the system, with each gluon being attached one after the other. The measure we use to define the increase in string length, for a particular attachment, is [3, 81]

$$\Delta\lambda = \ln\left[\frac{2}{m_0^2}\frac{(p_{c_1}\cdot p_{\bar{c}_1})\,(p_{c_2}\cdot p_{\bar{c}_2})}{(p_{\bar{c}_1}\cdot p_{c_2})}\right], \tag{5.4}$$

where m_0 is a normalization constant, which drops out when comparing the string lengths of two different gluon attachments, and c_1 ($\bar{c}_2$) represents the final-state parton carrying the colour (anticolour) index of the attached gluon. To illustrate, figure 19 shows the partons that enter the above expression for a specific example. Before the attachment, a single string piece is spanned between the final-state partons that carry the colour indices denoted $\bar{c}_1$ and c_2. After the attachment, there are two string pieces, one that is spanned between c_1 and $\bar{c}_1$, the other between c_2 and $\bar{c}_2$, hence the *increase* in string length is given by the expression eq. (5.4).

As above, however, note that neither primordial $k_\perp$ nor beam remnant longitudinal momenta have yet been assigned at this stage. Simplified kinematics are therefore set up, to be used only for the purpose of determining the colour connections: the momentum remaining in the beam remnant on each side is divided evenly among the respective remnant partons (junctions are here treated simply as ‘fictitious partons’, receiving the same momentum as the ‘real’ remnant partons), and primordial $k_\perp$ effects are ignored. Thereby, parton pairs involving (at least) two partons in one of the beam remnants will come to have zero invariant mass, hence the total $\Delta\lambda$ will be negative infinity for such pairs. Obviously, this is not desirable; one string piece with vanishing invariant mass should not affect the comparison, hence we impose a minimum invariant mass of m_0 for each string piece. If knowledge of the full

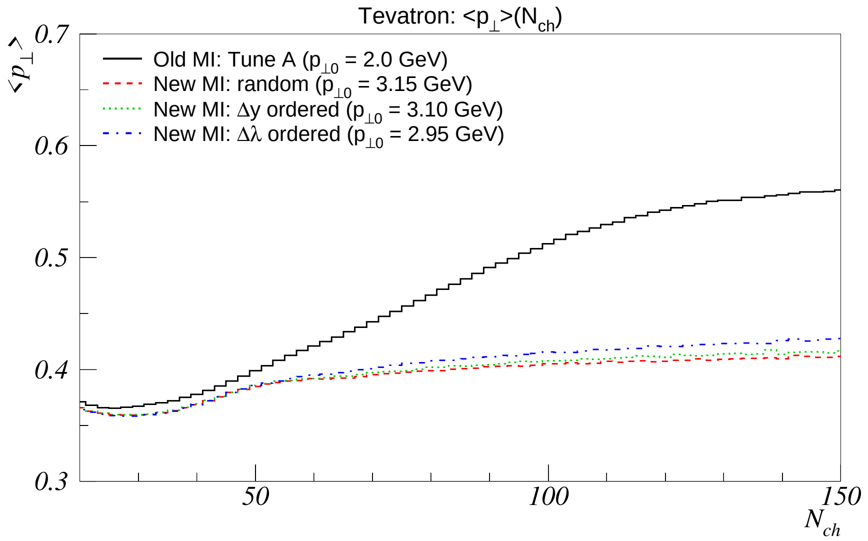


Figure 20: $\langle p_\perp \rangle$ vs. n_{ch} at the Tevatron for Tune A (solid lines), and for the new model with random (dashed lines), rapidity ordered (solid lines), and string length ordered (dotted lines) correlations in the initial state. Note that the origo of the plot is *not* at (0,0). For each of the new MI scenarios, $p_{\perp 0}$ was selected to give the same average charged multiplicity as Tune A, with the same impact parameter dependence as Tune A (i.e. a double gaussian matter distribution).

kinematics of the final state was available a better choice could of course be made here. However, these two aspects are intertwined. The kinematics of the final state may depend on the colour connections assumed for the initial state (see section 4.3 above), and vice versa. Our choice has been to determine the initial-state colour connections first, and then subsequently construct the final-state kinematics, hence some approximation is necessary at this point.

A variable which we have found to be sensitive to the colour connections in an event is the mean $p_\perp$ vs. charged multiplicity, $\langle p_\perp \rangle(n_{\mathrm{ch}})$ [1]. In scenarios with large string lengths, each additional interaction would result in a large increase in hadron multiplicity. This large multiplicity per interaction means that, in such scenarios, observed average charged multiplicities are reproduced with comparatively large values of $p_{\perp 0}$, i.e. with only a few parton-parton interactions taking place per event. Hence, correspondingly little perturbative $p_\perp$ is generated. On the other hand, in scenarios with smaller string lengths, comparatively more interactions would be required to produce the same multiplicity, hence more perturbative $p_\perp$ would be generated per charged particle, bringing $\langle p_\perp \rangle(n_{\mathrm{ch}})$ up.

In figure 20, we show the $\langle p_\perp \rangle$ vs. n_{ch} distribution for each of the possibilities described above. In all cases, $p_{\perp 0}$ was first selected so as to give identical average multiplicities, corresponding to the multiplicity obtained with Tune A. Since the $\Delta\lambda$ ordering results in the largest average number of interactions for a given multiplicity, it has the largest average $p_\perp$ per particle of the new scenarios, while the random ordering results in the smallest number of interactions.

One also notes that Tune A, which more or less agrees with recent experimental data [61, 56, 58], shows an even steeper rise with n_{ch} than any of the new scenarios. This tune of the old model is such that the partons produced by subsequent scatterings will almost always be hooked up to the existing configuration in the way that minimizes the total string length. This is more or less like the $\Delta\lambda$ ordering described above, but with the essential difference that the $\Delta\lambda$ ordering only concerns the colour lines that are present in the *initial state*, while the ordering of parton attachments in the old model occurs in the *final state*, without any attempt at constructing a consistent colour flow in the event.

From these observations, an interesting inference can be made. By the failure of even the $\Delta\lambda$ ordering of the colour lines in the initial state to describe the $\langle p_{\perp}\rangle(n_{\mathrm{ch}})$ distribution, it appears that the colour flow in physical events cannot be correctly described by merely arranging the colour lines present in the initial state. We imagine two possible causes for this. Firstly, the initial-state showers associated with each scattering are constructed by backwards DGLAP evolution of each scattering initiator separately, down to the shower cutoff scale. This does not take into account the possibility that the showers could be intertwined, i.e. that a parton at low virtuality, but above the shower cutoff scale, could have branched to give rise to *two* higher-virtuality scattering initiators. Secondly, one or more mechanisms causing colour exchanges between the showers may be active, both in the initial state as well as in the final state. Below we present some first studies related to the topic of colour exchanges, well aware that no simple solutions are to be expected.

5.3 Final-state colour (re-)connections

To investigate how much more we need to 'mess around with the colours', we study a crude model of colour exchanges in the final state. Essentially, we rearrange the colour connections between the final-state partons in a manner that, taken to the extreme, will converge on that string configuration which has the smallest total 'string length', according to the λ measure introduced above. This corresponds roughly to a minimization of the multiplicity produced when the parton system hadronizes. Note that we only apply this procedure to events where at least two interactions have occurred, to avoid the reconnections leading to large central rapdity gaps, i.e. diffractive topologies.

For a given configuration of final-state partons in momentum and colour space, we apply an iterative procedure that successively brings down the total string length of the system, by the following steps:

1. First, we have assigned Les Houches Accord style colour tags [82] to all partons, so that each colour tag in the final state is matched by exactly one corresponding anti-colour tag in the final state, with one string piece spanned between them. Junctions are special, since colour lines end there. In the following, we do not consider string pieces ending on junctions.

2. Secondly, we decide on a fraction, F, of the colour tags present in the event for which we will attempt to make a reassignment. Note that F can be larger than one, since several different reassignments ($n(n-1)/2$ for n colour tags, neglecting junctions) are normally possible.

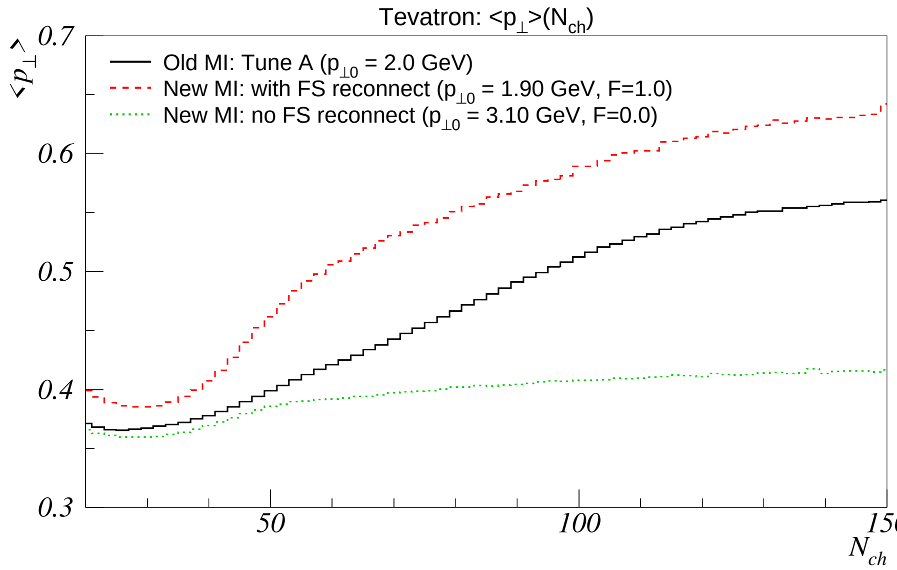


Figure 21: $\langle p_\perp \rangle$ vs. n_{ch} at the Tevatron for Tune A (solid line), and for the new model with (dashed line) and without (dotted line) final-state reconnections allowed. Both of the new models use rapidity ordering of the colour lines in the initial state and give the same average charged multiplicity as Tune A, with the same impact parameter dependence as Tune A.

3. Next, we select two colour tags at random, c_1 and c_2. Denoting the final-state parton carrying c_1 colour (anticolour) by i_1 (j_1) and the one carrying c_2 colour (anticolour) by i_2 (j_2), we compute the combined string length, λ, for the two string pieces i_1–j_1, i_2–j_2:
$$\lambda = \ln\left(\frac{2p_{i_1}\cdot p_{j_1}}{m_0^2}\frac{2p_{i_2}\cdot p_{j_2}}{m_0^2}\right). \tag{5.5}$$
By swapping e.g. the anticolours, a different string topology arises, i_1–j_2, i_2–j_1, with length
$$\lambda' = \ln\left(\frac{2p_{i_1}\cdot p_{j_2}}{m_0^2}\frac{2p_{i_2}\cdot p_{j_1}}{m_0^2}\right). \tag{5.6}$$
If $\lambda' < \lambda$, the colour reassignment is accepted, otherwise the original assigments are kept. If the fraction of colour tags tried so far is smaller than F, a new pair of random colour tags is selected.

4. Once the fraction F of colour tags has been tried, two things can happen. If at least one reconnection was made, then the colour topology now looks different, and the entire iteration is restarted. If no reconnection was made, the iteration ends.

Briefly summarized, we thus introduce the fraction F as a free parameter that controls the strength of colour reconnections in the final state.

As stated, this method is very crude and should not be interpreted as representing physics *per se*, but it does allow us to study whether a significant effect can be achieved by manipulating the colour correlations to reduce the string lengths. As illustrated by the dashed histogram in figure 21, this is very much the case. Here, we have allowed a large amount of reconnections to occur, $F = 1$, adjusting $p_{\perp 0}$ down so as to reproduce the average charged multiplicity of Tune A. As could be expected with this somewhat extreme choice of parameters, the new model now lies well above the data. (We shall return to more realistic tunings in section 6.)

However, this result should not be taken as evidence for the existence of colour reconnections in physical events. Rather it allows us to infer that, by changing the colour structure of events, it should be possible to obtain agreement with the data within our framework. It is encouraging that, by studying and attempting to describe this distribution, we may learn interesting lessons concerning the highly non-trivial issue of colour flow in hadronic interactions. We therefore plan to go further, to construct more physically motivated models for colour rearrangement between partons, both in the initial state and in the final state, and also to allow for the possibility of intertwining the initial-state showers.

Although all of these issues appear almost hopelessly complicated from the point of view of pure QCD, the salient features of the resulting physics may not be all that hard to penetrate. For instance, we imagine that the probability for two hard-scattering initiators to have originated from a common branching should be proportional to the probability that their spatial wavefunctions overlap, i.e. two very high-virtuality initial-state partons associated with different scatterings are most likely uncorrelated, since they only resolve very small distance scales in their parent hadron. On the other hand, two low-virtuality partons, even though associated with different scatterings, may very well have come from one and the same parent parton, since their wavefunctions are comparatively much larger. We will explore such and other ideas in a future study.

6. Model studies

In this section we concentrate on illustrating the properties of models that, as a baseline, roughly reproduce the charged multiplicity distribution of Tune A in $\mathrm{p\overline{p}}$ collisions at a centre-of-mass energy of 1800 GeV. Our studies only concern inelastic nondiffractive events, i.e. essentially the same as the experimentally defined (trigger-dependent) “min-bias” event sample; we will here use the two concepts interchangeably.

A general feature of the new multiple interactions modeling is that the added parton showers and the less efficient string energy minimization result in a higher multiplicity per interaction than Tune A. In order to arrive at the same average hadron multiplicity as Tune A, without too much colour reconnection required in the final state, a generally larger $p_{\perp 0}$ cutoff should be used. Table 1 lists three different tunes of the new framework, with successively smaller $p_{\perp 0}$ values and with different schemes for the initial-state colour correlations.

- The “Ran” model is based on a random ordering of the initial-state colour correlations, with a fairly large suppression of initiator gluon attachments to colour lines wholly within the beam remnant. Since only a minimal ordering of the colour correlations in the initial state are thus imposed, each additional interaction will *ab initio* give rise to a relatively large increase in hadron multiplicity. Therefore, a comparatively large cutoff $p_{\perp 0}$ is used, and the F parameter — controlling the amount of

Model name	IS Colour Ordering	$p_{\perp 0}$ [GeV]	F	Tevatron $\langle n_{\mathrm{int}} \rangle$	$\langle n_{\mathrm{part}} \rangle$	$\langle n_{\mathrm{rcp}} \rangle$	LHC $\langle n_{\mathrm{int}} \rangle$	$\langle n_{\mathrm{part}} \rangle$	$\langle n_{\mathrm{rcp}} \rangle$
Ran	random	2.50	0.55	3.3	21.5	18	4.2	40.2	45
Rap	Δy	2.40	0.55	3.6	22.8	19	4.5	43.5	49
Lam	$\Delta\lambda$	2.30	0.65	3.9	24.5	20	4.8	45.8	52
Tune A	–	2.00	–	5.7	19.2	–	6.9	27.7	–

Table 1: Parameters of the three models investigated in the text, and for Tune A where applicable. Also shown for each model is the average number of parton-parton interactions, $\langle n_{\mathrm{int}} \rangle$, the average number of final-state partons, $\langle n_{\mathrm{part}} \rangle$, and the average number of colour reconnections taking place, $\langle n_{\mathrm{rcp}} \rangle$, per min-bias collision at the Tevatron and at the LHC.

final-state reconnections — is likewise chosen fairly large, so as to get the correct average charged multiplicity.

- The "Rap" model uses the Δy measure introduced above to order the intial-state colour connections. $p_{\perp 0}$ can thus here be slightly smaller, allowing more interactions on the average (with the same F fraction) for the same average charged multiplicity.
- The "Lam" model employs the $\Delta\lambda$ ordering of the initial-state colour correlations. In principle, this model should provide the most ordered initial state of the three, and thus allow a smaller $p_{\perp 0}$ and/or F. Unfortunately, the earlier-mentioned limitations, that the beam remnant kinematics is not fully fixed when the minimization is performed, leads to final string lengths which are not significantly shorter than for the Δy ordering. Choosing a smaller $p_{\perp 0}$ for this tune, the F fraction is consequently also required to be slightly higher, in order to reproduce the Tune A average charged multiplicity.

Observe that all three models have a significant number of reconnections per event, cf. $\langle n_{\mathrm{rcp}} \rangle$ in table 1. As a fraction of the total number of potential colour rearrangements it is below the 10% level, but one should keep in mind that several clusters of partons appear in reasonably collimated jets, where reconnections would not be expected anyway. In this perspective, the amount of reconnections is quite significant.

Common for the new models is a rather smooth overlap profile ExpOfPow(1.8), as compared to the more peaked double gaussian of Tune A, this to better reproduce the shape of the Tune A multiplicity distribution. In addition, all of the new models assume that primordial $k_{\perp}$ kicks are compensated uniformly among all other initiator partons, that composite objects can only be formed in the beam remnant by valence quarks, and that initial-state colour connections breaking up the beam remnant are suppressed, so that the relative probability of attaching a gluon between two remnant partons (i.e. breaking up the remnant) as compared to an attachment where at least one 'leg' is outside the remnant is 0.01, whenever the latter type of attachment is possible.

Figure 22 shows the Tevatron multiplicity distributions of these models, as compared to Tune A. It is apparent that, while the average charged multiplicity is the same, the shape of the Tune A multiplicity distribution is not exactly reproduced by any of the models here

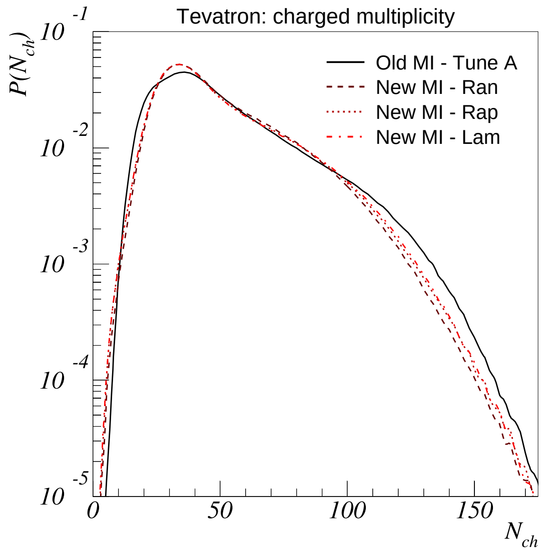


Figure 22: Multiplicity distributions for the Tevatron as obtained with Tune A (solid), and the Ran (dashed), Rap (dotted), and Lam (dash-dotted) models defined in table 1. In all cases, the average charged multiplicity is 49.5 (within ± 0.5).

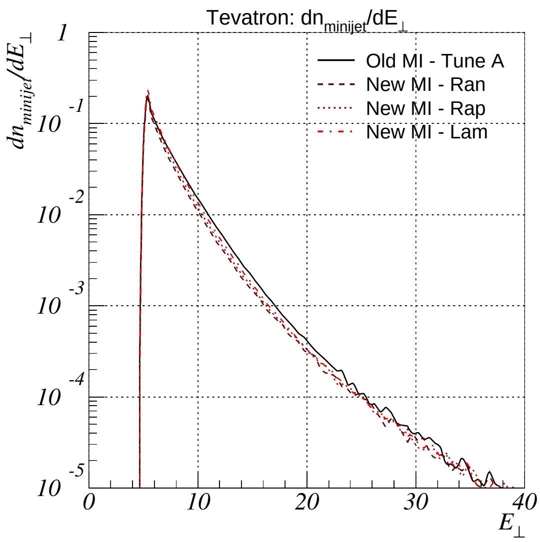


Figure 23: The number of jets as a function of jet $E_\perp$ (for $E_\perp > 5$ GeV) in min-bias collisions at the Tevatron. Results are shown for Tune A (solid), Ran (dashed), Rap (dotted), and Lam (dash-dotted), as defined in table 1.

investigated. This should not be taken too seriously; our aim is not to present full-fledged tunes, rather it is to explore the general properties of the new framework, and how these compare with those of Tune A.

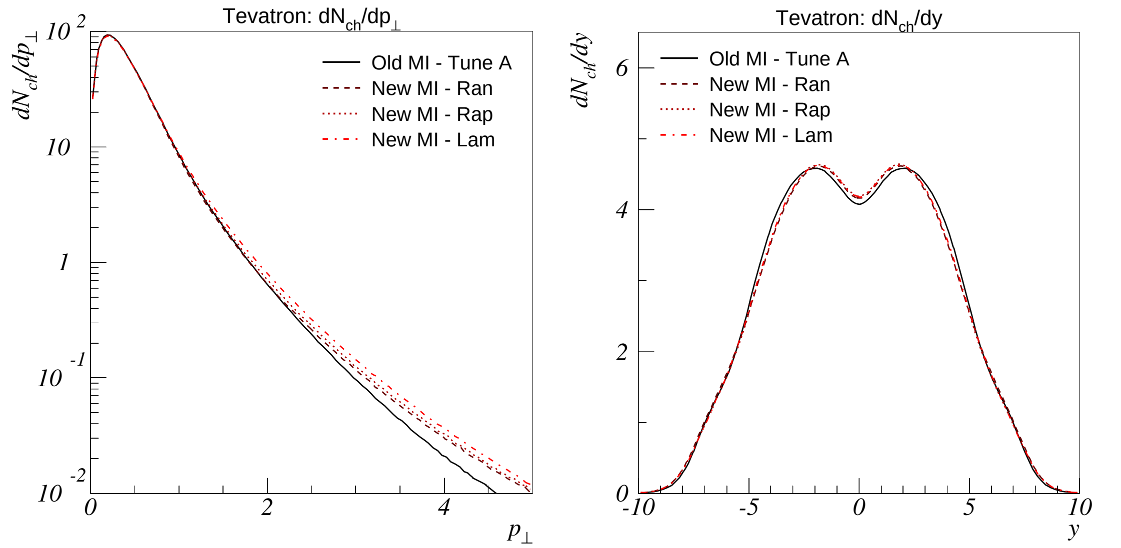
Below, we first present comparisons for $\mathrm{p\bar{p}}$ min-bias collisions at 1.8 TeV CM energy, highlighting the differences (and similarities) between the new models and Tune A. Thereafter, we apply the same models to the case of pp min-bias events at 14 TeV CM energy.

6.1 Comparisons at the Tevatron

Despite the differences between the old and new frameworks, the starting point in both cases is still that of a perturbative sequence of $p_\perp$-ordered scatterings. Especially for the hardest partons there should thus be next to no difference between the old and new frameworks. An illustration of this is given in figure 23, where the probability of finding a jet with transverse energy $E_\perp$ is plotted against $E_\perp$. A simple cone algorithm with cone size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.7$ has been used to cluster the jets, and only particles with $|\eta| < 2.5$ are included. As can be observed, there is hardly any difference between Tune A and the new models here. Further, the models exhibit similar charged particle spectra both in transverse momentum and in rapidity, figure 24, with a slightly harder $p_\perp$ spectrum and, consequently, a slightly more central y one in the new models.

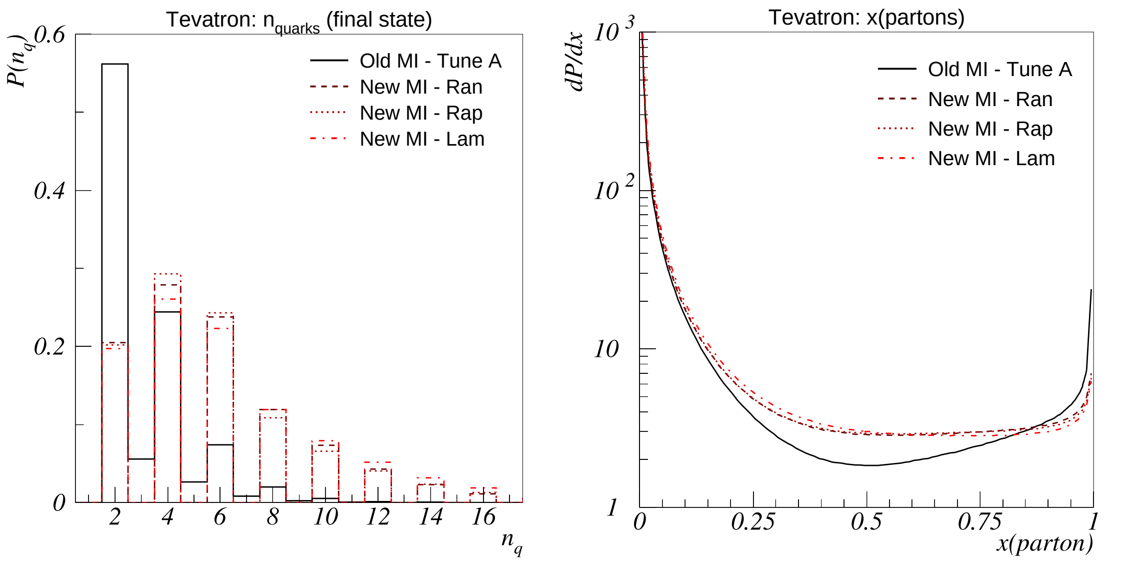
Having thus convinced ourselves that the overall features of the models agree, we turn to the aspects in which the new and old scenarios are expected to differ. One significant change is the possibility to knock out several valence quarks from the beam hadron. To quantify, the number of quarks (excluding diquarks) in the final state is illustrated in

JHEP03(2004)053



(*a*) (*b*)

Figure 24: Charged particle $p_\perp$ and y spectra at the Tevatron. Results are shown for Tune A (solid), Ran (dashed), Rap (dotted), and Lam (dash-dotted), as defined in table 1.



(*a*) (*b*)

Figure 25: (*a*) Number of final-state quarks (not counting diquarks) and (*b*) final-state parton x values at the Tevatron. Results are shown for Tune A (solid), Ran (dashed), Rap (dotted), and Lam (dash-dotted), as defined in table 1.

figure 25*a*. As can readily be observed, final-state quarks are much more abundant in the new scenarios. Valence quark interactions are, however, not the only cause of this. Diquarks are not included in figure 25*a*, and (as discussed in section 4.2) these are less frequently formed in the new scenarios, hence more of the quark content here appears as individual quarks in the final state. Further illustration of this is given by figure 25*b*, where

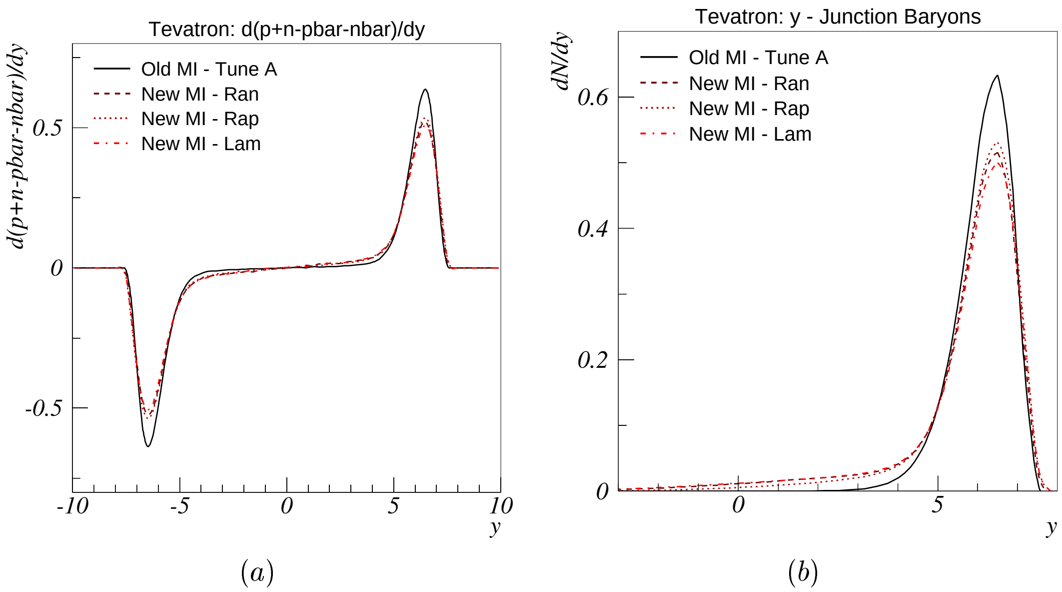


Figure 26: (*a*) $n(\mathrm{p}) + n(\mathrm{n}) - n(\overline{\mathrm{p}}) - n(\overline{\mathrm{n}})$ and (*b*) junction baryon rapidity distributions at the Tevatron. Results are shown for Tune A (solid), Ran (dashed), Rap (dotted), and Lam (dash-dotted), as defined in table 1.

the $x = 2E/\sqrt{s}$ values of all final-state partons, including diquarks, are shown. The peak towards low values comes mainly from gluons and is the same for Tune A and the new models. However, some of the content of the $x = 1$ peak in Tune A has vanished in the new models and has been replaced by a larger plateau at intermediate x, since a fraction of large-x diquarks has been split up into individual valence-like quarks.

Another point where the models differ is in the treatment of the beam remnant, especially concerning the flow of baryon number. Figure 26*a* shows the distribution of baryons minus the distribution of antibaryons as a function of rapidity, illustrating one way of experimentally probing the location of the beam baryon numbers in the final state. As expected, the distribution is more peaked in the old scenario, where the beam baryon number is 'locked' inside the escaping remnant diquark. Further illustration of the baryon number stopping is given by figure 26*b*, which shows the rapidity distribution of the final-state baryon carrying the baryon number of the incoming proton. Note especially the long tail in the new models, even extending to negative rapidities. The height of this tail depends on a set of model parameters, such as the choice of initial- and final-state colour correlations and the rules for diquark formation, cf. figure 11 fore more extreme scenarios.

The fact that the baryon number is not so closely associated with the valence quarks in the new model has an interesting side effect: apart from migrating in physical space, the initial-state baryon number can also 'migrate' in flavour to a much larger extent than before. When the junction is resolved, the flavour selection rules of ordinary string fragmentation take over in determining the flavour composition of the junction baryon. Thereby, it becomes possible for the junction baryon to have a larger strangeness number S than before, as shown in figure 27. The last bin, $|S| = 3$, is actually empty for Tune A, since in the old model it is impossible to produce an Ω^- from the incoming beam baryon number.

JHEP03(2004)053

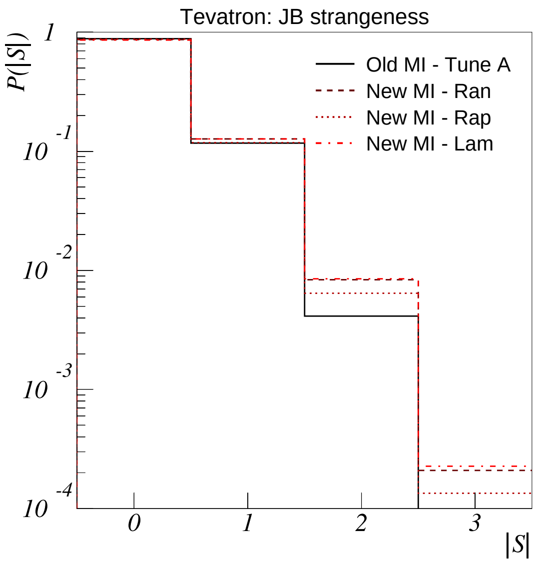


Figure 27: Junction baryon strangeness. Results are shown for Tune A (solid), Ran (dashed), Rap (dotted), and Lam (dash-dotted), as defined in table 1.

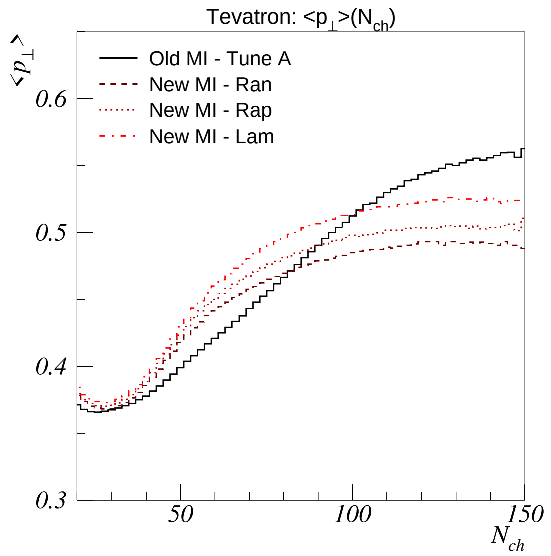


Figure 28: Average $p_\perp$ as a function of charged multiplicity for min-bias collisions at the Tevatron. Results are shown for Tune A (solid), Ran (dashed), Rap (dotted), and Lam (dash-dotted), as defined in table 1.

We note that heavy-ion experiments do observe a ratio $\Omega^-/\Omega^+ > 1$ [83], which is in contradiction with the old string model but which would be more in line with expectations based on the junction scenario introduced here.

Finally, figure 28 shows the $\langle p_\perp \rangle(n_{\mathrm{ch}})$ distributions. As previously noted, figure 21, this distribution is very sensitive to the colour correlations present at the hadronization stage. Since this is one of the major open issues remaining, it is not surprising that the agreement here is far from perfect: the new models exhibit a too early rise to a too low plateau, as compared to Tune A. While it is possible to obtain a much better agreement by varying the scheme adopted for the colour reconnections in the final state, our attempts in this direction have so far led to poorer descriptions of other distributions, the charged multiplicity distribution in particular. Moreover, the colour reconnection scheme adopted here is meant only as an instructive example, not as a model of the physics taking place. Our plan is to continue the study of colour correlations in more depth. In the context of such studies, it is encouraging to note that the acute sensitivity of the $\langle p_\perp \rangle(n_{\mathrm{ch}})$ distribution to these aspects makes its proper description a prime test for any physical model of the colour flow.

6.2 Comparisons at the LHC

Turning now to the situation at the LHC, the lack of experimental constraints increases the uncertainties. We here focus on only a subset of these, assuming the same energy scaling of the $p_{\perp 0}$ cutoff as for Tune A, proportional to $E_{\mathrm{cm}}^{0.25}$, and using the same parton distributions in all cases. We note that these aspects do constitute important sources

JHEP03(2004)053

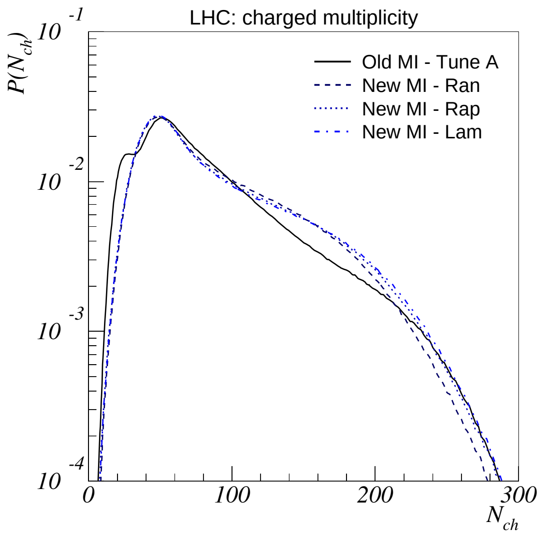


Figure 29: Multiplicity distributions for the LHC as obtained with Tune A (solid), and the Ran (dashed), Rap (dotted), and Lam (dash-dotted) models defined in table 1. For Tune A, the average charged multiplicity is $\langle n_{\mathrm{ch}}\rangle = 81$, whereas for the new models it is in the range 87-89.

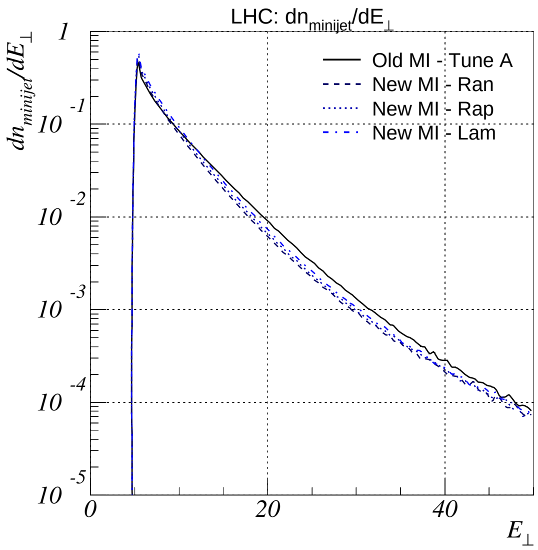


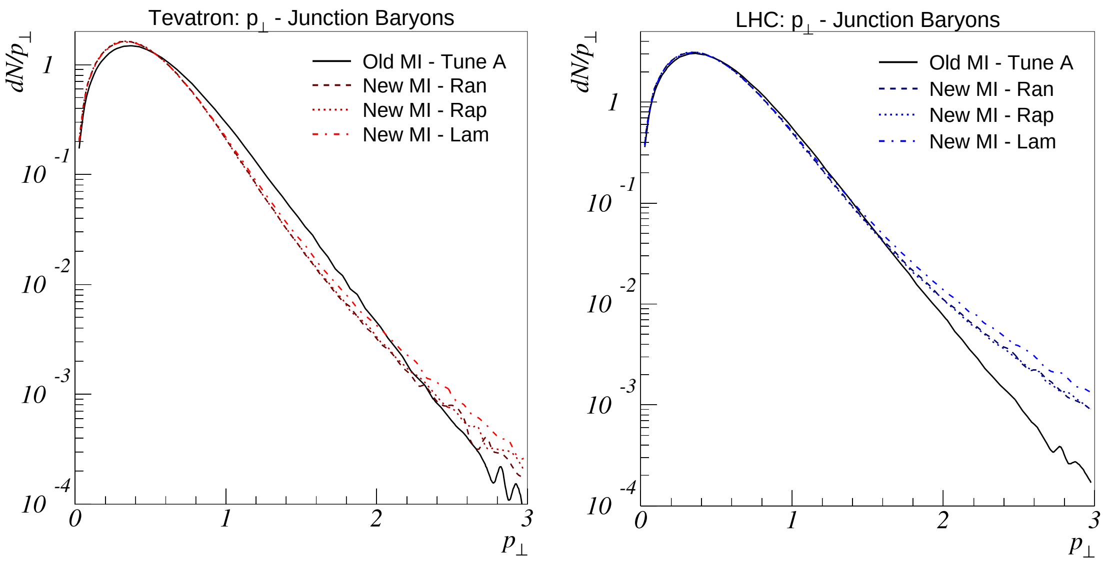
Figure 30: The number of jets as a function of jet $E_{\perp}$ (for $E_{\perp} > 5\,\mathrm{GeV}$) in min-bias collisions at the LHC. Results are shown for Tune A (solid), Ran (dashed), Rap (dotted), and Lam (dash-dotted), as defined in table 1.

of uncertainty in our ability to make trustworthy 'forecasts' for the LHC. Here, however, our aim is merely to compare alternative scenarios under identical boundary conditions.

Figure 29 shows the LHC charged multiplicity distribution for the same models as used for the Tevatron, but with the $p_{\perp 0}$ cutoff scaled to the LHC CM energy. The new models exhibit average multiplicities of 6-8 more charged particles per event than the Tune A value, $\langle n_{\mathrm{ch}}\rangle = 81$. This illustrates a general effect in the new models, which is due to the increased shower activity arising from associating also the sub-leading interactions with initial- and final-state cascades. The larger available phase space at higher energies implies that showers are more important at LHC, cf. the average number of final-state partons, $\langle n_{\mathrm{part}}\rangle$, in table 1. All else being equal this causes the multiplicity to increase more rapidly with energy than in the old model. (The strange bump on the Tune A distribution at low multiplicities is merely an artifact of the way parton distributions at low Q^2 are handled in that model.)

The addition of parton showers also increases the total amount of partonic transverse energy, but owing to a partial cancellation of the effects of radiation in the initial state (boosting some partons to larger $p_{\perp}$) and in the final state (jet broadening), the jet rates come out similar, as depicted in figure 30 (adopting the same cone algorithm and $|\eta| < 2.5$ region as before).

Figure 31 compares the junction baryon $p_{\perp}$ distributions at the Tevatron (left plot) and at the LHC (right plot), for Tune A and the new models. An interesting difference is



(a) (b)

Figure 31: Junction baryon $p_\perp$ spectrum at (a) the Tevatron and (b) the LHC. Results are shown for Tune A (solid), Ran (dashed), Rap (dotted), and Lam (dash-dotted), as defined in table 1.

that the junction baryon can be significantly harder in $p_\perp$ at the LHC than at the Tevatron in the new models, whereas Tune A exhibits spectra which are almost identical between the two energies. This is due to the intrinsic difference between the way primordial $k_\perp$ is treated in the two frameworks. In the old model, the width of the primordial $k_\perp$ distribution for the parton initiating the hardest scattering is fixed, to 1 GeV by default, hence there is no mechanism that would allow the junction baryon spectrum to depend on the CM energy (at sufficiently high energies that energy-momentum conservation effects can be neglected). In the new model, the amount of primordial $k_\perp$ given to initiators depends on the Q^2 of their associated hard scattering. With the increased phase space at the LHC, more primordial $k_\perp$ is thus imparted by recoil effects to the junction baryon than at the Tevatron, hence the $p_\perp$ spectrum becomes harder.

Also the junction baryon longitudinal migration shows some difference. Comparing the Tevatron junction baryon rapidity distribution, figure 26b above, with the LHC one, figure 32, we may distinguish two components. One is the peak at large rapidities, which corresponds to an (effective) diquark fragmentation and which is only shifted outwards in rapidity relative to the Tevatron by the increased energy. The other is the tail to central rapdities, which corresponds to baryon stopping. This tail *does* increase with energy, following the increase in the average number of interactions.

Finally, we show the $\langle p_\perp \rangle (n_{\mathrm{ch}})$ distributions in figure 33. The same qualitative behaviour as at the Tevatron is apparent: the new models exhibit an earlier rise to a lower plateau, as compared to Tune A. Again, it is premature to draw any strong conclusions, in view of the still simple-minded description of the colour flow that we have included here. Further and more detailed studies of possible colour correlation mechanisms in hadronic collisions will be required in order to fully understand these aspects.

JHEP03(2004)053

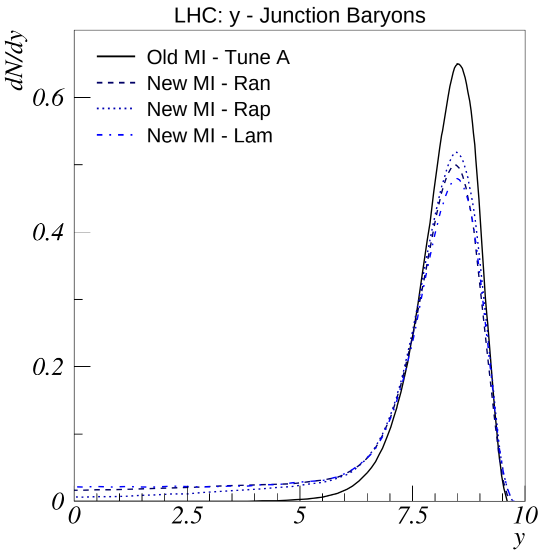


Figure 32: Junction baryon rapidity distributions at the LHC. Note: at the LHC *both* beam baryon numbers are included in the figure, whereas in the Tevatron plots, figure 26, the antibaryon number is not. Results are shown for Tune A (solid), Ran (dashed), Rap (dotted), and Lam (dash-dotted), as defined in table 1.

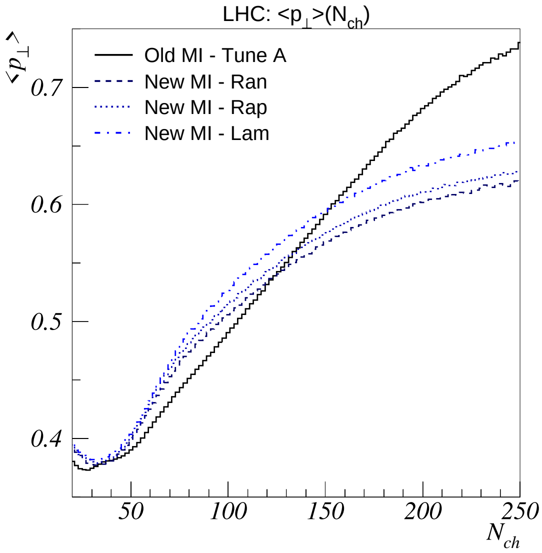


Figure 33: Average $p_\perp$ as a function of charged multiplicity for min-bias collisions at the LHC. Results are shown for Tune A (solid), Ran (dashed), Rap (dotted), and Lam (dash-dotted), as defined in table 1.

7. Conclusion and outlook

Only in the last few years have multiple interactions gone from being a scientific curiosity, by most assumed relevant only for some rare topologies of four-jet events, to being accepted as the key element for understanding the structure of underlying events. However, this leaves a lot of questions to be addressed, such as:

1. What is the detailed mechanism and functional form of the dampening of the perturbative cross section at small $p_\perp$?
2. What is the energy dependence of the mechanism(s) involved?
3. How is the internal structure of the proton reflected in an impact-parameter-dependent multiple interactions rate, as manifested e.g. in jet pedestal effects?
4. How can the set of colliding partons from a hadron be described in terms of correlated multiparton distribution functions of flavours and longitudinal momenta?
5. How does a set of initial partons at some low perturbative cutoff scale evolve into such a set of colliding partons? Is standard DGLAP evolution sufficient, or must BFKL/CCFM effects be taken into account?
6. How would the set of initiators correlate with the flavour content of, and the longitudinal momentum sharing inside, the left-behind beam remnant?

7. How are the initiator and remnant partons correlated by confinement effects, e.g. in primordial $k_\perp$?

8. How are all produced partons, both the interacting and the beam-remnant ones, correlated in colour? Is the large number-of-colours limit relevant, wherein partons can be hooked up into strings representing a linear confinement force?

9. How is the original baryon number of an incoming proton reflected in the colour topology?

10. To what extent would a framework with independently fragmenting string systems, as defined from the colour topology, be modified by the space-time overlap of several strings?

Tentative answers to some of the questions are provided by the Tune A of the Pythia multiple interactions framework. Thus we now believe that:

- The matter overlap when two hadrons collide can be described by an impact-parameter dependence more spiked than a gaussian but less so than an exponential.
- The $p_{\perp 0}$ regularization scale does increase with energy.
- The colours of final-state partons are not random but correlated, somehow, to give a reduced string length.

This still leaves many questions unanswered. Worse, existing event generators would not even address many of the relevant issues, at least not in a deliberate or realistic fashion. In this article we have therefore tried to take the next step towards a better understanding of the structure of a hadronic event, addressing several of the points above. This in particular has concerned the correlations between initiator and remnant partons in the hadron beams, in terms of flavour, longitudinal and transverse momenta. Colour correlations have also been studied, and here it appears that the final-state partons need be involved as well. The complexity of the colour issues is tremendous, however, and we do not consider the studies finished in this area.

A specific new topic addressed is that of baryon number flow. Data from hadronic collisions, and even more from heavy-ion ones, show large excesses of baryon over antibaryon production in the central rapidity region of events [84], suggesting a significant influx of baryon number from the high-rapidity colliding beams, more than would be expected from standard quark/diquark fragmentation models. When the junction is introduced as a topological feature of the colour field in the baryon, however, the fate of the baryon number of an incoming beam particle may partly or wholly decouple from that of the valence quarks [1, 85]. We have here demonstrated that the junction topology in combination with multiple interactions can induce quite large rapidity shifts, of the desired kind.

The problem may actually be the opposite, i.e. not to move the baryon number by too much. To this end, we have assumed a suppression of interactions that affect several of the three colour chains that connect the valence quarks to the junction. This could be an impact-parameter-related effect, that not the whole proton is involved in the hard processes. If so, the suppression should be less pronounced in heavy-ion collisions, where the interactions of a proton with several nucleons in the other nucleus could occur at different positions in the transverse plane and thereby affect different chains. Obviously it would be a major undertaking to construct a complete model for heavy-ion collisions to study these ideas, but we hope in the future to be able to present a simple study of the baryon number flow.

Another open issue is that of intertwined initial-state showers, whereby two seemingly unrelated partons, each undergoing a hard scattering, reconstruct back to come from a common shower ancestor. With the new $p_\perp$-ordered showers now being implemented in Pythia [86] we intend to introduce enough flexibility that such issues could be addressed.

This will also further constrain the initial-state colour flow. The possibility of final-state colour reconnections remains, however, and has been proposed as a mechanism to introduce diffractive topologies in a number of processes [87]. One here needs to better understand how much reconnections are allowed/required, and of what character.

We see that much work remains, before the physics of the underlying event is truly understood. Progress will not be possible without a constructive dialogue between theory and experiment. We have frequently had reason to mention Tune A as a role model here, because it offers a convenient reference that more sophisticated models can be tested against, without the need to know the details of the CDF detector. However, only a few distributions went into the tune, and so we do not know what to aim for in many other respects.

To give one specific example, it would be valuable to have information on the 'lumpiness' of the underlying event, such as n-jet rates as a function of some jet resolution parameter, similarly to e^+e^--annihilation QCD analyses. One would there hope for an intermediate resolution region, between the coarse one that is dominated by the perturbative QCD structure and the fine one that mainly is sensitive to hadronization details, where the structure of the multiple interactions would play a key role. An understanding of this lumpiness is related to the fluctuations in the jet pedestal, and thereby to the smearing of jet energies in SUSY searches, say. It all hangs together ...

In summary, striving for a better understanding of the physics of the underlying event is both interesting and useful. Interesting because it forces us to consider many issues normally swept under the carpet, and to confront dramatically different scenarios. Useful because it ties in with so many other physics analyses at hadron colliders. So there is plenty of interesting and useful work ahead of us before the picture has clarified completely!

Acknowledgments

The authors gratefully acknowledge the stimulating atmosphere at the Les Houches 2003 Workshop on Physics at TeV Colliders, the CERN Workshop on Monte Carlo tools for the LHC, and the Collider Physics 2004 Workshop at KITP, UCSB. We have benefitted from discussions with R.D. Field, J. Huston, A. Moraes, and many more. We are also grateful to the NorduGRID project, for use of computing resources.

This research was supported in part by the National Science Foundation under Grant No. PHY99-07949, and by The Royal Physiographic Society in Lund.

References

[1] T. Sjöstrand and M. van Zijl, *A multiple interaction model for the event structure in hadron collisions*, *Phys. Rev.* **D 36** (1987) 2019.

[2] C. Seez, CMS TN/93–115;
CMS collaboration, *Technical proposal*, CERN/LHCC 94–38, p. 50;
CMS collaboration, *ECAL technical design report*, CERN/LHCC 97–33, p. 330;
U. Egede, *The search for a standard model Higgs at the LHC and electron identification using transition radiation in the ATLAS tracker*, Ph.D. thesis, Lund University LUNFD6/(NFFL–7150) 1997;
ATLAS collaboration, *Detector and physics performance technical design report*, vol. I, CERN/LHCC 99–14, p. 228.

[3] T. Sjöstrand and P.Z. Skands, *Baryon number violation and string topologies*, *Nucl. Phys.* **B 659** (2003) 243 [hep-ph/0212264].

[4] B. Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand, *Parton fragmentation and string dynamics*, *Phys. Rept.* **97** (1983) 31;
B. Andersson, *The Lund model*, Cambridge University Press, Cambridge 1998.

[5] P. Skands and T. Sjöstrand, *Progress on multiple interactions: modelling the underlying event in hadron hadron collisions*, in the proceedings of *International europhysics conference on high-energy physics (HEP 2003)*, Aachen, Germany, 17–23 Jul 2003, Eur. Phys. J. C Direct, DOI: 10.1140/epjcd/s2003-03-520-7 [hep-ph/0310315];
T. Sjöstrand and P. Skands, *Multiple interactions and beam remnants*, hep-ph/0401060 to appear in proceedings of *3rd Les Houches workshop: physics at TeV colliders*, Les Houches, France, 26 May - 6 Jun 2003.

[6] CTEQ collaboration, H.L. Lai et al., *Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions*, *Eur. Phys. J.* **C 12** (2000) 375 [hep-ph/9903282].

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

[8] J. Dischler and T. Sjöstrand, *QCD interconnection studies at linear colliders*, *Eur. Phys. J. Direct.* **C 2** (2001) 1.

[9] T. Sjöstrand, P. Edén, C. Friberg, L. Lönnblad, G. Miu, S. Mrenna and E. Norrbin, *High-energy physics event generation with Pythia 6.1*, *Comput. Phys. Commun.* **135** (2001) 238 [hep-ph/001001];
T. Sjöstrand, L. Lönnblad, S. Mrenna and P. Skands, *Pythia 6.3 physics and manual*, hep-ph/0308153.

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

[11] V.V. Sudakov, *Vertex parts at very high-energies in quantum electrodynamics*, *Zh. Eksp. Teor. Fiz.* **30** (1956) 87.

[12] G.A. Schuler and T. Sjöstrand, *Towards a complete description of high-energy photoproduction*, *Nucl. Phys.* **B 407** (1993) 539; *A scenario for high-energy $\gamma\gamma$ interactions*, *Z. Physik* **C 73** (1997) 677 [hep-ph/9605240];
C. Friberg and T. Sjöstrand, *Total cross-sections and event properties from real to virtual photons*, *J. High Energy Phys.* **09** (2000) 10.

JHEP03(2004)053

[13] H.-M. Chan et al., *Diffractive scattering in the dual model*, *Nucl. Phys.* **B 86** (1975) 479;
H.-M. Chan, J.E. Paton, S.T. Tsou and S.W. Ng, *Regge parameters from duality and unitarity*, *Nucl. Phys.* **B 92** (1975) 13;
G.F. Chew and C. Rosenzweig, *Dual topological unitarization: an ordered approach to hadron theory*, *Phys. Rept.* **41** (1978) 263.

[14] A. Capella, U. Sukhatme, C.-I. Tan and J. Tran Thanh Van, *Jets in small p_T hadronic collisions, universality of quark fragmentation and rising rapidity plateaus*, *Phys. Lett.* **B 81** (1979) 68; *Dual parton model*, *Phys. Rept.* **236** (1994) 225;
H. Minakata, *Universal quark-jet fragmentation in soft hadronic reactions*, *Phys. Rev.* **D 20** (1979) 1656;
G. Cohen-Tannoudji, A. El Hassouni, J. Kalinowski, O. Napoly and R. Peschanski, *Partons at low p_T*, *Phys. Rev.* **D 21** (1980) 2699;
A. Capella and J. Tran Thanh Van, *Hadron-nucleus interactions and the leading particle effect in a dual parton model*, *Z. Physik* **C 10** (1981) 249;
A. Capella and J. Tran Thanh Van, *Multiplicity distributions up to $\sqrt{s} = 540\,GeV$ in the dual parton model*, *Phys. Lett.* **B 114** (1982) 450;
K. Fialkowski and A. Kotanski, *Hadron multiplicity distributions in a dual model*, *Phys. Lett.* **B 107** (1981) 132;
P. Aurenche and F.W. Bopp, *Rapidity spectra in proton proton and proton - anti-proton scattering up to 540 GeV in a dual parton model*, *Phys. Lett.* **B 114** (1982) 363;
A.B. Kaidalov, *The quark - gluon structure of the pomeron and the rise of inclusive spectra at high-energies*, *Phys. Lett.* **B 116** (1982) 459;
K. Fialkowski and A. Kotanski, *Forward-backward correlations in pp interactions in a dual model*, *Phys. Lett.* **B 115** (1982) 425;
A.B. Kaidalov and K.A. Ter-Martirosian, *Pomeron as quark-gluon strings and multiple hadron production at SPS collider energies*, *Phys. Lett.* **B 117** (1982) 247;
P. Aurenche, F.W. Bopp and J. Ranft, *Particle production in hadron hadron collisions at collider energies in an exclusive multistring fragmentation model*, *Z. Physik* **C 23** (1984) 67;
J. Ranft, P. Aurenche and F.W. Bopp, *The multistring fragmentation model and violation of KNO scaling*, *Z. Physik* **C 26** (1984) 279; F.W. Bopp, P. Aurenche and J. Ranft, *The transverse momenta in minimum bias events at collider energies*, *Phys. Rev.* **D 33** (1986) 1867;
V.M. Chudakov and V.V. Lugovoi, *Dual parton model of inelastic pp interactions and cascade breaking of string*, *Z. Physik* **C 59** (1993) 511.

[15] L.V. Gribov, E.M. Levin and M.G. Ryskin, *Semihard processes in QCD*, *Phys. Rept.* **100** (1983) 1;
E.M. Levin and M.G. Ryskin, *High-energy hadron collisions in QCD*, *Phys. Rept.* **189** (1990) 267.

[16] V.N. Gribov, *A reggeon diagram technique*, *Sov. Phys. JETP* **26** (1968) 414;
V.A. Abramovski, O.V. Kancheli and V.N. Gribov, *Character of inclusive spectra and fluctuations produced in inelastic processes by multi-pomeron exchange*, *Sov. J. Nucl. Phys.* **18** (1974) 308;
G. Veneziano, *Regge intercepts and unitarity in planar dual models*, *Nucl. Phys.* **B 74** (1974) 365; *Some aspects of a unified approach to gauge, dual and Gribov theories*, *Nucl. Phys.* **B 117** (1976) 519;
J. Bartels and M.G. Ryskin, *The space-time picture of the wee partons and the AGK cutting rules in perturbative QCD*, *Z. Physik* **C 76** (1997) 241 [hep-ph/9612226].

[17] H. Baer, F.E. Paige, S.D. Protopopescu and X. Tata, *Isajet 7.48: a Monte Carlo event generator for pp, $\bar{p}p$ and e^+e^- reactions*, hep-ph/0001086.

[18] P. Aurenche et al., *Multiparticle production in a two component dual parton model*, *Phys. Rev.* **D 45** (1992) 92;
P. Aurenche, F.W. Bopp, R. Engel, D. Pertermann, J. Ranft and S. Roesler, *DTUJET-93: sampling inelastic proton proton and anti-proton-proton collisions according to the two component dual parton model*, *Comput. Phys. Commun.* **83** (1994) 107.

[19] R. Engel, *Photoproduction within the two component dual parton model, 1. Amplitudes and cross-sections*, *Z. Physik* **C 66** (1995) 203;
R. Engel and J. Ranft, *Hadronic photon-photon interactions at high energies*, *Phys. Rev.* **D 54** (1996) 4244 [hep-ph/9509373].

[20] J. Ranft, *The dual parton model at cosmic ray energies*, *Phys. Rev.* **D 51** (1995) 64;
S. Roesler, R. Engel and J. Ranft, *The Monte Carlo event generator DPMJET-III*, hep-ph/0012252.

[21] P.V. Landshoff and J.C. Polkinghorne, *Calorimeter triggers for hard collisions*, *Phys. Rev.* **D 18** (1978) 3344;
C. Goebel, F. Halzen and D.M. Scott, *Double Drell-Yan annihilations in hadron collisions: novel tests of the constituent picture*, *Phys. Rev.* **D 22** (1980) 2789;
M.L. Mangano, *Four jet production at the Tevatron collider*, *Z. Physik* **C 42** (1989) 331.

[22] F. Takagi, *Multiple production of quark jets off nuclei*, *Phys. Rev. Lett.* **43** (1979) 1296;
N. Paver and D. Treleani, *Multi-quark scattering and large p_T jet production in hadronic collisions*, *Nuovo Cim.* **A70** (1982) 215; *Multiple parton processes in the TeV region*, *Z. Physik* **C 28** (1985) 187; *Multiple parton interactions and multi-jet events at collider and Tevatron energies*, *Phys. Lett.* **B 146** (1984) 252;
B. Humpert, *Are there multi-quark interactions?*, *Phys. Lett.* **B 131** (1983) 461;
B. Humpert and R. Odorico, *Multiparton scattering and QCD radiation as sources of four jet events*, *Phys. Lett.* **B 154** (1985) 211;
G. Pancheri and Y.N. Srivastava, *Low p_T jets and the rise with energy of the inelastic cross-section*, *Phys. Lett.* **B 182** (1986) 199;
N. Brown, *Mini-jets, unitarity and multiple parton scattering*, *Mod. Phys. Lett.* **A 4** (1989) 2447.

[23] R.E. Ecclestone and D.M. Scott, *Hadroproduction of $\psi\psi$*, *Z. Physik* **C 19** (1983) 29;
B. Humpert, *The production of gauge boson pairs by $p\bar{p}$ colliders*, *Phys. Lett.* **B 135** (1984) 179;
M. Mekhfi, *Multiparton processes: an application to double Drell-Yan*, *Phys. Rev.* **D 32** (1985) 2371;
F. Halzen, P. Hoyer and W.J. Stirling, *Evidence for multiple parton interactions from the observation of multi-muon events in Drell-Yan experiments*, *Phys. Lett.* **B 188** (1987) 375;
R.W. Robinett, *$\psi\psi$ production and multiple parton interactions at supercollider energies*, *Phys. Lett.* **B 230** (1989) 153;
R.M. Godbole, S. Gupta and J. Lindfors, *Double parton scattering contribution to W + jets*, *Z. Physik* **C 47** (1990) 69;
M. Drees and T. Han, *Signals for double parton scattering at the Fermilab Tevatron*, *Phys. Rev. Lett.* **77** (1996) 4142 [hep-ph/9605430];
A. Del Fabbro and D. Treleani, *A double parton scattering background to Higgs boson production at the LHC*, *Phys. Rev.* **D 61** (2000) 077502 [hep-ph/9911358];

JHEP03(2004)053

A. Kulesza and W.J. Stirling, *Like-sign W boson production at the LHC as a probe of double parton scattering*, *Phys. Lett.* **B 475** (2000) 168 [hep-ph/9912232].

[24] T.T. Chou and C.N. Yang, *Model of elastic high-energy scattering*, *Phys. Rev.* **170** (1968) 1591; *Elastic hadron hadron scattering at ultrahigh-energies and existence of many dips*, *Phys. Rev.* **D 19** (1979) 3268;
C. Bourrely, J. Soffer and T.T. Wu, *Impact picture expectations for very high-energy elastic pp and $p\bar{p}$ scattering*, *Nucl. Phys.* **B 247** (1984) 15; *Impact-picture phenomenology for $\pi^{\pm}p$, $K^{\pm}p$ and pp, $\bar{p}p$ elastic scattering at high energies*, *Eur. Phys. J.* **C 28** (2003) 97 [hep-ph/0210264];
P. L'Heureux, B. Margolis and P. Valin, *Quark-gluon model for diffraction at high-energies*, *Phys. Rev.* **D 32** (1985) 1681.

[25] L. Durand and P. Hong, *QCD and rising total cross-sections*, *Phys. Rev. Lett.* **58** (1987) 303;
A. Capella, J. Tran Thanh Van and J. Kwiecinski, *Mini-jets, QCD and unitarity*, *Phys. Rev. Lett.* **58** (1987) 2015;
M.M. Block, F. Halzen and B. Margolis, *How large is the total cross-section at supercollider energies?*, *Phys. Rev.* **D 45** (1992) 839;
R.M. Godbole and G. Pancheri, *Hadronic cross-sections in $\gamma\gamma$ processes and the next linear collider*, *Eur. Phys. J.* **C 19** (2001) 129 [hep-ph/0010104].

[26] J. Kuti and V.F. Weisskopf, *Inelastic lepton-nucleon scattering and lepton pair production in the relativistic quark parton model*, *Phys. Rev.* **D 4** (1971) 3418;
K. Konishi, A. Ukawa and G. Veneziano, *Jet calculus: a simple algorithm for resolving QCD jets*, *Nucl. Phys.* **B 157** (1979) 45;
R. Kirschner, *Generalized Lipatov-Altarelli-Parisi equations and jet calculus rules*, *Phys. Lett.* **B 84** (1979) 266;
H.R. Gerhold, *On the generalized Kuti-Weisskopf model for parton distributions*, *Nuovo Cim.* **59A** (1980) 373;
V.P. Shelest, A.M. Snigirev and G.M. Zinovev, *The multiparton distribution equations in QCD*, *Phys. Lett.* **B 113** (1982) 325;
A.M. Snigirev, *Double parton distributions in the leading logarithm approximation of perturbative QCD*, *Phys. Rev.* **D 68** (2003) 114012 [hep-ph/0304172].

[27] M. Mekhfi, *Correlations in color and spin in multiparton processes*, *Phys. Rev.* **D 32** (1985) 2380;
M. Mekhfi and X. Artru, *Sudakov suppression of color correlations in multiparton scattering*, *Phys. Rev.* **D 37** (1988) 2618.

[28] A. Bialas and E. Bialas, *Impact parameter analysis of multiplicity distribution in high-energy pp collisions*, *Acta Phys. Polon.* **B5** (1974) 373;
T.T. Chou and C.N. Yang, *Geometrical model of multiparticle production in hadron-hadron collisions*, *Phys. Rev.* **D 32** (1985) 1692;
S. Barshay, *Strong forward-backward multiplicity correlations in hadron-hadron collisions and their absence in electron-positron annihilation*, *Z. Physik* **C 32** (1986) 513;
W.R. Chen and R.C. Hwa, *Branching, geometrical scaling and KNO scaling*, *Phys. Rev.* **D 36** (1987) 760.

[29] A.H. Mueller and J.-w. Qiu, *Gluon recombination and shadowing at small values of x*, *Nucl. Phys.* **B 268** (1986) 427.

JHEP03(2004)053

[30] G. Corcella et al., *HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *J. High Energy Phys.* **01** (2001) 010 [hep-ph/0011363]; *Herwig 6.5 release note*, hep-ph/0210213.

[31] UA5 collaboration, G.J. Alner et al., *The UA5 high-energy $\bar{p}p$ simulation program*, *Nucl. Phys.* **B 291** (1987) 445.

[32] J.M. Butterworth, J.R. Forshaw and M.H. Seymour, *Multiparton interactions in photoproduction at HERA*, *Z. Physik* **C 72** (1996) 637 [hep-ph/9601371].

[33] I. Borozan and M.H. Seymour, *An eikonal model for multiparticle production in hadron hadron interactions*, *J. High Energy Phys.* **09** (2002) 015 [hep-ph/0207283].

[34] E.A. Kuraev, L.N. Lipatov and V.S. Fadin, *The Pomeranchuk singularity in nonabelian gauge theories*, *Sov. Phys. JETP* **45** (1977) 199;
I.I. Balitsky and L.N. Lipatov, *The Pomeranchuk singularity in quantum chromodynamics*, *Sov. J. Nucl. Phys.* **28** (1978) 822.

[35] M. Ciafaloni, *Coherence effects in initial jets at small Q^2/s*, *Nucl. Phys.* **B 296** (1987) 49;
S. Catani, F. Fiorani and G. Marchesini, *Small x behavior of initial state radiation in perturbative QCD*, *Nucl. Phys.* **B 336** (1990) 18.

[36] B. Andersson, G. Gustafson and J. Samuelsson, *The linked dipole chain model for DIS*, *Nucl. Phys.* **B 467** (1996) 443;
B. Andersson, G. Gustafson and H. Kharraziha, *Investigations into the BFKL mechanism with a running QCD coupling*, *Phys. Rev.* **D 57** (1998) 5543 [hep-ph/9711403];
H. Kharraziha and L. Lönnblad, *The linked dipole chain Monte Carlo*, *J. High Energy Phys.* **03** (1998) 006 [hep-ph/9709424].

[37] G. Gustafson, L. Lönnblad and G. Miu, *Gluon distribution functions in the k_T-factorization approach*, *J. High Energy Phys.* **09** (2002) 005 [hep-ph/0206195].

[38] G. Gustafson, L. Lönnblad and G. Miu, *Hadronic collisions in the linked dipole chain model*, *Phys. Rev.* **D 67** (2003) 034020 [hep-ph/0209186].

[39] K. Kajantie, P.V. Landshoff and J. Lindfors, *Minijet production in high-energy nucleus-nucleus collisions*, *Phys. Rev. Lett.* **59** (1987) 2527;
K.J. Eskola, K. Kajantie and J. Lindfors, *Quark and gluon production in high-energy nucleus-nucleus collisions*, *Nucl. Phys.* **B 323** (1989) 37;
K.J. Eskola, K. Kajantie and K. Tuominen, *Heavy ion collision multiplicities and gluon distribution functions*, *Nucl. Phys.* **A 700** (2002) 509 [hep-ph/0106330].

[40] Axial Field Spectrometer collaboration, T. Åkesson et al., *Double parton scattering in pp collisions at $\sqrt{s} = 63$ GeV*, *Z. Physik* **C 34** (1987) 163.

[41] UA2 collaboration, J. Alitti et al., *A study of multi-jet events at the CERN $\bar{p}p$ collider and a search for double parton scattering*, *Phys. Lett.* **B 268** (1991) 145.

[42] CDF collaboration, F. Abe et al., *Study of four jet events and evidence for double parton interactions in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, *Phys. Rev.* **D 47** (1993) 4857.

[43] CDF collaboration, F. Abe et al., *Measurement of double parton scattering in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV*, *Phys. Rev. Lett.* **79** (1997) 584;
CDF collaboration, F. Abe et al., *Double parton scattering in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV*, *Phys. Rev.* **D 56** (1997) 3811.

JHEP03(2004)053

[44] ZEUS collaboration, C. Gwenlan, *Multijets in photoproduction at HERA*, *Acta Phys. Polon.* **B33** (2002) 3123.

[45] D0 collaboration, V.M. Abazov et al., *Multiple jet production at low transverse energies in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, *Phys. Rev.* **D 67** (2003) 052001 [hep-ex/0207046].

[46] UA1 collaboration, C.-E. Wulz, *Low transverse energy jets in $p\bar{p}$ collisions at $\sqrt{s}$ in the 0.2 TeV–0.9 TeV range and their interpretation in terms of QCD*, in proceedings of the *22nd Rencontres de Moriond*, Les Arcs, France, 15–21 March 1987.

[47] Z. Koba, H.B. Nielsen and P. Olesen, *Scaling of multiplicity distributions in high-energy hadron collisions*, *Nucl. Phys.* **B 40** (1972) 317.

[48] UA5 collaboration, G.J. Alner et al., *Scaling violation favoring high multiplicity events at 540 GeV CMS energy*, *Phys. Lett.* **B 138** (1984) 304;
UA5 collaboration, R.E. Ansorge et al., *Charged particle multiplicity distributions at 200 GeV and 900 GeV center-of-mass energy*, *Z. Physik* **C 43** (1989) 357.

[49] E735 collaboration, T. Alexopoulos et al., *The role of double parton collisions in soft hadron interactions*, *Phys. Lett.* **B 435** (1998) 453.

[50] E.D. Malaza and B.R. Webber, *Multiplicity distributions in quark and gluon jets*, *Nucl. Phys.* **B 267** (1986) 702;
E.D. Malaza, *Multiplicity distributions in quark and gluon jets to $O(\alpha_s)$*, *Z. Physik* **C 31** (1986) 143.

[51] ZEUS collaboration, M. Derrick et al., *Neutral strange particle production in deep inelastic scattering at HERA*, *Z. Physik* **C 68** (1995) 29 [hep-ex/9505011];
H1 collaboration, S. Aid et al., *Strangeness in photoproduction and deep-inelastic scattering at HERA*, contribution 479 to EPS HEP95, Brussels;
T. Sjöstrand, *Soft photoproduction physics*, *J. Phys.* **G 22** (1996) 709;
H1 Collaboration, Abstract 1102, submitted to ICHEP02, Amsterdam.

[52] R608 collaboration, A.M. Smith et al., *Evidence for pomeron single quark interactions in proton diffraction at the ISR*, *Phys. Lett.* **B 163** (1985) 267.

[53] UA5 collaboration, R.E. Ansorge et al., *Charged particle correlations in $\bar{p}p$ collisions at c.m. energies of 200 GeV, 546 GeV and 900 GeV*, *Z. Physik* **C 37** (1988) 191;
E735 collaboration, T. Alexopoulos et al., *Charged particle multiplicity correlations in $p\bar{p}$ collisions at $\sqrt{s} = 0.3$ TeV to 1.8 TeV*, *Phys. Lett.* **B 353** (1995) 155.

[54] UA1 collaboration, C. Albajar et al., *Production of low transverse energy clusters in $\bar{p}p$ collisions at $\sqrt{s} = 0.2$ TeV to 0.9 TeV and their interpretation in terms of QCD jets*, *Nucl. Phys.* **B 309** (1988) 405.

[55] CDF collaboration, T. Affolder et al., *Charged jet evolution and the underlying event in proton anti-proton collisions at 1.8 TeV*, *Phys. Rev.* **D 65** (2002) 092002.

[56] CDF collaboration, R.D. Field, *The underlying event in hard scattering processes*, in the proceedings of *Snowmass 2001 eConf* **C010630** (2001) P501 [hep-ph/0201192], talks available from webpage `http://www.phys.ufl.edu/~rfield/cdf/`;
V. Tano, *The underlying event in hadron hadron collisions*, hep-ex/0205023 to appear in the proceedings of the *37th Rencontres de Moriond*, Les Arcs, France, 16–23 March 2002.

[57] H1 collaboration, S. Aid et al., *Jets and energy flow in photon-proton collisions at HERA*, *Z. Physik* **C 70** (1996) 17 [hep-ex/9511012].

[58] Ames-Bologna-CERN-Dortmund-Heidelberg-Warsaw collaboration, A. Breakstone et al., *Multiplicity dependence of transverse momentum spectra at ISR energies*, *Phys. Lett.* **B 132** (1983) 463;
UA1 collaboration, C. Albajar et al., *A study of the general characteristics of proton-anti-proton collisions at $\sqrt{s} = 0.2$ TeV to 0.9 TeV*, *Nucl. Phys.* **B 335** (1990) 261;
E735 collaboration, T. Alexopoulos et al., *Multiplicity dependence of transverse momentum spectra of centrally produced hadrons in $\bar{p}p$ collisions at 0.3 TeV, 0.54 TeV, 0.9 TeV and 1.8 TeV center-of-mass energy*, *Phys. Lett.* **B 336** (1994) 599;
CDF collaboration, D. Acosta et al., *Soft and hard interactions in $p\bar{p}$ collisions at $\sqrt{s} = 1800$ GeV and 630 GeV*, *Phys. Rev.* **D 65** (2002) 072005.

[59] UA1 collaboration, G. Arnison et al., *Hadronic jet production at the CERN proton-anti-proton collider*, *Phys. Lett.* **B 132** (1983) 214;
D0 collaboration, S. Abachi et al., *Transverse energy distributions within jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, *Phys. Lett.* **B 357** (1995) 500;
ZEUS collaboration, J. Breitweg et al., *Dijet cross sections in photoproduction at HERA*, *Eur. Phys. J.* **C 1** (1998) 109 [hep-ex/9710018].

[60] CDF collaboration, F. Abe et al., *Measurement of the B/D0 - anti-B/D0 flavor oscillation frequency and study of same side flavor tagging of B mesons in $p\bar{p}$ collisions*, *Phys. Rev.* **D 59** (1999) 032001 [hep-ex/9806026];
P. Bartalini et al., *Tuning of multiple interactions generated by Pythia*, in the proceedings of the *Workshop on standard model physics (and more) at the LHC*, CERN 2000-004, p. 293.

[61] R.D. Field, *Underlying-event studies/tunes for Run 2*, presentations at the *Matrix element and Monte Carlo tuning workshop*, Fermilab, 4 October 2002 and 29–30 April 2003, talks available from webpage `http://cepa.fnal.gov/CPD/MCTuning/`, and further recent talks available from `http://www.phys.ufl.edu/~rfield/cdf/`.

[62] A. Kharchilava, *D0 min bias comparisons*, presentation at the *Matrix Element and Monte Carlo Tuning Workshop*, Fermilab, 4 October 2002, talk available from webpage `http://cepa.fnal.gov/CPD/MCTuning/`.

[63] A. Moraes, C. Buttar, I. Dawson and P. Hodgson, ATLAS internal notes notes and talks available from webpage `http://amoraes.home.cern.ch/amoraes/`.

[64] J.M. Butterworth and S. Butterworth, *Jetweb: a WWW interface and database for Monte Carlo tuning and validation*, *Comput. Phys. Commun.* **153** (2003) 164 [hep-ph/0210404], see also `http://jetweb.hep.ucl.ac.uk`.

[65] P.V. Landshoff, *Pomeron physics: an update*, *Nucl. Phys.* **99A** *(Proc. Suppl.)* (2001) 311 [hep-ph/0010315].

[66] Yu.L. Dokshitzer, V.A. Khoze, A.H. Mueller and S.I. Troyan, *Basics of perturbative QCD*, Editions Frontières, Gif-sur-Yvette 1991.

[67] T. Sjöstrand, *A model for initial state parton showers*, *Phys. Lett.* **B 157** (1985) 321;
M. Bengtsson, T. Sjöstrand and M. van Zijl, *Initial state radiation effects on W and jet production*, *Z. Physik* **C 32** (1986) 67.

[68] C. Balázs, J. Huston and I. Puljak, *Higgs production: a comparison of parton showers and resummation*, *Phys. Rev.* **D 63** (2001) 014021 [hep-ph/0002032].

JHEP03(2004)053

[69] E. Thomé, *Perturbative and nonperturbative effects in transverse momentum generation*, hep-ph/0401121;
J. Huston, I. Puljak, T. Sjöstrand and E. Thomé, *Resummation and shower studies*, hep-ph/0401145.

[70] European Muon collaboration, M. Arneodo et al., *Jet production and fragmentation properties in deep inelastic muon scattering*, *Z. Physik* **C 36** (1987) 527;
L. Apanasevich et al., *k_T effects in direct photon production*, *Phys. Rev.* **D 59** (1999) 074007 [hep-ph/9808467].

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

[72] B. Andersson, G. Gustafson and T. Sjöstrand, *How to find the gluon jets in e^+e^- annihilation*, *Phys. Lett.* **B 94** (1980) 211;
JADE collaboration, W. Bartel et al., *Particle distribution in three jet events produced by e^+e^- annihilation*, *Z. Physik* **C 21** (1983) 37; *Experimental study of jets in electron-positron annihilation*, *Phys. Lett.* **B 101** (1981) 129; *Test of fragmentation models by comparison with three jet events produced in $e^+e^- \to$ hadrons*, *Phys. Lett.* **B 134** (1984) 275;
TPC/Two Gamma collaboration, H. Aihara et al., *Tests of models for quark and gluon fragmentation in e^+e^- annihilation $\sqrt{s} = 29\,GeV$*, *Z. Physik* **C 28** (1985) 31;
TASSO collaboration, M. Althoff et al., *A study of three jet events in e^+e^- annihilation into hadrons at 34.6 GeV center-of-mass energy*, *Z. Physik* **C 29** (1985) 29.

[73] L. Montanet, G.C. Rossi and G. Veneziano, *Baryonium physics*, *Phys. Rept.* **63** (1980) 149.

[74] X. Artru, *String model with baryons: topology; classical motion*, *Nucl. Phys.* **B 85** (1975) 442.

[75] M. Imachi, S. Otsuki and F. Toyoda, *Orientable hadron structure*, *Prog. Theor. Phys.* **54** (1975) 280;
Y. Igarashi et al., *Constituent rearrangement model and hadron reactions*, *Prog. Theor. Phys. Suppl.* **63** (1978) 49;
G.C. Rossi and G. Veneziano, *A possible description of baryon dynamics in dual and gauge theories*, *Nucl. Phys.* **B 123** (1977) 507.

[76] G.'t Hooft, *The topological mechanism for permanent quark confinement in a nonabelian gauge theory*, *Physica Scripta* **25** (1982) 133.

[77] P.Z. Skands, *Baryon number violation and string hadronization*, in *Hamburg 2002, supersymmetry and unification of fundamental interactions*, vol. 2 pp. 831–837, proceedings of *10th international conference on supersymmety and unification of fundamental interactions (SUSY02)*, Hamburg, Germany, 17–23 Jun 2002 [hep-ph/0209199].

[78] H.-U. Bengtsson, *The Lund Monte Carlo for high p_T physics*, *Comput. Phys. Commun.* **31** (1984) 323.

[79] N. van Remortel, *Bose-Einstein correlations in W W pair production at LEP*, hep-ex/0305014, to appear in the proceedings of 38th Rencontres de Moriond on QCD and High-Energy Hadronic Interactions, Les Arcs, Savoie, France, 22-29 Mar 2003;
A. Straessner, *QCD and W mass determination at LEP*, hep-ex/0304031.

[80] UA1 collaboration, C. Albajar et al., *Bose-Einstein correlations in $\bar{p}p$ interactions at $\sqrt{s} = 0.2$ TeV to 0.9 TeV*, *Phys. Lett.* **B 226** (1989) 410;
T. Alexopoulos et al., *A study of source size in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV using pion interferometry*, *Phys. Rev.* **D 48** (1993) 1931.

JHEP03(2004)053

[81] B. Andersson, G. Gustafson and B. Söderberg, *A general model for jet fragmentation*, *Z. Physik* **C 20** (1983) 317;
B. Andersson, S. Mohanty and F. Söderberg, *The λ-measure and the generalized dipoles in the Lund model*, *Nucl. Phys.* **B 646** (2002) 102 [hep-ph/0207025].

[82] E. Boos et al., *Generic user process interface for event generators*, in the proceedings of the *Workshop on physics at TeV colliders*, 21 May – 1 June 2001, Les Houches, France, P. Aurenche et al. eds., hep-ph/0109068.

[83] WA97 collaboration, F. Antinori et al., *Production of strange and multistrange hadrons in nucleus nucleus collisions at the SPS*, *Nucl. Phys.* **A 661** (1999) 130;
STAR collaboration, C. Suire, *Ω^- and $\bar{\Omega}^+$ production in Au + Au collisions at $\sqrt{s_{NN}} = 130\,GeV$ and $200\,GeV$*, *Nucl. Phys.* **A 715** (2003) 470 [nucl-ex/0211017];
NA57 collaboration, F. Antinori et al., *Hyperon production in 40-A GeV/c collisions from the NA57 experiment*, *J. Phys.* **G 30** (2004) S199.

[84] Pisa-Stony Brook-British-Scandinavian-MIT collaboration, G. Bellettini et al., *Associated multiplicities in proton proton collisions with a charged hadron at $x = 0$ at the CERN ISR*, *Nuovo Cim.* **A 42** (1977) 85;
British-Scandinavian collaboration, B. Alper et al., *Production spectra of $\pi^\pm$, $k^\pm$, $\rho^\pm$ at large angles in proton proton collisions in the CERN intersecting storage rings*, *Nucl. Phys.* **B 100** (1975) 237;
H1 collaboration, paper 556, submitted to ICHEP 98, and in preparation;
Ames-Bologna-CERN-Dortmund-Heidelberg-Warsaw collaboration, A. Breakstone et al., *A diquark scattering model for high p_T proton production in pp collisions at the ISR*, *Z. Physik* **C 28** (1985) 335;
STAR collaboration, C. Adler et al., *Mid-rapidity anti-proton to proton ratio from Au + Au collisions at $\sqrt{s_{NN}} = 130\,GeV$*, *Phys. Rev. Lett.* **86** (2001) 4778 [nucl-ex/0104022];
PHENIX collaboration, K. Adcox et al., *Measurement of the lambda and antilambda particles in Au + Au collisions at $\sqrt{s_{NN}} = 130\,GeV$*, *Phys. Rev. Lett.* **89** (2002) 092302 [nucl-ex/0204007];
PHOBOS collaboration, B.B. Back et al., *Ratios of charged antiparticles to particles near mid-rapidity in Au + Au collisions at $\sqrt{s_{NN}} = 200\,GeV$*, *Phys. Rev.* **C 67** (2003) 021901 [nucl-ex/0206012];
BRAHMS collaboration, I.G. Bearden et al., *Rapidity dependence of charged antihadron to hadron ratios in Au + Au collisions at $\sqrt{s_{NN}} = 200\,GeV$*, *Phys. Rev. Lett.* **90** (2003) 102301.

[85] D. Kharzeev, *Can gluons trace baryon number?*, *Phys. Lett.* **B 378** (1996) 238 [nucl-th/9602027];
B. Kopeliovich and B. Povh, *Baryon asymmetry of the proton sea at low x*, *Z. Physik* **C 75** (1997) 693 [hep-ph/9607486];
S.E. Vance, M. Gyulassy and X.N. Wang, *Baryon number transport via gluonic junctions*, *Phys. Lett.* **B 443** (1998) 45 [nucl-th/9806008];
G.T. Garvey, B.Z. Kopeliovich and B. Povh, *Baryon number flow in high-energy collisions*, *Comments Mod. Phys.* **A2** (2001) 47.

[86] T. Sjöstrand, *New showers with transverse momentum ordering*, submitted to the proceedings of the *Workshop on physics at TeV colliders*, Les Houches, France, 26 May – 6 June 2003, hep-ph/0401061.

[87] R. Enberg, G. Ingelman and N. Timneanu, *Soft color interactions and diffractive hard scattering at the Tevatron*, *Phys. Rev.* **D 64** (2001) 114015 [hep-ph/0106246].