

RECEIVED: May 8, 2015

REVISED: July 3, 2015

ACCEPTED: July 12, 2015

PUBLISHED: August 3, 2015

String formation beyond leading colour

Jesper R. Christiansen^{a,b} and Peter Z. Skands^{b,c}

^a*Department of Astronomy and Theoretical Physics, Lund University,
Sölvegatan 14, Lund, Sweden*

^b*Theoretical Physics, CERN,
CH-1211, Geneva 23, Switzerland*

^c*School of Physics and Astronomy, Monash University,
VIC-3800, Australia*

E-mail: jesper.christiansen@thep.lu.se, peter.skands@monash.edu

ABSTRACT: We present a new model for the hadronisation of multi-parton systems, in which colour correlations beyond leading N_C are allowed to influence the formation of confining potentials (strings). The multiplet structure of $SU(3)$ is combined with a minimisation of the string potential energy, to decide between which partons strings should form, allowing also for “baryonic” configurations (e.g., two colours can combine coherently to form an anticolour). In e^+e^- collisions, modifications to the leading-colour picture are small, suppressed by both colour and kinematics factors. But in pp collisions, multi-parton interactions increase the number of possible subleading connections, counteracting their naive $1/N_C^2$ suppression. Moreover, those that reduce the overall string lengths are kinematically favoured. The model, which we have implemented in the PYTHIA 8 generator, is capable of reaching agreement not only with the important $\langle p_\perp \rangle$ (n_{charged}) distribution but also with measured rates (and ratios) of kaons and hyperons, in both ee and pp collisions. Nonetheless, the shape of their p_\perp spectra remains challenging to explain.

KEYWORDS: QCD Phenomenology, Monte Carlo Simulations

ARXIV EPRINT: [1505.01681](https://arxiv.org/abs/1505.01681)

Contents

1	Introduction	1
2	The model	5
2.1	Existing MPI models and colour space	6
2.2	Beyond leading colour	8
2.3	The new model	11
2.3.1	Colour flow in the beam remnant	13
2.3.2	Colour flow in the whole event	16
2.3.3	Hadronisation of multi-junction systems	21
2.3.4	Space-time structure	23
3	Constraints and tuning	25
3.1	Lepton colliders	25
3.2	Hadron colliders	27
3.3	Direct CR constraints?	31
3.4	Suggestions for new observables	33
4	Application to top mass measurements at hadron colliders	40
5	Summary and outlook	43
A	Model parameters	44

1 Introduction

The description of hadronic final states at high-energy colliders involves a complicated cocktail of physics effects, dominated by QCD [1–3]. For the calculation of *inclusive* hard-scattering cross sections, factorisation allows most of the complicated long-distance physics to be represented in the form of universal parton distribution functions (PDFs) [4], while the short-distance parts can be calculated perturbatively. Perturbative aspects, such as hard-process matrix elements, parton showers, and decay (chains) of short-lived resonances, are generally coming under increasingly good control, due to a combination of advances: better amplitude calculations (including better automation and better interfaces [5–12]), better parton-shower algorithms (e.g. ones based on QCD dipoles [13–19]), and better techniques for how to combine them (matching and merging, see [3, 12, 20–22] and references therein). These successes build on an extensive prior experience with perturbative approximations to QCD at both fixed and infinite order, and the tractable nature of the perturbative expansions themselves.

To describe the full (*exclusive*) event structure, however, several additional soft-physics effects must be accounted for, such as hadronisation, multiple parton interactions (MPI), Bose-Einstein correlation effects, and beam remnants. These are connected with the rich structure of QCD beyond perturbation theory and are vital, each in their own way, to the understanding of issues such as underlying-event/pileup effects on isolation and accurate jet calibrations, and the interpretation of identified-particle rates and spectra.

For these aspects, explicit calculations can only be performed in the context of simplified phenomenological models, constructed so as to capture the essential features of full (nonperturbative) QCD. An example relevant to this paper is the Lund string model of hadronisation [23, 24], whose cornerstone is the observation that the static QCD potential between a quark and an antiquark in an overall colour-singlet state grows linearly with the distance between them, for distances larger than about 0.5 fm [25]. This is interpreted as a consequence of the gluon field between the charges forming a high-tension “string” (with tension $\kappa \sim 1 \text{ GeV/fm}$), which subsequently fragments into hadrons.

While the details of the string-breaking process may be complicated (the Lund model invokes quantum tunnelling to describe this aspect [23]), the first question that any hadronisation model needs to address is therefore simply: *between which partons do confining potentials arise?* In string-based models, this is equivalent to answering the question between which partons string pieces should be formed. Traditionally, Monte Carlo event generators make use of the leading-colour (LC) approximation to trace the colour flow on an event-by-event basis (see [3, 26]), leading to partonic final states in which each quark is colour-connected to a single (unique) other parton in the event (equivalent to a leading-colour QCD dipole [27]). Gluons are represented as carrying both a colour and an anticolour charge, and are hence each connected to two other partons. At the level of strings, this is interpreted as gluons forming transverse “kinks” on strings whose endpoints are quarks and antiquarks [23]. Studies at ee colliders show this to be a quite reasonable approximation in that environment, and the traditional Lund string model, implemented in PYTHIA [28–30], is capable of delivering a good description of the vast majority of ee collider data (for recent studies, see, e.g., [31–34]).

The question of *colour reconnections* (CR) — broadly, whether other string topologies than the LC one could lead to non-negligible corrections with respect to the LC picture — was studied at LEP [35–42], chiefly in the context of CR uncertainties on W mass determinations in $ee \rightarrow WW$ [43], with conclusion that excluded the very aggressive models and disfavoured the no CR scenario at 2.8 standard deviation [44]. The uncertainty on the W mass from this source ended up at $\Delta m_W \sim 35 \text{ MeV}$, corresponding to about 0.05%.

There are strong physical reasons to think that CR effects *should* be highly suppressed at LEP, however. Firstly, there is a “trivial” parametric suppression of beyond-LC effects of order $1/N_C^2 \sim 10\%$. Secondly, the two W decay systems are separate colour-singlet systems, with a space-time separation of order of the inverse W width, $\Gamma_W \sim 2 \text{ GeV}$. This separation implies that interference effects between the two systems should be highly suppressed for wavelengths shorter than $1/\Gamma_W$, i.e., there can be essentially no perturbative cross-talk between them. This line of argument motivated the phrasing of CR models that operate only at the non-perturbative level as the most physically reasonable [43], an

observation that we shall also adhere to in the present work. Thirdly, the QCD coherence of perturbative parton cascades implies that, inside each W (or Z) decay system, angles of successive QCD emissions tend to be ordered from large to small [45], so that there is very little space-time overlap between the QCD dipoles inside each system. This means that, even if one were to allow to set up confining potentials between non-LC-connected partons, these would tend to correspond to *larger* opening angles and therefore they would have a higher total potential energy (longer strings) than the equivalent LC ones. The LC topology should therefore also be *dynamically* favoured over any possible non-LC ones. All these factors contribute to an expectation of quite small effects, at least in the context of e^+e^- collisions.

Moving to pp collisions (and using pp as a shorthand to for any generic hadron-hadron collision, including in particular also $p\bar{p}$ ones), the situation changes dramatically. Trivially, one must now include coloured initial-state partons, with associated coloured beam remnants. But more importantly, the modern understanding of the underlying event (UE) and of soft-inclusive (minimum-bias/pileup) physics in general, especially at high particle multiplicities, is that they are dominated by contributions from multiple parton interactions (MPI) [46]. In a pp event that contains several MPI systems, there is a non-negligible possibility of phase-space overlaps between final states from different MPI systems. Moreover, since the MPI scattering centres must all reside within the proton radius, which is of the same order as the transverse size of QCD strings, the initial-state (beam) jets will all “sit” right on top of each other, a situation which should affect the fragmentation especially at high rapidities. Finally, unlike the case for angular-ordered partons inside a jet, there is no perturbative principle that predisposes colour-connected partons from different MPI or beam-remnant systems to have small opening angles; indeed a recent study [47] found that such “inter-MPI/remnant” invariant masses (denoted i -type and n -type in [47]) tend to be among the largest in the events, corresponding to a high potential energy in a string context, and hence with the most to gain from potential reconnections. For these reasons, we expect qualitatively larger effects in pp collisions.

There are also tantalising hints from hadron-collider data that nontrivial physics effects are present at the hadronisation stage in pp collisions. The most important such clue is furnished by the dependence of the average (charged) particle p_\perp on the particle multiplicity, $\langle p_\perp \rangle (n_{\text{ch}})$. Measurements of this quantity in minimum-bias events, first made at the ISR [48] and since by UA1 [49], CDF [50, 51] and the LHC experiments [52–54], reveal that $\langle p_\perp \rangle$ grows with n_{ch} , as can be seen in the plots in figure 1 (from mcplots.cern.ch [55]). This cannot be accounted for by independently hadronising MPI systems, for which the expectation would be that $\langle p_\perp \rangle (n_{\text{ch}})$ should be almost flat, as is also illustrated by the “no CR” curves in figure 1. (If each MPI hadronises independently, then per-particle quantities such as $\langle p_\perp \rangle$ should be independent of the number of MPI, which is correlated with n_{ch} [46].) The observation that $\langle p_\perp \rangle$ increases with n_{ch} therefore strongly suggests that some form of collective hadronisation phenomenon is at play, correlating partons from different MPI systems.

Given these arguments, and the realisation [57] that precision kinematic extractions of the top quark mass at hadron colliders (see e.g., [58–64] for experimental methods and [65]

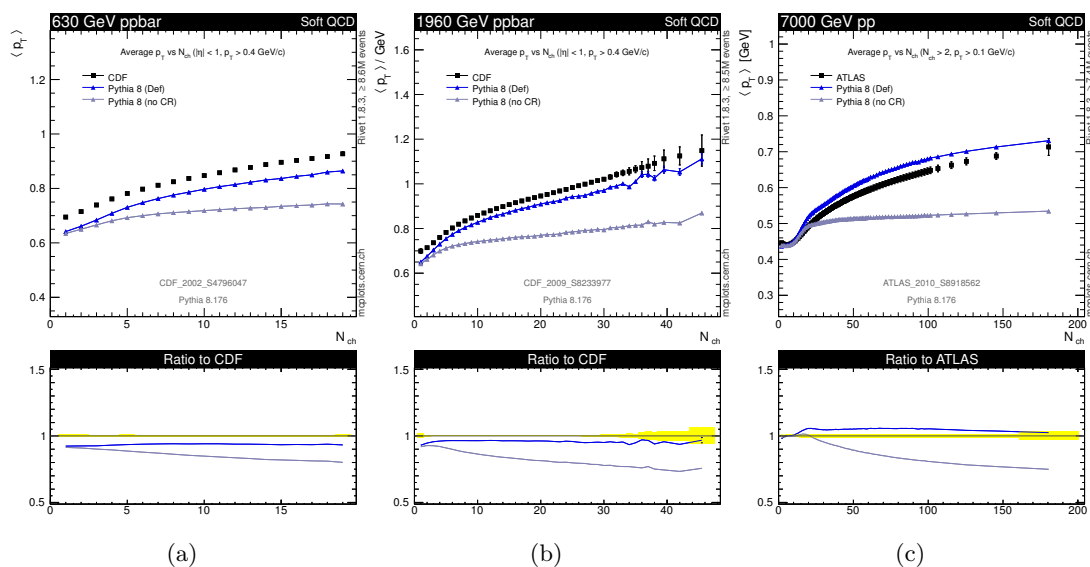


Figure 1. Measurements of $\langle p_{\perp} \rangle (n_{ch})$ in minimum-bias events at 630 GeV [50] (a), 1960 GeV [51] (b), and 7000 GeV [53] (c), compared to PYTHIA 8.175 [29] (tune 4C [56]), with and without colour reconnections switched on. (Plots from mcplots.cern.ch [55].)

for a recent phenomenology review) can be significantly affected by colour reconnections,¹ several toy models have appeared [47, 57, 68–71], relying mainly on potential-energy minimisation arguments to reconfigure the partonic colour connections for hadronisation. Although these models have had some success in describing the $\langle p_{\perp} \rangle (n_{ch})$ distribution (as e.g., in figure 1), the lack of rigorous underpinnings have implied that large uncertainties remain, which still contribute about a 500 MeV uncertainty on the hadronic top mass extraction [62, 65, 71]. In this paper, we take a first step towards creating a more realistic model, combining the earlier string-length minimisation arguments with selection rules based on the colour algebra of SU(3). Our treatment amounts to taking the LC connections produced by the shower as a starting point, complemented by an SU(3)-weighted randomization over the set of possible subleading topologies that would have been present in a full-colour treatment. The missing colour information should thereby be restored, at least in a statistical sense.

An alternative line of argument, pursued in particular in the EPOS model [72], invokes the notion of hydrodynamic collective flow to explain the $\langle p_{\perp} \rangle (n_{ch})$ distribution (as well as the so-called CMS “ridge effect” [73, 74] and a host of other pp observables [72]). Certainly, the presence of hydro effects in pp is a hypothesis that, if confirmed, would have far-reaching consequences, and it will be an important task for future experimental and phenomenological studies to find ways of disentangling CR effects from hydro ones. In this context, our paper should therefore also be viewed as an attempt to see how far one can get *without* postulating genuine (pressure-driven) collective-flow effects in pp . Within

¹For completeness we note that, similarly to above, much smaller effects are expected in e^+e^- environments [66, 67].

this context, it is important to note that CR can mimic flow effects to some extent, via the creation of boosted strings [75]. Alternatively, it is possible that the effective string tension could be rising, as in the idea of colour ropes [76], with recent work along these lines reported on by the Lund group [77]. Finally, we note that non-hydro rescattering has also been proposed [78] as a potential mechanism contributing to the rise of $\langle p_{\perp} \rangle (n_{\text{ch}})$, though the explicit model of parton-parton rescattering effects presented in [78] found only very small effects. The possibility of Boltzmann-like elastic (or even inelastic) final-state hadron-hadron rescattering is still open. As usual, nature’s solution is likely to involve an interplay of effects at different levels. Nevertheless, before exploring further effects at the hadron level, we believe it makes good sense to first examine the hadronisation process itself, which is the topic of this work.

Finally, we note that colour flows beyond LC have also been invoked in the context of J/ψ formation [79–82], and as a potential mechanism to generate diffractive topologies in ep and pp collisions [83, 84].

In section 2, we briefly recapitulate the treatment of colour space for the existing MPI models in PYTHIA, and present the new model that we have developed, combining the minimisation of the string potential with the multiplet structure of QCD. In section 3, we constrain the resulting free model parameters on a selection of both ee and pp data, discussing the physics consequences of the new colour-space treatment as we go along. In section 4, we consider implications for precision extractions of the top quark mass at hadron colliders. Finally, in section 5, we summarise and give an outlook.

2 The model

In this section, we present the colour-space model that we have developed, which allows strings to form not only between LC-connected partons, but also between specific non-LC-connected ones, following combination rules that approximate the multiplet structure of full-colour QCD. We begin with a brief summary of the current modelling, in section 2.1. We then turn to a general discussion of coherence effects beyond leading N_C in section 2.2. Finally, in section 2.3, we present the detailed implementation of the new model.

We emphasise that there is a conceptual difference between colour-space *ambiguities*, such as those explored in this work, and physical colour *reconnections*. The subleading-colour effects we discuss here arise naturally in “full-colour” $SU(3)$ and do not involve any physical exchange of colour or momentum (although explicit algorithms may of course still employ an iterative-reconnection scheme to find the potential-energy minimum). Strictly speaking, the term colour *reconnections* should be reserved to describe effects related to *dynamical* reconfigurations of the colour/string space that involve explicit exchange of colours and momentum, via perturbative gluon exchanges or non-perturbative string interactions. Effects of this type are not explored directly in this work, instead we refer the interested reader to the SK string-interaction models presented in [43, 66, 67]. Somewhat sloppily, we follow the entrenched convention in the field and use the acronym “CR” for effects of either kind here.

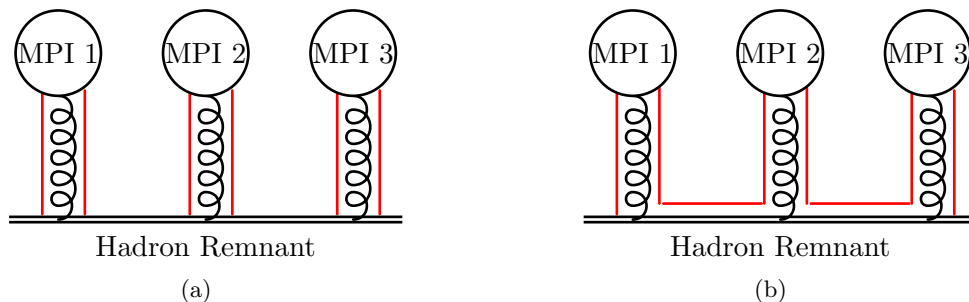


Figure 2. (a): in a strict LC picture, each MPI initiator gluon increases the “colour charge” of the beam remnant by two units. (b): allowing different MPI initiators to be connected in a colour chain reduces the total colour charge of the remnants. Here for example, no strings will be stretched directly between the shown remnant and the final states of MPI 2. (Note that the colour assignments shown are for illustration only, and would be represented by Les Houches Colour Tags [5, 6] in a real event generator.)

2.1 Existing MPI models and colour space

In a naive LC picture, each MPI scattering system is viewed as separate and distinct from all other systems in colour space. The very simplest colour-space options in the old PYTHIA 6 MPI model [46] and the first HERWIG (and HERWIG++) MPI models [85, 86] go a step further, representing each MPI final state as two quarks (or gluons), colour-connected directly to each other, i.e., treating each MPI system as a separate hadronising colour-singlet system. However, this ignores that the incoming partons are coloured, and hence that the total colour charge of each MPI scattering system is in general non-zero. These particular models therefore violate colour conservation and are unphysical.

To be LC-correct one must take into account that each MPI-initiator parton should cause one or two strings to be stretched to its remnant (one for quarks, two for gluons). This conserves colour, but still has the implication that no strings would be stretched *between* different MPI systems. This situation is illustrated in figure 2(a). Physically, this can lead to arbitrarily many strings being stretched across the central rapidity region, one or two for each MPI (corresponding to adding their total colour charges together as scalar quantities, rather than as SU(3) vectors).

However, already in the context of earlier works [46, 87], it was noted that even this picture cannot be quite physically correct. Since all the MPI initiators on each side are extracted from one and the same (colour-singlet) beam particle, and since they are extracted at a rather low scale of order the perturbative evolution cutoff $p_{\perp 0} \sim$ one to a few GeV, there is presumably some overlap and accompanying saturation effects, implying that they are not completely independent. Not knowing the exact form of the correlations, a pragmatic solution is to minimise the total colour charge of the remnant (and hence the number of strings stretched to it), by allowing the different MPI systems to be colour-connected to each other along a “chain” in colour space, as illustrated in figure 2(b). Variations of this are used in the current forms of both the PYTHIA [46, 87] and HERWIG++ [47] MPI models, reducing the number of strings/clusters especially in the remnant-fragmentation region at high rapidities. It is, however, still fundamentally ambiguous exactly which systems to

connect and how. In the example of figure 2, it is arbitrary that it happens to be the colour of MPI 1 and the anticolour of MPI 3 which end up connected to the remnant. For a more detailed discussion of this aspect, see e.g. [87]. An interesting physics point is that, in this picture, the particle production at very forward rapidities is controlled essentially by how large one allows the colour charge of the remnant to become, which in turn depends on the number of MPI and their mutual colour correlations. This could presumably be revealed by studies correlating the particle production in the central region (sensitive to the number of MPI) with that in the forward region (sensitive to the total charge of the remnant).

In the absence of any further CR effects, the relationship between the number of MPI and the average particle multiplicity at central rapidities is still approximately linear. Consequently, the per-particle spectra in high-multiplicity events (with many MPI) are similar to those in (non-diffractive) low-multiplicity events (with few MPI).² This is what leads to the simple expectation of the flat $\langle p_{\perp} \rangle (n_{\text{ch}})$ spectrum exhibited by the “no CR” curves that were shown in figure 1 in the previous section. However, as was also remarked on there, the experimental data convincingly rule out such a constant behaviour. This observation is the main reason additional non-trivial final-state CR effects have been included in both HERWIG++ and PYTHIA.

In the original (non-interleaved) MPI model in PYTHIA 6 [28, 46], the parameters `PARP(85)` and `PARP(86)` allowed to force a fraction of the MPI final states to be two gluons colour-connected to their nearest neighbours in momentum space. The physical picture was that the hardest interaction built up a “skeleton” of string pieces, onto which a fraction of the gluons from MPI were grafted (by brute force) in the places where they caused the least amount of change of string length. This effectively minimised the increase in string length from those gluons. An important factor contributing to the revival of the question of CR in hadron collisions was the tuning studies of this model, carried out by Rick Field on underlying-event and minimum-bias data from the CDF experiment at the Tevatron [50, 88, 89]. His resulting “Tune A” and related tunes [90, 91] were the first to give good fits to the available data at the time, but the surprising conclusion was that in order to do so this “colour-space grafting” had to be done nearly all the time.

An alternative set of CR models, which relied on physical analogies with overlapping strings in superconductors, were developed only in the context of e^+e^- collisions [43, 66, 67], chiefly with the aim of studying potential CR uncertainties on the W mass, see [92] and references therein. As far as we are aware, this class of models has not yet been applied in the context of the more complex environment of hadron-hadron collisions.

In the new (interleaved) MPI model in PYTHIA 6 [14, 28], showers and MPI were carried out in parallel, with physical colour flows. This was too complicated to handle with the old CR model. A new “colour annealing” CR scenario was developed [57, 69, 93] which, after the shower evolution had finished, allowed for a fraction of partons to “forget” their LC colour connections, with new ones determined based on the string area law (shorter

²For very low multiplicities, well-understood bias effects cause the average particle p_{\perp} to increase (if the event is required to contain only one particle, then that particle must be carrying all the scattered energy), while for high multiplicities, the contribution from hard-jet fragmentation also generates slightly harder spectra.

strings are preferred), following a simplified annealing-like algorithm, in a similar spirit to an earlier model by Rathsman, called the “Generalized Area-Law” (GAL) model [68]. The fraction of partons that forgot their LC colour connections was assumed to grow with the number of MPI, with a per-MPI probability given by the parameter `PARP(78)`. A further parameter, `PARP(77)`, allowed to suppress the reconnection probability for fast-moving partons. Although still intended as a toy model, the new colour-annealing models obtained good agreement with the Tevatron minimum-bias and underlying-event data, e.g. in the form of the Perugia family of tunes [94, 95]. The most recent incarnations, the Perugia 2011 and 2012 tunes, also included LHC data and were among the main reference tunes used during Run 1 of the LHC [95]. However, a study comparing independent MPI+CR tunings at different collider energies revealed different preferred CR parameter values at different CM energies [96], implying that the modelling of this aspect, or at least its energy dependence, was still inadequate.

In PYTHIA 8, the default MPI colour-space treatment is similar to that of the original PYTHIA 6 model, although starting out from a more detailed modelling of the colour flow in each MPI. With a certain probability, controlled by the parameter `ColourReconnection:range`, all the gluons of each lower- p_{\perp} interaction can be inserted onto the colour-flow dipoles of a higher- p_{\perp} one, in such a way as to minimise the total string length [71]. The effects of this model was already illustrated in figure 1. A set of alternative CR scenarios was also presented in [71], but were still mostly intended as toy models in the context of estimating uncertainties on the top-quark mass.

Finally, in the most recent developments of the HERWIG++ MPI model, an explicit scenario for colour reconnections has likewise been introduced [47], based on a simulated-annealing algorithm that minimises (sums of) cluster masses. In the context of the cluster hadronisation model [97], the minimisation of cluster masses fulfils a similar function as the minimisation of string lengths above. The two minimisations differ in that the string length measure is closely related to the product of the invariant masses rather than the sum used in the cluster model. The main model parameter is the probability to accept a favourable reconnection, p_{reco} . The study in [47] emphasised in particular that the largest pre-reconnection cluster masses are spanned between hard partons and the remnants (denoted n -type clusters), with inter-MPI ones (spanned directly between partons from different MPI systems and denoted i -type) having the second-largest masses. The former again indicates that there is a non-trivial interplay with the non-perturbative hadronisation of the beam remnant, while the latter reflects the lack of a priori knowledge about the colour correlations *between* different MPI systems. Similarly to the qualitative conclusions made with the PYTHIA CR models, the HERWIG++ study found that quite large values of $p_{\text{reco}} \sim 0.5$ were required to describe hadron-collider data.

2.2 Beyond leading colour

To illustrate the colour-space ambiguity between different MPI systems, and between them and the beam remnant, let us take the simple case of double-parton scattering (DPS), with all the initiator partons being gluons. What happens in colour space when we extract *two* gluons from a proton? Even if we imagine that the two gluons are completely uncorrelated,

QCD gives several possibilities for their superpositions:

$$\mathbf{8} \otimes \mathbf{8} = \mathbf{27} \oplus \mathbf{10} \oplus \overline{\mathbf{10}} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} . \quad (2.1)$$

The highest-charge multiplet, here the **27** (a “viginti-septet”), effectively corresponds to the LC configuration: symmetric addition of the two gluons, each carrying two units of (LC) colour charge (one colour and one anticolour), for a total of 4 string pieces required to be attached to the remnant.³ However, note that the probability for this to occur is

$$P_{LC} = \frac{27}{64} < 50\% , \quad (2.2)$$

hence the naive expectation that subleading topologies should be suppressed by $1/N_C^2$ is badly broken already in this very simple case.⁴ The decuplets (octets) correspond to coherent-superposition topologies with a lower total colour charge and consequently only three (two) string pieces attached to the remnant. The singlet represents the special case in which the two MPI-initiator gluons have identical and opposite colours, with total colour charge 0 (generating a diffractive-looking topology from the point of view of the remnant). In QCD, for two random (uncorrelated) gluons, there is a $1/64$ probability for this to happen purely by chance.

The other possible two-parton combinations are:

$$\mathbf{3} \otimes \mathbf{8} = \mathbf{15} \oplus \overline{\mathbf{6}} \oplus \mathbf{3} , \quad (2.3)$$

$$\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1} , \quad (2.4)$$

$$\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \overline{\mathbf{3}} , \quad (2.5)$$

where strict LC would correspond to populating only the **15** (quindecuplet), **8** (octet), and **6** (sextet), respectively. The relative weights (probabilities) for each multiplet in each of these combinations are illustrated in figure 3, along with diagrams exemplifying corresponding colour flows (with thick lines indicating partons, thin ones colour-flow lines). For each multiplet, three vertical bars indicate the probability associated with that multiplet in strict Leading Colour (LC), in our model (defined below), and in SU(3) (QCD), respectively. The filled circles represent the *ratio* between our model and QCD, so for those *unity* indicates perfect agreement. Note that, since the subleading multiplets are absent in LC, only two non-zero bars appear for them. Below, we shall also consider the probability for three uncorrelated triplets to form an overall singlet, which is $1/27$ in QCD:

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} , \quad (2.6)$$

while in our simplified model it will come out to be $2/81 = (1 - \frac{1}{3})/27$.

We emphasise that we only use these composition rules for colour-unconnected partons, which in the context of our model we approximate as being totally uncorrelated. LC-connected partons are always in a singlet with respect to each other, and colour neighbours

³Assuming each string can only carry one unit of flux, or equivalently a 4-unit “colour rope”, see [77].

⁴In general, the highest-charge multiplet in the combination of k gluons represents a fraction $(k+1)^3/8^k$ of the possibilities.

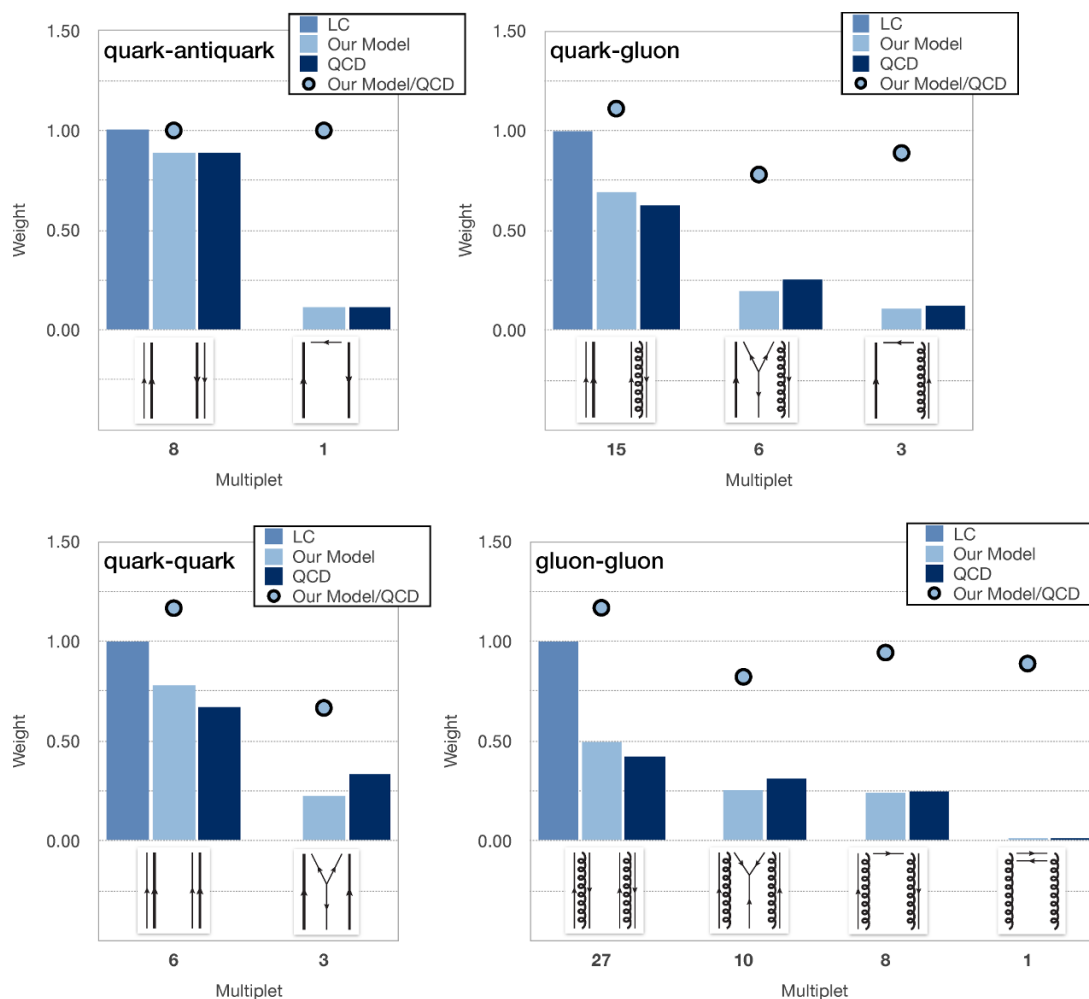


Figure 3. Illustration of the possible colour states of two random (uncorrelated) partons. In strict LC, only the completely incoherent superposition is populated. Our model (described below) gives a systematically better approximation. Filled circles show the ratio between our model and full SU(3). The diagrams below the histograms attempt to illustrate corresponding colour-flow configurations, with thick and thin lines denoting partons and colour-flow lines, respectively.

(e.g., the two colour lines of a gluon or those of a $q\bar{q}$ pair produced by a $g \rightarrow q\bar{q}$ splitting) are never in a singlet with respect to each other.

The approximation of colour-unconnected partons being totally uncorrelated, combined with a set of specific colour-space parton-parton composition rules, such as those of SU(3) or the simplified ones defined below, allow us to build up an approximate picture of the possible colour-space correlations that a complicated parton system can have, including randomised coherence effects beyond Leading Colour. Due to the subleading correlations, there are many possible string topologies that could represent such a parton system, including but not limited to the LC one. The selection principle that determines how the system collapses into a specific string configuration will be furnished by the minimisation of the string potential, as we shall return to below.

Our model thus consists of two stages. First, we generate an approximate picture of the possible colour states of a parton system. Then, we select a specific realisation of that state in terms of explicit string connections. This is done at the time when the system is prepared for hadronisation, i.e., after parton showering but before string fragmentation.

By maintaining the structure of the (LC) showers unchanged, we neglect any possibility of reconnections occurring already at the perturbative level. Though perturbative gluon exchanges and/or full-colour shower effects might mediate such effects in nature, we expect their consequences to be suppressed relative to the non-perturbative ones considered here. This is partly due to the coherence and collinear-enhancement properties already acting to minimise the mass of LC dipoles inside each perturbative cascade, and partly due to the space-time separation between different systems (be they different MPI systems, which are typically separated by transverse distances of order $1/\Lambda_{\text{QCD}}$ inside the proton, or different resonance-decay systems separated by $1/\Gamma_{\text{res}}$). Thus, at high $Q \gg \Lambda_{\text{QCD}}$ or $Q \gg \Gamma_{\text{res}}$ we don't expect any cross-talk between different MPI or different resonance systems, respectively. The case can be made that perturbative reconnection effects could still be active at longer wavelengths, but we expect that such semi-soft effects can presumably be absorbed in the non-perturbative modelling without huge mistakes.

2.3 The new model

Our simplified colour-space model is defined as follows. Rather than attempting to capture the full correlations (which we have emphasised are not a priori known anyway and would require a cumbersome matrix-based formalism), we note that the main subleading parton-parton combination possibilities of real QCD can be encoded in a single “colour index”, running from 1 to 9 (with corresponding indices for anti-colours).

Quarks are assigned a single such colour index, antiquarks a single anticolour index, and gluons have one of each, with the restriction that their colour and anticolour indices cannot be the same. Thus, formally our model has 9 different quark colour states and 72 kinds of gluon states. We emphasise that these indices should not be confused with the ordinary 3-dimensional SU(3) quark colour indices (red, green, and blue); rather, our index labels the possible colour states of two-parton (and in some cases three-parton) *combinations*. Thus, for example, a quark and an antiquark are in an overall colour-singlet state if the colour index of the former equals the anticolour index of the latter, otherwise they are in an octet state, cf. eq. (2.4). We note that a similar index was used already in the models of “dipole swing” presented in [98, 99], though here we generalise to parton combinations involving colour-epsilon structures as well, cf. figure 3.

Confining potentials will be allowed to form between any two partons that have matching colour and anticolour indices. Since LC-connected partons are forced to have matching colour and anticolour indices, the “original” (LC) string topology always remains possible, but now further possibilities also exist involving partons that *accidentally* have matching indices, illustrated in figure 4.

Furthermore, two *colour* indices are allowed to sum coherently to a single *anticolour* index within three separate closed index groups: [1,4,7], [2,5,8], and [3,6,9]. E.g., two quarks carrying indices 2 and 5 respectively, are allowed to appear to the rest of the event

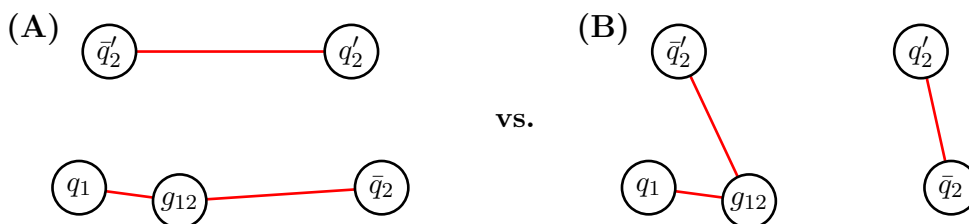


Figure 4. Illustration of a multi-parton state with a rather simple colour-space ambiguity. Subscripts indicate colour-space indices. (A): the “original” (LC) string topology. (B): an alternative string topology, allowed by the accidentally matching “2” indices.

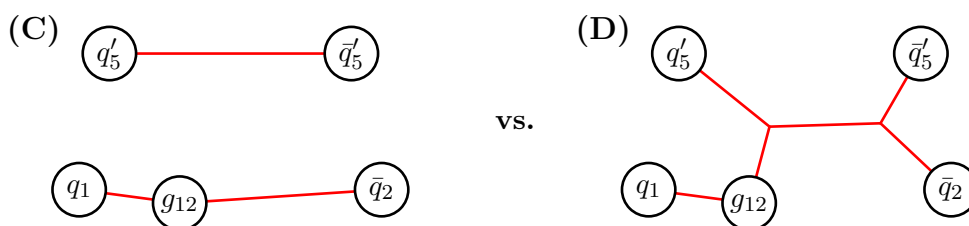


Figure 5. Illustration of a similar multi-parton state as in figure 4, but now with index assignments resulting in a junction-type colour-space ambiguity. The orientation of the top $q'_5 - \bar{q}'_5$ dipole has also been reversed relative to figure 4. (C): the original (LC) string topology. (D): an alternative string topology with a junction and an antijunction, allowed by the cyclically matching “2” and “5” indices.

as carrying a single combined anti-8 index. These index combinations represent the anti-symmetric ε_{ijk} colour combinations that were pictorially represented as Y-shaped “colour junctions” in figure 3. A junction can therefore be interpreted as the string extension of a baryon with the baryon number (ε_{ijk}) located in the centre of the Y-shape [100]. An explicit example of a parton system whose colour state includes such a possibility is shown in figure 5. A model for string hadronisation of such topologies was developed in [100] and has subsequently also been applied to the modelling of baryon beam remnants [87]. We reuse it here for hadronisation of junction-type colour-index combinations.

The new model can be divided into two main parts: a new treatment of the colour flow in the beam remnant and a new CR scheme. The two models are independent and can therefore in principle be combined both with each other as well as with other models. (Note, however, that the old PYTHIA 8 CR scheme is inextricably linked with the colour treatment of the beam remnant and therefore only works together with the old beam-remnant model.) Both of the models occasionally result in complicated multi-junction configurations that the existing PYTHIA hadronisation cannot handle. Rather than attempting to address these somewhat pathological topologies in detail, this problem is circumvented by a clean-up method that simplifies the structure of the resulting systems to a level that PYTHIA can handle.

The next two sections describe respectively the details of the new beam-remnant model and the new CR scheme, including technical aspects and the algorithmic implementation. Afterwards the junction clean-up method is described.

2.3.1 Colour flow in the beam remnant

It being an inherently non-perturbative object, we do not expect to be able to use perturbative QCD to understand the structure of the beam remnant. Instead, we rely on conservation laws; the partons making up the beam remnant must, together with those that have been kicked out by MPI, sum up to the total energy and momentum of the beam particle, be in an overall colour-singlet state, with unit baryon number (for a proton beam), carrying the appropriate total valence content for each quark flavour, with equal numbers of sea quarks and antiquarks. The machinery used to conserve all these quantities should be consistent with whatever knowledge of QCD we possess, such as the standard single-parton-inclusive PDFs to which our framework reduces in the case of single-parton scattering.

In this work the focus is on the formation of colour-singlet states, including the use of SU(3) epsilon tensors. This naturally leads to a modification of the treatment of baryon number conservation, due to the close link between baryons and the epsilon tensors in SU(3). The modelling of energy/momentum and flavour conservation is not touched relative to the existing modelling of those aspects, and thus only a small review is presented here (for more details see [87]).

The overall algorithm can be structured as follows:

1. Determine the colour structure of the already scattered partons.
2. Add the minimum amount of partons needed for flavour conservation.
3. Add the minimum amount of gluons required to obtain a colour-singlet state.
4. Connect all colours.
5. With all the partons determined find their energy fractions.

The conservation of baryon number is not listed as a separate point, but naturally follows from the formation of junctions. Let us now consider each of these points individually starting from the top.

To calculate the colour structure of the beam remnant, let us return to the DPS example of earlier. With a probability of 27/64, the two gluons form a completely incoherent state, leaving four colour charges to be compensated for in the beam remnant (two colours and two anticolours). However the three valence quarks alone are insufficient to build up a **27** (eq. (2.6)), and therefore a minimum of one additional gluon is needed. Then two of the quarks can combine to form a $\bar{\mathbf{3}}$, which can form an **8** with the remaining valence quark, which then can enter in a **27** with the added gluon. Conversely, if the two gluons had been in an octet state instead, the additional gluon would not have been needed. Thus, in order to determine the minimal number of gluons needed in the beam remnant, we need to know the overall colour representation of all the MPI initiators combined.

While it could be possible to choose this representation purely statistically, based on the (simplified or full) SU(3) weights, we note two reasons that a lower total beam-remnant charge is likely to be preferred in nature. Firstly, to determine the most preferred configuration the string length needs to be considered. Since the beam remnants reside in

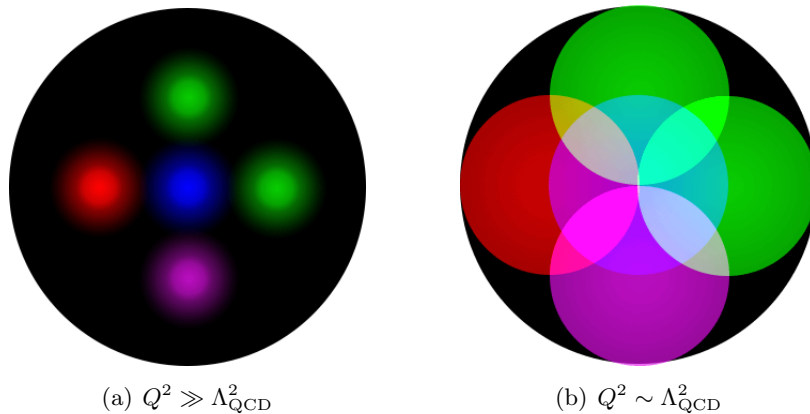


Figure 6. A proton (black) with five distinct colour sources (e.g., four MPI-initiator partons plus one representing the beam remnant), chosen so that they add to a singlet (with magenta = antigreen). Shown are two different resolution scales representative of (a) the perturbative stage, during which the MPI systems are considered as being uncorrelated in colour space, and (b) the nonperturbative stage, at which the beam remnant is considered and we assign higher weights to states with lower total QCD charge in order to mimic saturation effects.

the very forward regime, strings spanned between the remnant and the scattered gluons tend to be long, and as such a good approximation is to minimise the number of strings spanned to the beam. This corresponds to preferring a low-charge colour-multiplet state for the remnant, and as a consequence also minimises the number of additional gluons required. Secondly, a purely stochastic selection corresponds to the assumption that the scattered partons are uncorrelated in colour space. For hard MPIs (at $Q \gg \Lambda_{\text{QCD}}$), this is presumably a good approximation, since the typical space-time separation of the collisions are such that two independent interactions do not have time to communicate. This is illustrated by figure 6(a). However, after the initial-state radiation is added, the lower evolution scale implies larger spatial wavefunctions, allowing for interference between different interactions, illustrated by figure 6(b). An additional argument is that at a low evolution scale the number of partons is low, thus to combine to an overall singlet the correlation between the few individual partons needs to be large. To provide a complete description of this cross-talk, multi-parton densities for arbitrarily many partons would be needed, ideally including colour correlations and saturation effects. Although correlations in double-parton densities has been the topic of several recent developments [101–109], the field is still not at a stage at which it would be straightforward to combine explicit double-parton distributions with the standard single-parton-inclusive ones (which a code like PYTHIA must be compatible with), nor to generalise them to arbitrarily many partons. The only formalism we are aware of that addresses all of these issues (in particular flavour and momentum correlations for arbitrarily many partons) while reducing to the single-parton ones for the hardest interaction remains the one developed in the context of the current PYTHIA beam-remnant model [87]. In this study, we supplement the momentum- and flavour-correlation model of [87] with a simple model of colour-space saturation effects appropriate to the SU(3)-multiplet language used in this work. Noting that saturation should lead to a suppression

of higher-multiplet states, we use a simple ansatz of exponential suppression with multiplet size, M :

$$p(M) = \exp(-M/k_{\text{saturation}}) \tag{2.7}$$

where p is the probability to accept a multiplet of size M and $k_{\text{saturation}}$ is a free parameter that controls the amount of suppression.

Everything stated above for the two-gluon case can be generalised to include quarks and an arbitrary number of MPI. The calculation just extends to slightly more complicated expressions:

$$\mathbf{8} \otimes \mathbf{3} \otimes \bar{\mathbf{3}} \otimes \mathbf{8} \otimes \dots = \dots \tag{2.8}$$

where quarks enter as triplets, antiquarks as antitriplets and gluons as octets. The statistical probability to choose any specific multiplet can be calculated in a similar fashion, either in full or simplified QCD. There is however still an ambiguity in how the colours are connected. For instance consider two quarks and one antiquark forming an overall triplet state. The colour calculation will not tell us which of the quarks are in a singlet with the antiquark. In the case of ambiguities, the implementation is to choose randomly. The CR algorithm applied later may anyhow change the initial colour topology, lessening the effect of the above choice.

At the level of the technical implementation, the choice of colour state for the scattered partons is transferred to the final state particles of the event using the LC structure of the MPI and PS. For example, if the two gluons are in the octet state, one of the LHE colour tags (see [5, 6]) is changed accordingly, and is propagated through to the colour of the final state particles.

As a special case the overall colour structure of the beam remnant is not allowed to be a singlet. This is to avoid double-counting between diffractive and non-diffractive events. In the DPS example, if the gluons form a colour singlet, they essentially make up (part of) a pomeron, and thus should fall under the single-diffractive description. It would be interesting to look into the interplay between MPI and diffraction in more detail using the colour-multiplet language developed here, but this would require its own dedicated study, beyond the scope of this work.

The conservation of flavour is relative straightforward and follows [87]. The principle used is to add the minimum needed flavour. For example if only an s quark is scattered from a proton, the remnant will consist of an \bar{s} plus the three valence quarks.

With the flavour structure and colour multiplet of the beam remnant known, it is now possible to calculate explicitly how many gluons need to be added to obtain a colour-singlet state. Again the idea is to add the minimum number of gluons to the beam remnant. Colour-junction structures, which we have argued can arise naturally in the colour structure of the scattered partons, complicate this calculation slightly. To achieve an overall colour-singlet state the number of junctions minus the number of antijunctions has to match that of the beam particle. Taking this into account the minimal number of gluons is given by

$$N_{\text{gluons}} = \max\left(0, \frac{(N_{\text{colour}} - N_{\text{quarks}} + \|N_{\text{junctions}} - N_{\text{antijunctions}} - b\|)}{2}\right); \tag{2.9}$$

where b is the beam baryon number (1 if the beam is a baryon, 0 if the beam is a meson, and -1 if the is an antibaryon). The division by two is due to the gluons carrying two colour lines. Without junctions, the number of gluons is simply the number of colour lines to the remnant minus the number of available quarks to connect those colour lines to. It is easiest to understand how junctions change this, by noting that the creation of a junction basically takes two colour charges and turn into one anticolour. Thus the number of required connections goes down by one for each additional junction needed.

After the gluons are added, all the colour connections and junction structures are assigned randomly between the remaining colours, with one exception: if the beam particle is a baryon and a junction needs to be constructed (similarly for an antibaryon and an antijunction), two of the valence quarks will be used to form the junction structure (possibly embedded in a diquark), if they have not already been scattered in the MPIs.

With finally the full parton structure known, including both flavours and explicit colours, the last step of the construction of the beam remnant is the assignment of energy fractions (x values) to each remnant parton, according to modified PDFs. To obtain overall energy-momentum conservation, the individual partons are scaled by an overall factor. The scaling becomes slightly more complicated by the introduction of primordial k_{\perp} . Details on the modified PDF versions and the scaling can be found in [87].

2.3.2 Colour flow in the whole event

As discussed above, the CR model is applied after the parton-shower evolution has finished (and after inclusion of the beam-remnant partons as described in the preceding subsection), just before the hadronisation. The model builds on two main principles: a simplified SU(3) structure of QCD, based on indices from 1 to 9, to tell which configurations are possible; and the potential energy of the resulting string systems, as measured by the so-called λ measure [23], to choose between the allowed configurations.

The starting point for the model is the LC configuration emerging from the showers + beam remnants. Thus between each LC-connected pair of partons a tentative dipole is constructed. This configuration is then changed by allowing two (or three) dipoles to reconnect, and this procedure is iterated until no more reconnections occur. In each step of the algorithm, four different types of reconnections can occur, illustrated in figure 7:

1. simple dipole-type reconnections involving two dipoles that exchange endpoints (figure 7(a));
2. two dipoles can form a junction-antijunction structure (figure 7(b));
3. three dipoles can form a junction-antijunction structure (figure 7(c));
4. two multi-parton string systems can form a junction and an antijunction at different points along the string and connect them via their gluons (figure 7(d)).

Note that, although mainly dipoles between quarks are shown in the illustrations, all dipoles ($q-\bar{q}$, $q-g$, $g-\bar{q}$ and $g-g$) are treated in the same manner in the implementation. Within an LC dipole, the quark and antiquark are assumed to be completely colour coherent, so

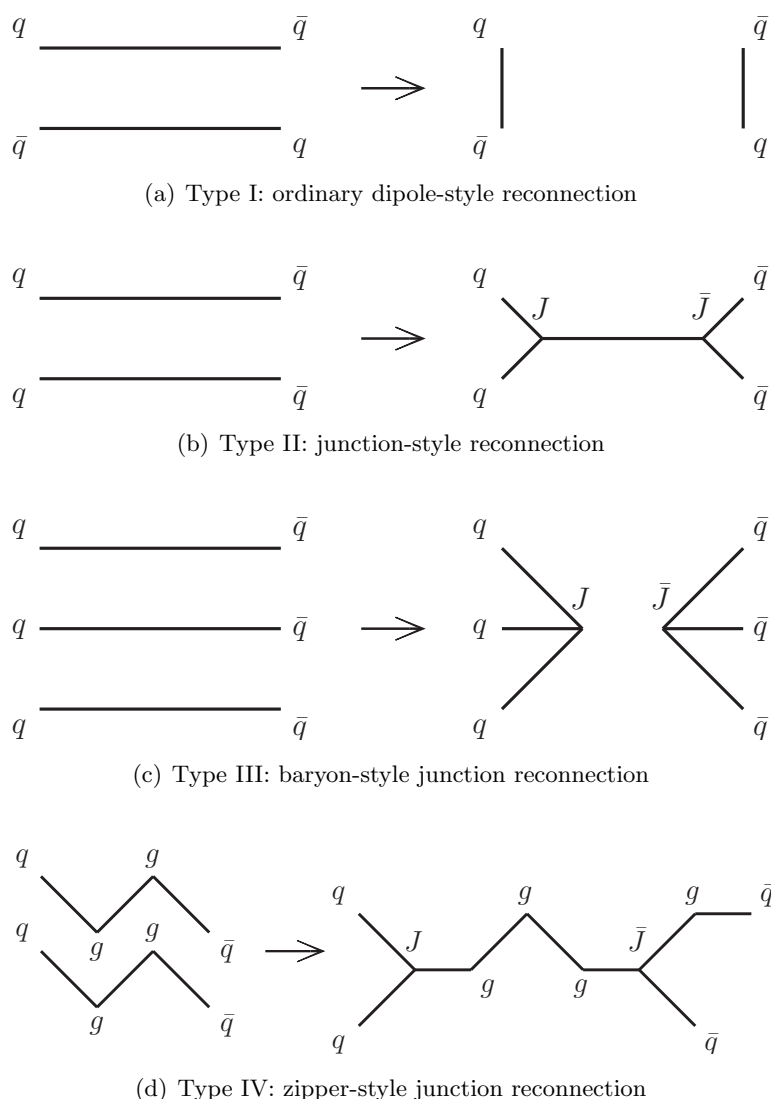


Figure 7. The four different allowed reconnection types. Type I (a) is the ordinary string reconnection. Type II (b) is the formation of a connected junction antijunction pair. Type III (c) is the formation of junction and antijunction, which are not directly connected. Type IV (d) is similar to type II except that it allows for gluons to be added between the two junctions.

that the probabilities for two dipoles to be in a colour-coherent state can be found by the standard SU(3) products. In full QCD, the probabilities for type I (dipole) and II (junction) reconnections for $q\bar{q}$ dipoles are given by eq. (2.4) and (2.5) as $P_I^{q\bar{q}} = 1/9$ and $P_{II}^{q\bar{q}} = 1/3$, respectively. For gg dipoles, the calculation is complicated slightly by the fact that eq. (2.1) takes into account both the colour and anticolour charges of both of the gluons. With a probability of $P_I^{gg} = 8/64 = 1/8$ each, either “side” (colour or anticolour) of the gluons are allowed to reconnect (for a $1/64$ probability that CR is allowed on both sides). And with a total probability of $P_{II}^{gg} = 20/64 = 5/16$ either one *or* the other side is allowed a junction-type reconnection (both sides would be equivalent to a dipole-style reconnection already

counted above). For simplicity, the index rules described in the beginning of this section have been defined to treat $q\bar{q}$, qg , and $g\bar{q}$ dipoles all on an equal footing. The result is a compromise of $P_I = 1/9$ for all dipole-type reconnections and $P_{II} = 2/9$ for all junction-type reconnections. Differences between $q\bar{q}$ and $g\bar{q}$ combinations still arise due to gluons being prevented from having the same colour and anticolour indices, and since the combination of two type-II reconnections is equivalent to a type-I reconnection. A comparison between the weights resulting from our simplified treatment and the multiplet weights in full QCD for the simple case of two-parton combinations was illustrated in figure 3. Note also that the probability for a type-III reconnection among three uncorrelated $q\bar{q}$ dipoles (essentially creating a baryon from three uncorrelated quarks) is $P_{III}^{qqq} = 1/27$ in full QCD (among the 27 different ways to combine 3 quarks, only one is a singlet) while it is only $2/3$ as large in our model; $P_{III} = 2/9 \times 1/9 = 2/81$. Although our model should be a significant step in the right direction, we therefore still expect a tendency to underestimate baryon production. When tuning the model below, we shall see that we are able to compensate for this by letting junction-type reconnections appear somewhat more energetically favourable compared with dipole-type reconnections.

To recapitulate the model implementation, each dipole is assigned a random index value between 1 and 9. Two dipoles are allowed to do a type-I reconnection if the two numbers are equal, providing the $1/9$ probability. Type-II reconnections are allowed if the two numbers modulo three are equal and the indices are different (e.g. 1-4 and 1-7), thereby providing the $2/9$ probability. Three dipoles are allowed to do the type-III reconnection if they all have the same index modulo three and are all different (1-4-7). The type-IV reconnection follows the same principles: a dipole needs to have the same index, and a junction needs to have different-but-equal-under-modulo-three index. The exact probability for type-IV reconnections depend on the number of gluons in the string.

The number of allowed colour indices can in principle be changed (the 9 above), e.g. to vary the strength of CR. However, the type-II, -III, and -IV reconnections rely on the use of modulo, thus care should be taken if junction formation is allowed. A different method to control the strength of the CR will be discussed below.

The above colour considerations only tell which new colour configurations are *allowed* and not whether they are *preferable*. To determine this, we invoke a minimisation of the λ string-length measure. The λ measure can be interpreted as the potential energy of a string, more detailed it is the area spanned by the string prior to hadronisation. It is closely connected to the total rapidity span of the string, and thereby also its total particle production. The minimisation is carried out by only allowing reconnections that lower the λ measure, which ensures that a local minimum is reached.

A further complication is that, while the λ -measure for a quark-antiquark system with any number of gluon kinks in between is neatly defined by an iterative procedure [23], the measure defined there did not include junction structures. The first extension to handle these were achieved by starting from the simple measure between a quark antiquark dipole [100]:

$$\lambda^{q\bar{q}} = \ln \left(1 + \frac{s_{q\bar{q}}}{2m_0^2} \right) \tag{2.10}$$

where $s_{q\bar{q}}$ is the dipole mass squared and m_0 is a constant with dimensions of energy, of order Λ_{QCD} . For high dipole masses, the “1” in eq. (2.10) can be neglected, splitting the λ -measure neatly into two parts: one from the quark and one from the antiquark end (in the dipole rest frame):

$$\lambda^{q\bar{q}} \xrightarrow{s \gg m_0^2} \ln\left(\frac{s}{2m_0^2}\right) = \ln\frac{\sqrt{2}E_q}{m_0} + \ln\frac{\sqrt{2}E_{\bar{q}}}{m_0}. \quad (2.11)$$

The extension to handle a junction system used the same method, going to the junction rest-frame and adding up the “ λ -measures” from all the three (anti-)quark ends. The end result became

$$\lambda^{q_1q_2q_3} = \ln\frac{\sqrt{2}E_1}{m_0} + \ln\frac{\sqrt{2}E_2}{m_0} + \ln\frac{\sqrt{2}E_3}{m_0} \quad (2.12)$$

where the energies are calculated in the junction rest frame.⁵ This procedure worked well in the scenarios considered in that study, since all the dipoles had a relative large mass. However, in the context of our CR model, we will often be considering dipoles that have quite small masses. In that case, continuing to ignore the “1” in eq. (2.10) can lead to arbitrarily large negative λ measures. Among other things, such a behaviour could allow soft particles with vanishing string lengths to have a disproportionately large impact on dipoles with a large invariant mass. Generalising this behaviour to soft junction structures results in similar effects, namely that soft particles can have a disproportionately large effect.

An alternative measure is here proposed to remove the problem with negative string lengths,

$$\lambda' = \ln\left(1 + \frac{\sqrt{2}E_1}{m_0}\right) + \ln\left(1 + \frac{\sqrt{2}E_2}{m_0}\right) \quad (2.13)$$

where the energies are calculated in the rest-frame of the dipoles. This measure is always positive definite. In the case of massless particles the λ' -measure can be rewritten to

$$\lambda' = \ln\left(1 + \frac{s}{2m_0^2} + \frac{\sqrt{2s}}{m_0}\right) \quad (2.14)$$

where again s is the invariant mass squared of the dipole and m_0 is a constant. The two measures agree in the limit of large invariant masses ($s \gg m_0$). The implementation includes a few alternative measures as options, but the above is chosen as the default measure and therefore also the one that the parameters are tuned for.

A final complication regarding the λ measure is that the form above cannot be used to describe the distance between two directly connected junctions. Instead the same measure as described in [100] is also used in this study ($\lambda = \beta_1\beta_2 + \sqrt{(\beta_1\beta_2)^2 - 1}$, where β_1 and β_2 are the 4-velocities of the two junction systems).

Since the λ -measure for junctions introduces additional approximations, a tuneable parameter is added to control the junction production. Several options for this parameter are possible and we settled on a $m_{0j} \neq m_0$ in the λ -measure for junctions. A higher m_{0j}

⁵Note: we use a slightly different definition of m_0 here compared to the original paper [100].

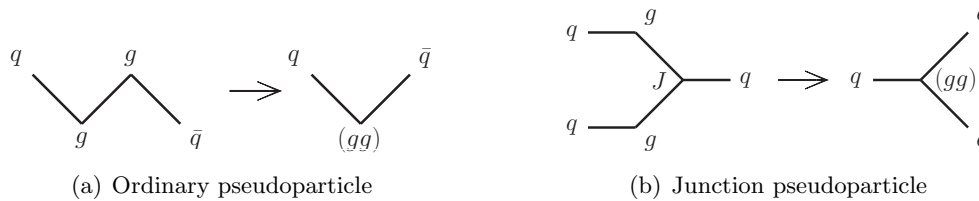


Figure 8. The figure shows how two gluons are turned into a pseudoparticle depending on whether they are connected via an ordinary string (a) or a junction (b). The (gg) represents the formed pseudoparticle.

means a lower λ measure, resulting in an enhancement of the junction production. We cast the free parameter as the ratio,

$$\text{junctionCorrection} : C_j = m_{0j}/m_0, \quad (2.15)$$

thus a value C_j above unity indicates an enhancement in junction production, and vice versa. The possibility of a junction enhancement can be seen as providing a crude mechanism to compensate for the intrinsic suppression of junction topologies in the colour-space model. Indeed, in the section on tuning below we find that values above unity are preferred in order to fit the observed amounts of baryon production.

In the context of CR, it is generally the dipoles with the largest invariant masses which are the most interesting; they are the ones for which reconnections can produce the largest reductions of the λ measure. However, as evident from the above discussion, dipoles with *small* invariant masses can actually be the most technically problematic to deal with. It was therefore decided to remove dipoles with an invariant mass below m_0 from the colour reconnection. Technically this is achieved by combining the small-mass pair into a new pseudo-particle. For an ordinary dipole this is a trivial task (figure 8(a)), but if the dipole is connected to a junction the technical aspects becomes more complicated (figure 8(b)). The easiest way to think of this is as an ordinary diquark, but in addition to these we can have digluons, which will have three ordinary (anti-)colour tags. Note that we do not intend these to represent any sort of weakly bound state; we merely use them to represent a low-invariant-mass collection of partons whose internal structure we consider uninteresting for the purpose of CR. The pseudo-particles are formed after the LC dipoles are formed, and also after any colour reconnections if the new dipoles have a mass below m_0 . Increasing m_0 will therefore lower the amount of CR. Only small effects occur for variations around the Λ_{QCD} scale, however increasing m_0 beyond 1 GeV introduces a significant reduction of CR.

The complete algorithm for the colour reconnection can be summarised as below.

1. Form dipoles from the LC configuration.
2. Make pseudoparticles of all dipoles with mass below m_0 .
3. Minimise λ -measure by normal string reconnections.
4. Minimise λ -measure by junction reconnections.
5. If any junction reconnections happened return to point 3.

The choice to first do the normal string reconnections before trying to form any junctions is due to the algorithm not allowing to remove any junction pairs.

Since each reconnection is required to result in a lower λ -measure than the previous one, the minimisation procedure is only expected to reach a local minimum. A possible extension to reach the global minimum would be to use simulated annealing [110]. This is, e.g., the approach adopted in the HERWIG++ CR model [47]. However this would also require the implementation of inverse reconnections (i.e. a junction and an antijunction collapsing to form strings, and the unfolding of pseudo-particles.). Secondly the computational time needed to find the global minimum would slow down the event generation speed very significantly. For purposes of this implementation, we therefore restrict ourselves to a local deterministic minimisation here, noting that an algorithm capable of reproducing the full expected area-law exponential would be a desirable future refinement.

2.3.3 Hadronisation of multi-junction systems

The existing junction hadronisation model [87] was developed mainly for the case of string systems containing a single junction (in the context of baryon-number violating SUSY decays like $\tilde{\chi}^0 \rightarrow qq\bar{q}$). For such systems, the strategy of is to take the two legs with lowest energy in the junction rest-frame and hadronise them from their respective quark ends inwards towards the junction, until a (low) energy threshold is reached, at which point the two endpoints are combined into a diquark (which contains the junction inside). This diquark then becomes the new endpoint of the last string piece, which can then be fragmented as usual.

The case of a junction-antijunction system was also addressed in [87] (arising e.g., in the case of $e^+e^- \rightarrow \tilde{t}\tilde{t}^* \rightarrow \bar{q}\bar{q}qq$), but the new treatment of the beam remnants presented here, as well as the new CR model, can produce configurations with any number of colour-connected junctions and antijunctions. This goes beyond what the existing model can handle.

The systems of equations describing such arbitrarily complicated string topologies are likely to be quite involved, with associated risks of instabilities and pathological cases. Rather than attempting to address these issues in full gory detail, we here adopt a simple “divide-and-conquer” strategy, slicing the full system into individual pieces that contain only one junction each, via the following 3 steps:

- If a junction and an antijunction are connected with a single gluon between them, that gluon is forced to split into two light quarks (u,d and s) that each equally share the 4-momentum of the gluon (corresponding to $z = \frac{1}{2}$). Since the gluon is massless, the two quarks will have to be parallel (figure 9(a)).
- If a junction and an antijunction are connected with at least 2 gluons in between, the gluon pair with the highest invariant mass is found, and is split according to the string-fragmentation function. The highest invariant mass is chosen due to it having the largest phase space and being the most likely to have a string breakup occur. The split is done in such a way that the two gluons are preserved but each of them

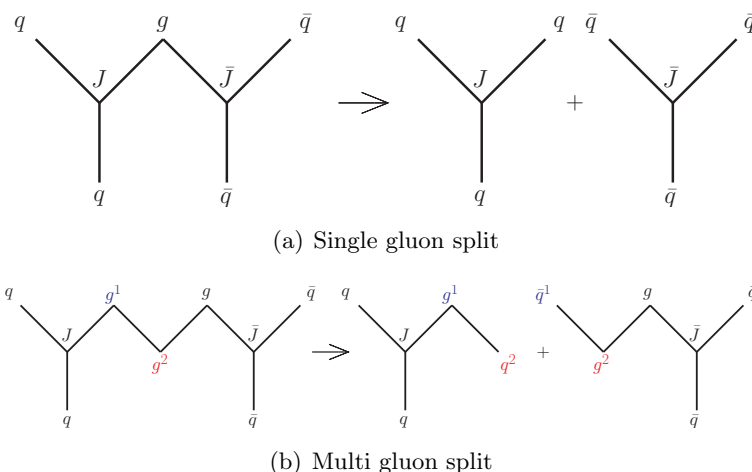


Figure 9. The figure shows how connected junctions are separated if they are connected by respectively a single gluon (a) or multiple gluons (b). The indices indicate where the split happens and which particles each gluon splits into.

give up part of their 4-momentum to the new quark pair. The new quark that is colour-connected to one of the gluons will be parallel to the other gluon (figure 9(b), where the indices indicate who is parallel with whom.).

- After the two rules above have been applied, only directly-connected junction-antijunctions are left. If all three legs of both junctions are connected to each other, the system contains no partons and can be thrown away. If two of the legs are directly connected, the junction-antijunction system is equivalent to a single string piece and is replaced by such, see figure 10(a). Finally, the case of a single direct junction-antijunction connection is dealt with differently, depending on whether the system contains further junction-antijunction connections or not. In the former case, illustrated in figure 10(b), the maximum number of junctions are formed from the partons directly connected to the junction system. The remaining particles are formed into normal strings. In the example of figure 10(b), three quarks are first removed to form a junction system; the remaining q and \bar{q} then have no option but to form a normal string. The current method randomly selects which outgoing particles to connect with junctions. One extension would be to use the string measure to decide who combines with whom. (However the effect of this might be smaller than expected, since the majority of the multi-junction configuration comes from the beam remnant treatment, which later undergoes CR.)

For cases with a single direct junction-antijunction string piece and no further junctions in the system, illustrated in figure 10(c), the λ -measure is used to determine whether the two junctions should annihilate or be kept [87] (essentially by determining whether the strings pulling on the two junctions cause them to move towards each other, towards annihilation, or away from each other). If the junctions survive, a new $q\bar{q}$ pair is formed by taking momentum from the other legs of the junction. Otherwise the junction topology is replaced by two ordinary strings. An option to always keep the junctions also exists.

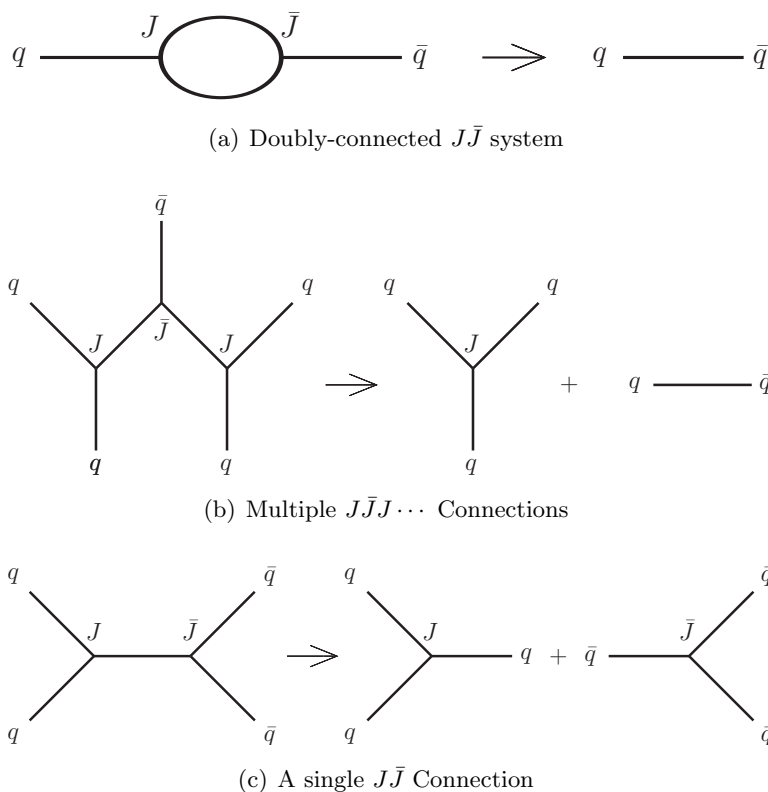


Figure 10. The figure shows how directly connected junctions are separated. Figure (a) shows the replacement of a doubly-connected junction-antijunction system by an ordinary string piece. Figure (b) shows the method used to reduce systems with more than two interconnected junctions. Figure (c) shows the split-up of a system containing exactly one junction-antijunction connection, into two separate junction systems.

2.3.4 Space-time structure

By default, we do not account for any space-time separation between different MPI systems. This is motivated by the observation that, physically, the individual MPI vertices can at most be separated by transverse distances of order the proton radius, which by definition is small compared with the length of any string long enough to fragment into multiple particles.

We do note, however, that in order for reconnections to occur between two string pieces, they should be in *causal contact*; if either string has already hadronised before the other forms, there is no space-time region in which reconnections between them could physically occur. In the rest frame of a hadronising string piece, we take the formation time of the corresponding QCD dipole to be given roughly by the inverse of its invariant mass, $\tau_{\text{form}} \sim 1/m_{\text{string}}$. Alternative measures (e.g., the k_{\perp} evolution variable of the PS) could also have been used, and to allow at least a range of variations of the exact definition, a free parameter is introduced. The time at which the string piece begins to hadronise is related to the inverse of Λ_{QCD} , $\tau_{\text{had}} \sim 1/\Lambda_{\text{QCD}}$. In order for reconnections to be possible between two string pieces, we require that they must be able to resolve each other during the time

between formation and hadronisation, taking time-dilation effects caused by relative boosts into account. There are several ways in which this requirement can be formulated at the technical level, and accordingly we have implemented a few different options in the code. In principle, the two strings can be defined to be in causal contact if the relative boost parameter fulfils:

$$\gamma \tau_{\text{form}} < C_{\text{time}} \tau_{\text{had}} \Rightarrow \frac{\gamma c}{m_{\text{string}} r_{\text{had}}} < C_{\text{time}} \quad (2.16)$$

where C_{time} is a tuneable parameter and $r_{\text{had}} (= \tau_{\text{had}} c \equiv 1 \text{ fm})$ is a fixed constant given by the typical hadronisation scale. There are however two major problems with this definition: first it is not Lorentz invariant; the two dipoles will not always agree on whether they are in causal contact or not. This can be circumvented, by either requiring both to be able to resolve each other (strict) or just either of them to be able to resolve the other (loose). Secondly, the emission of a soft gluon from an otherwise high-mass string changes m_{string} significantly for each of the produced string pieces, which gives an undesirable infrared sensitivity to this measure, reminiscent of the problems associated with defining the λ string-length measure itself. One way to avoid this problem is to consider the *first* formation time of each colour line, i.e. the dipole mass at the time the corresponding colour line was first created in the shower, which we have implemented as an alternative option. No matter the exact definition of formation time and hadronisation time, all models agree that reconnection between boosted strings should be suppressed. A final extremely simple way to capture this in a Lorentz-invariant way is to apply a cut-off directly on the boost factor γ , which thus provides a simple alternative to the other models.

These different methods have all been implemented and are available in PYTHIA, via the mode `ColourReconnection:timeDilationMode`. The C_{time} parameter introduced above is specified by `ColourReconnection:timeDilationPar` and controls the size of the allowed relative boost factor for reconnections to occur. As such it can be used to tune the amount of CR. Its optimal value will vary depending on the method used, but after the methods are tuned they produce similar results (see section 3 for details).

A final aspect related to space-time structure that deserves special mention is resonance decays. By default, these are treated separately from the rest of the event. Physically, this is well motivated for longer-lived particles (e.g., Higgs bosons), which are expected to decay and hadronise separately. For shorter-lived resonances the separation of the MPI systems and resonance decays is physically not so well motivated. E.g., most Z/W bosons and top quarks will decay before hadronisation takes place, $\Gamma \gg \Lambda_{\text{QCD}}$, and as such should be allowed to interact with the particles from the MPI systems, ideally with a slightly suppressed probability due to the decay time.

Currently, only two extreme cases are implemented, corresponding to letting CR occur *before* or *after* (all) resonance decays. The corresponding flag in PYTHIA is called `PartonLevel:earlyResDec`. When switched on, CR is performed *after* all resonance decays have occurred, and all final-state partons therefore participate fully in the CR. Since no suppression with resonance lifetime is applied, this gauges the largest possible impact on resonance decays from CR. When switched off, CR is performed *before* resonance decays, hence involving only the beam remnant and MPI systems. It is equivalent to

assuming an infinite lifetime for the resonances, and hence estimates the smallest possible impact on resonance decays from CR. An optional additional CR can be performed between the decay products of the resonance decays, with the physics motivation being $H \rightarrow WW \rightarrow q\bar{q}q\bar{q}$ studies.

To summarise, we acknowledge that the treatment of space-time separation effects and causality is still rather primitive in this model. The derivation of a more detailed formalism for these aspects would therefore be a welcome and interesting future development.

3 Constraints and tuning

The tuning scheme follows the same procedure as for the Monash 2013 tune [34]. However at a more limited scope, since only CR parameters, and ones strongly correlated with them, are tuned. As a natural consequence of this, the Monash tune was chosen as the baseline. As discussed in section 2.3.4, several options are available for the choice of CR time-dilation method, which naturally results in slightly different preferred parameter sets. Here, we consider the following three modes:

- Mode 0: no time-dilation constraints. m_0 controls the amount of CR (mode 0);
- Mode 2: time dilation using the boost factor obtained from the final-state mass of the dipoles, requiring all dipoles involved in a reconnection to be causally connected (strict);
- Mode 3: time dilation as in Mode 2, but requiring only a single connection to be causally connected (loose).

This allows to investigate the consequences of some of the ambiguities in the implementation of the model. For the purpose of later studies that may want to focus on a single model, we suggest to use mode 2 as the “standard” one for the new CR. The parameters described in this section will therefore correspond to that particular model, with parameters for the others given in appendix A. Note that this section only contains the main physical parameters; for a complete list we again refer to appendix A.

3.1 Lepton colliders

We begin with e^+e^- collisions. Only small effects are expected in this environment, due to the p_\perp -ordering of the shower and the absence of MPIs. Only CR and string-fragmentation variables were studied, since the shower was left untouched. The fragmentation model contains three main parameters governing the kinematics of the produced hadrons: the non-perturbative p_\perp produced in string breaks, controlled by the σ_\perp parameter (`StringPT:sigma`), and the two parameters, a and b , which control the shape of the longitudinal (z) fragmentation function. For pedagogical descriptions, see e.g. [2, 23, 30, 34]. Since the effects are expected to be small, we made the choice of keeping $\sigma_\perp = 0.335$ GeV unchanged, adjusting only the longitudinal (a and b) parameters. Changing the minimal number of parameters also helps to disentangle the effects of CR from the retuning. As a verification, a tune with a smaller σ_\perp (0.305 GeV) was considered, however after retuning a and b the two tunes described the LEP data with a similar fidelity. (The choice of testing

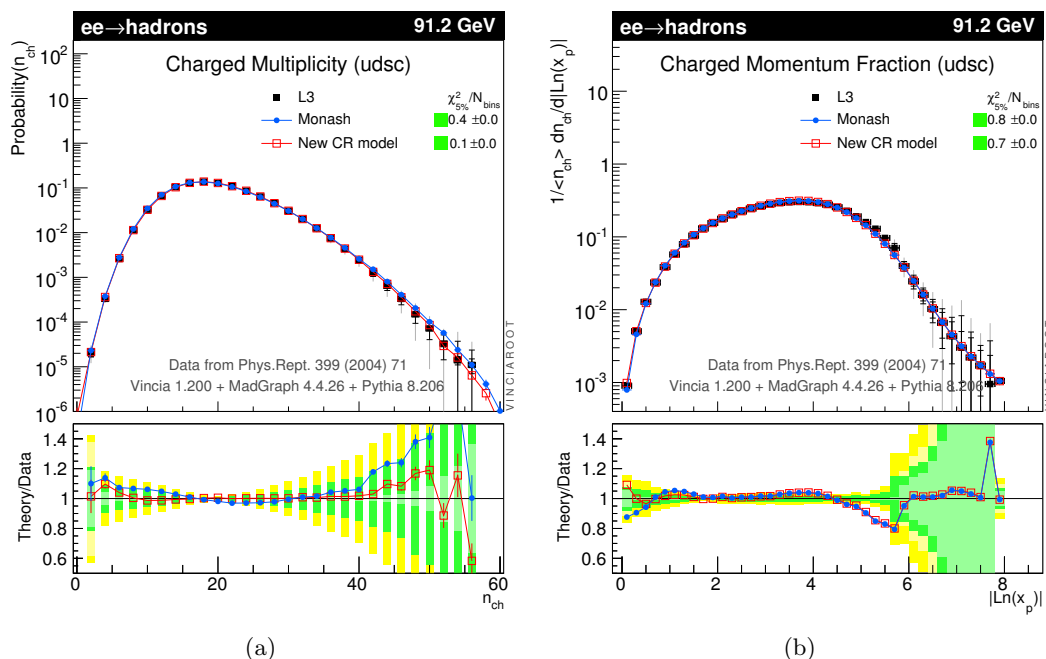


Figure 11. Charged-particle multiplicity (a) and momentum fraction (b) spectra, in light-flavour tagged data from the L3 collaboration [111]. (Plots made with VINCIAROOT [112]. The ratio panes follow the now-standard “Brazilian” colour conventions, with outer (yellow) bands corresponding to 2σ deviations and inner (green) bands corresponding to 1σ deviations.)

a lower σ_{\perp} was made since the CR model tends connect more collinear partons leading to shorter strings, but a harder p_{\perp} spectrum of the produced hadrons [75].)

The determination of the two parameters of the Lund fragmentation function, a and b , is complicated slightly by the fact that they are highly correlated; choosing both of them to be quite small often produces equally good descriptions of fragmentation spectra as choosing both of them large, corresponding to a relatively elongated and correlated χ^2 “valley”. By simultaneously considering both variables and comparing them to both multiplicity and momentum spectra, cf. figure 11 (with the “New CR model” curve showing our new model, and “Monash” the baseline Monash 2013 tune), we here settled on a low-valued pair, as compared with the default Monash values:

```
StringZ:aLund = 0.38 # was 0.68
StringZ:bLund = 0.64 # was 0.98
```

The new CR also alters the ratio between the identified-particle yields, especially so for baryon production due to the introduction of additional junctions. Thus the flavour-selection parameters of the string model also need to be retuned, by comparing with the total identified-particle yields, see e.g. [34]. As expected the effects are minimal in e^+e^- collisions, and only small changes are required. The modifications were therefore done with a view to providing a better description also for pp colliders, but staying within the uncertainties allowed by the LEP data. This resulted in an adjustment of the parameters

for the diquark over quark fragmentation probability and the strangeness suppression:

```
StringFlav:probQQtoQ = 0.078 # was 0.081
StringFlav:probStoUD = 0.2 # was 0.217
```

As expected the diquark over quark probability is reduced due to the introduction of junctions. More surprisingly is the increased suppression of strange quarks, since the model a priori should not influence flavour selection. The technical implementation of the junction hadronisation does, however, introduce a slight enhancement of the strangeness production, due to an even probability for a gluon to split into an u, d or s quarks when separating junction systems. This is not visible at LEP, but at pp colliders the slightly lower strangeness fragmentation is favoured.

The final set of fragmentation parameters we define is more technical. For junction systems and beam remnants, a separate set of parameters controls the choice of total spin when two, already produced, quarks are combined into a diquark. Unlike diquarks produced by ordinary string breaks (whose spin is controlled by the parameter `StringFlav:probQQ1toQQ0`), which can only contain the light quark flavours (u, d, s), and for which the significant mass splittings between the light-flavour spin-3/2 and spin-1/2 baryon multiplets necessitates a rather strong suppression of spin-1 diquark production (relative to the naive factor 3 enhancement from spin counting), junction systems in particular can allow the formation of baryons involving heavy flavours, which have smaller mass splittings and which therefore might require less suppression of spin-1 diquarks. We note also that diquarks produced in string breakups are produced within the linear confinement of the string, whereas junction diquarks come from the combination of two already uncorrelated quarks, so there is a priori little physics reason to assume the parameters must be identical.

With the limited amount of junctions in the old model, none for ee and at most two for pp , these parameters previously had almost no influence on measurable observables and were therefore largely irrelevant for tuning. With the additional junctions produced by our model, these parameters can now give larger effects. Measurements of higher-spin and heavy-flavour baryon states at pp colliders are still rather limited though, and so far we are not aware of published directly usable constraints from experiments. For the time being therefore, we choose to fix the parameters to be identical to those for the production of ordinary diquarks in string breakups:

```
StringFlav:probQQ1toQQ0join = 0.027,0.027,0.027,0.027
```

The four components give the suppression when the heaviest quark is u/d, s, c or b, respectively. We stress that this is merely a starting point, hopefully to be revised soon by comparisons with new data from the LHC experiments.

3.2 Hadron colliders

The retuning to hadron colliders consisted of tuning three main parameters:

- C_{time} (`ColourReconnection:timeDilationPar`): controls the overall strength of the colour-reconnection effect via suppression of high-boost reconnections, see section 2.3.4. Can be tuned to the $\langle p_{\perp} \rangle$ vs n_{ch} distribution.

- C_j (`ColourReconnection:junctionCorrection`): multiplicative factor, m_{0j}/m_0 , applied to the string-length measure for junction systems, thereby enhancing or suppressing the likelihood of junction reconnections. Controls the junction component of the baryon to meson fraction and is tuned to the Λ/K_s^0 ratio.
- p_{\perp}^{ref} (`MultiPartonInteractions:pT0Ref`): lower (infrared) regularisation scale of the MPI framework. Controls the amount of low p_{\perp} MPIs and is therefore closely related to the total multiplicity and can be tuned to the $d\langle n_{\text{ch}} \rangle/d\eta$ distribution.

By iteratively fitting each parameter to its respective most sensitive curve an overall good agreement with data was achieved (see figure 12) with the following parameters:

```
ColourReconnection:junctionCorrection = 1.2 # new parameter
ColourReconnection:timeDilationPar   = 0.18 # new parameter
MultiPartonInteractions:pT0Ref      = 2.15 # was 2.28
```

Note in particular this is the first time that PYTHIA has been able to describe the Λ/K_s ratio in pp collisions while remaining consistent with LEP bounds. We explore this in more detail in section 3.4.

The $C_j = 1.2$ parameter shows that a slight enhancement of junction reconnections (i.e., baryon production) is needed, relative to mesonic ones. However, given the approximations used in the implementation of especially the junction structures, such a difference is not unreasonable. Small differences between the modes can be seen in the $\langle p_{\perp} \rangle$ vs n_{ch} and more significant differences for multiplicity distributions, cf. figure 13. With respect to the latter, however, we note that the differences in the tails of the multiplicity distributions can be tuned away by modifying the assumed transverse matter density profile of the proton, which was kept fixed here to highlight the differences with the minimal number of retuned parameters.

The new colour treatment of the beam remnant (BR) introduces a single new parameter controlling the amount of saturation, cf. the discussion in section 2.3.1. Due to the low p_{\perp} of the BR particles, the effects are largest in the forward direction. We therefore use the forward charged multiplicity as measured by the TOTEM experiment [116] to compare different modelling choices of this aspect, see figure 14(a). The difference between no saturation⁶ ($k_{\text{saturation}} \rightarrow \infty$) and maximal saturation ($k_{\text{saturation}} = 0.1$) is about 10% and exhibits no shape difference over the TOTEM pseudo rapidity range. For illustration and completeness, we may also consider what happens over the full rapidity range, at least at the theory level. This is illustrated in figure 14(b). In the central region, the effect of applying saturation is a slight decrease of the particle yield, and thus would already have been tuned away by p_{\perp}^{ref} . It was therefore chosen to use a relative high saturation to mimic the effect of the earlier PYTHIA beam remnant model:

```
BeamRemnants:saturation = 5
```

We emphasise however that this is merely a starting point, and that a different balance between p_{\perp}^{ref} and $k_{\text{saturation}}$ may be preferred in future tuning efforts, especially ones taking

⁶Technically: `BeamRemnants:saturation = 1E9`.

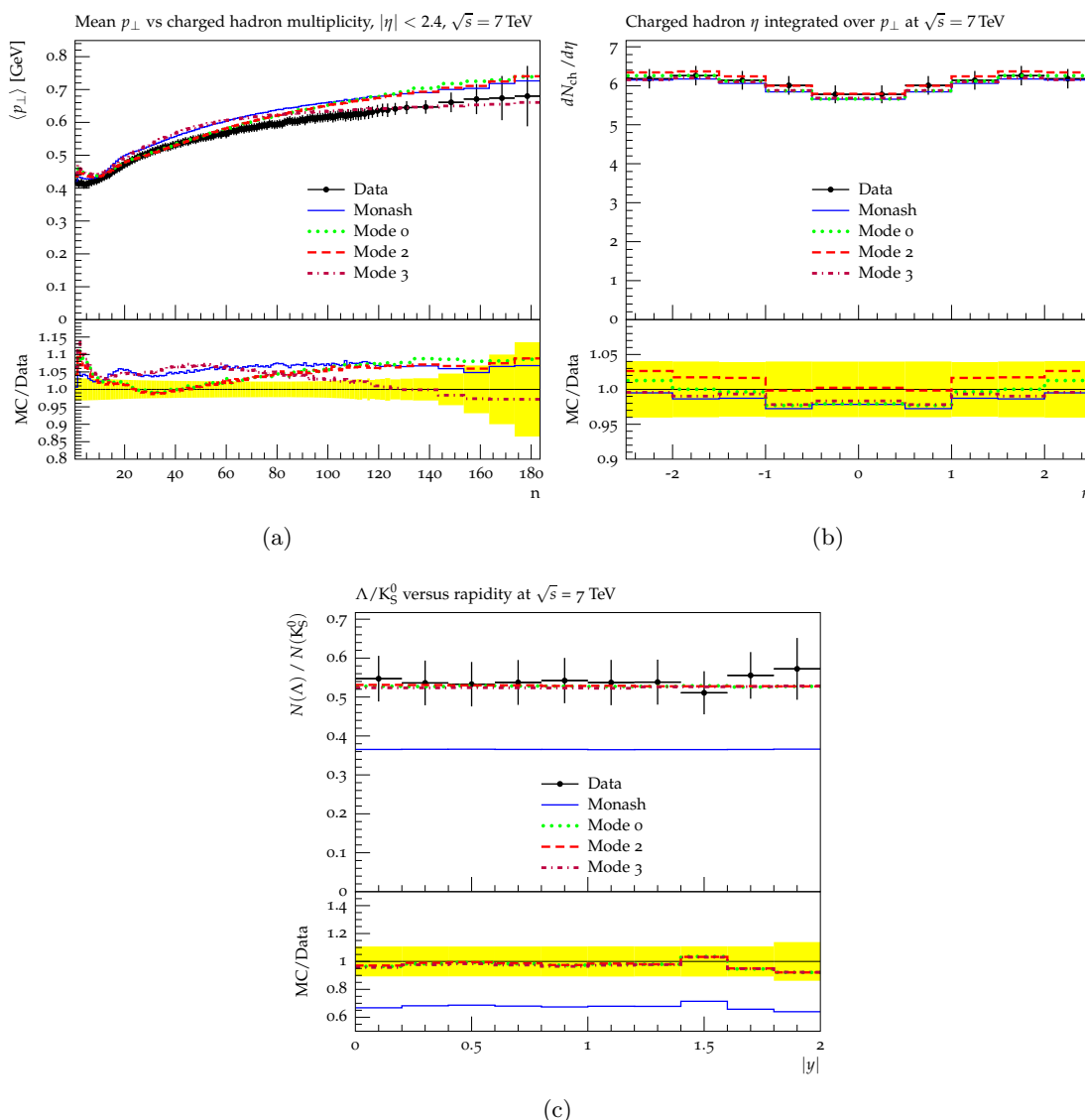


Figure 12. The average p_{\perp} as a function of multiplicity [52] (a), the average charged multiplicity as a function of pseudorapidity [113] (b), and the Λ/K_s ratio [114] (c). All observables from the CMS collaboration and plotted with the Rivet framework [115]. All PYTHIA simulations were non single diffractive (NSD) with a lifetime cut-off $\tau_{\max} = 10$ mm/ c and no p_{\perp} cuts applied to the final state particles. The yellow error band represents the experimental 1σ deviation.

a more dedicated look at the forward region. In such a study the sharing of momentum between the partons in the remnant should also be considered, since it is known to alter both particle production in the forward region and the multiplicity distributions.

An interesting signal that may help to break the relative degeneracy between p_{\perp}^{ref} and $k_{\text{saturation}}$, is to look for baryons at high rapidities, which, due to the introduction of junction structures in the BR can act as further tracers of the degree to which the BR has been disturbed. This is illustrated by the lower set of curves in figure 14(b). The effects

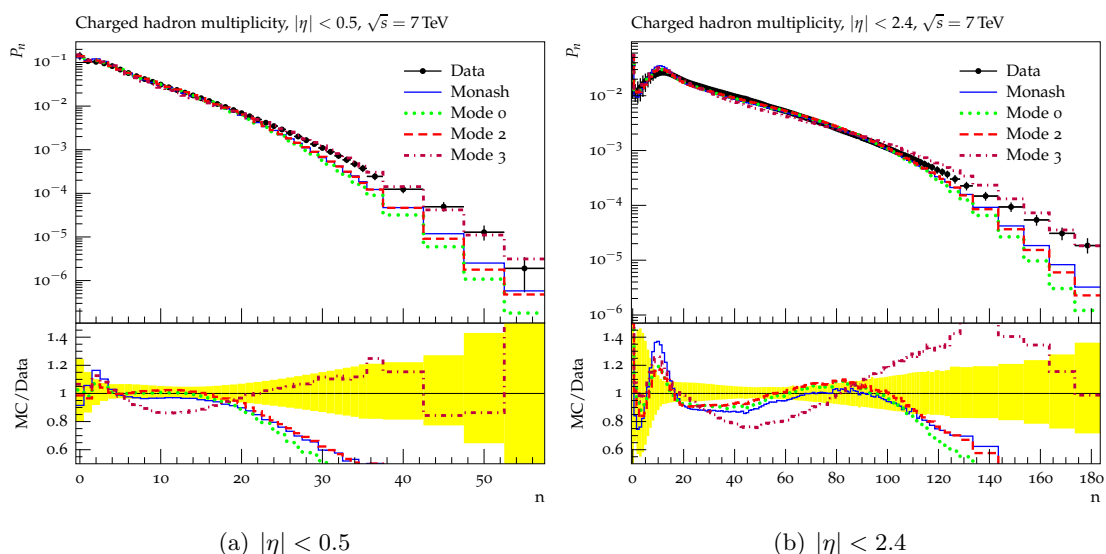


Figure 13. The two plots show the multiplicity distributions for respectively very central tracks (a) and the full CMS tracker coverage (b), compared with CMS data [52]. All PYTHIA simulations were NSD with a lifetime cut-off ($\tau_{\max} = 10$ mm/c) and no p_{\perp} cuts were applied to the final state particles. The yellow error band represents the experimental 1σ deviation.

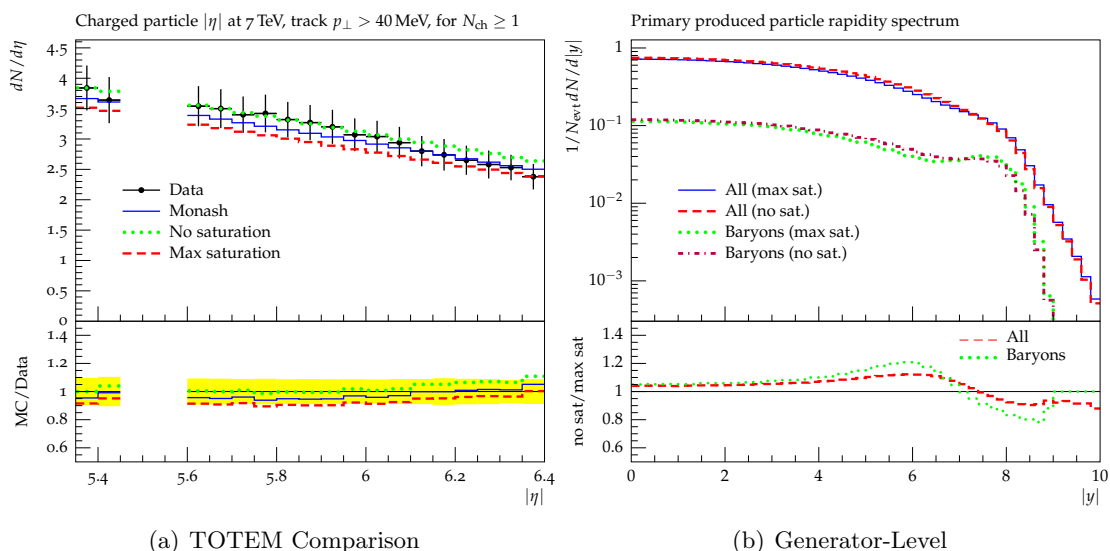


Figure 14. (a): different extreme saturation choices compared with the TOTEM forward multiplicity data [116]. (The PYTHIA simulation includes all soft-QCD processes and a particle lifetime cut-off $\tau_{\max} = 10$ mm/c.) The yellow error band represents the experimental 1σ deviation. (b): MC rapidity distributions for respectively all particles and baryons only. (For simplicity only non diffractive (ND) events were used, hadron decays were turned off to reflect primary hadron production, and no p_{\perp} cuts were imposed.)

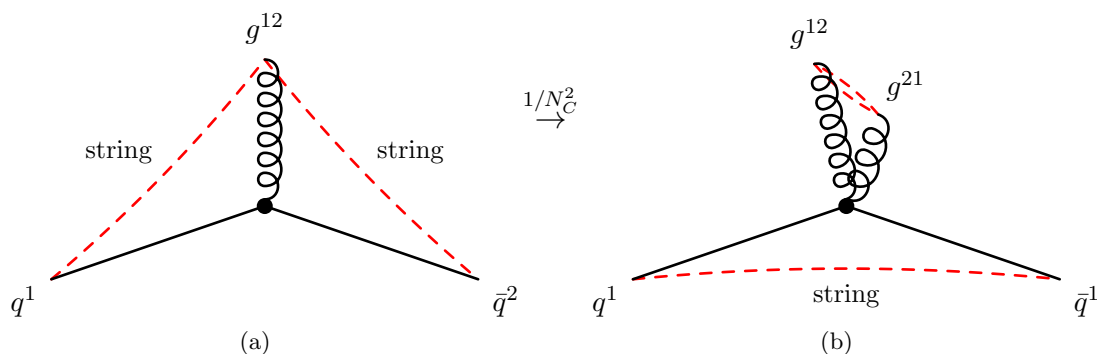


Figure 15. (a): illustration of the correspondence between colour connections and string pieces in an ordinary (LC) 3-jet topology. (b): if the gluon jet is composed of (at least) two gluons, there is probability for the $q\bar{q}$ system to be in an overall singlet. (The LC string topology remains a possibility as well, with the string-length measure λ used to decide between them.) Notation: g^{ac} denotes a gluon with colour (anticolour) index c (a). For (anti)quarks, we use $q^c \equiv q^{0c}$ and $\bar{q}^a \equiv \bar{q}^{a0}$.

are indeed seen to be slightly larger for baryons, however the total cross section is also significantly lower. From this simple MC study we are not able to say clearly whether such a measurement, which requires the additional non-trivial ingredient of baryon identification, would be experimentally feasible.

3.3 Direct CR constraints?

In the preceding sections, we constrained (“retuned”) the fragmentation parameters using observables like the charged-particle multiplicity and fragmentation spectrum, which are indirectly sensitive to CR effects via the modifications caused to these spectra by the minimisation of string lengths. But what about observables with more direct sensitivity to CR effects? There are two main categories of dedicated CR studies at LEP: fully hadronically decaying WW events (looking for reconnections between the two W systems), and colour-flow sensitive observables in three-jet events. In this study we restrict our attention to the latter of these. A follow-up dedicated study of CR effects at e^+e^- colliders, both earlier as well as possible future colliders, is planned.

Without CR, the three produced jets will in general be represented by a colour string stretched from the quark via the gluon to the antiquark, illustrated in figure 15(a). The string pieces spanned between the quark and gluon jets lead to a relatively large particle production between these jets, while there is a suppression of particle production in the phase-space region directly between the quark and antiquark jets. However, if CR is allowed, there is a $1/N_C^2$ chance that two (or more) sequentially emitted gluons end up cancelling each others’ colour charges. Thus, if at least one additional gluon was produced in the FSR, the “gluon jet” may effectively become overall colour neutral, allowing it to decouple from the quark-antiquark system in colour space. This is illustrated in figure 15(b). There is a caveat to the above, namely, if the two gluons originate from a single gluon, i.e., $g \rightarrow gg$, the two gluons must form a colour octet. In the gluon-collinear limit, this colour structure dominates and the probability for the jet to end up colour neutral should

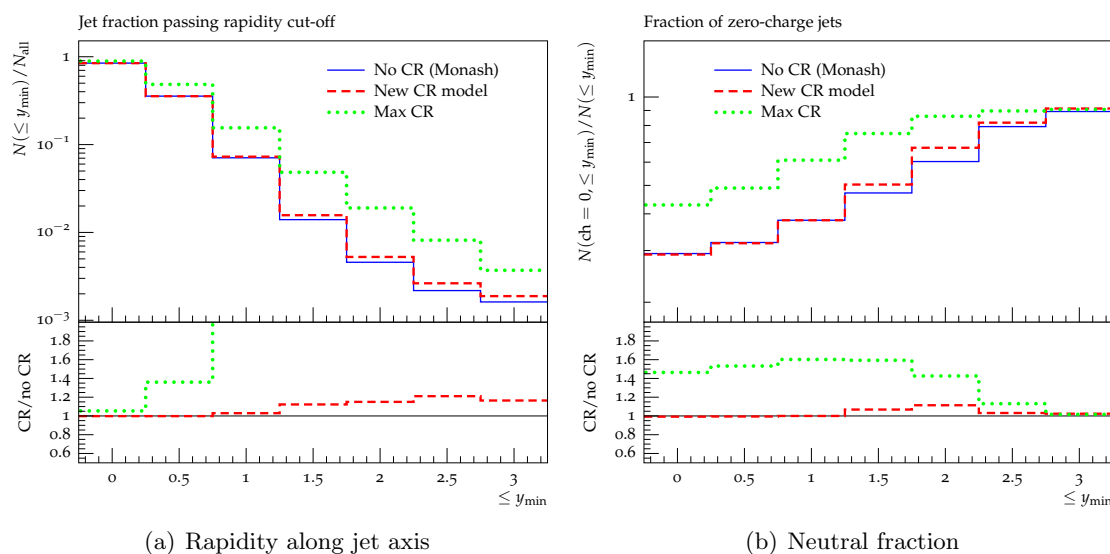


Figure 16. Observables constraining CR in 3-jet events at LEP. (a): the minimum rapidity of the constituents of the third jet with respect to its jet axis. (b): the fraction of the third jet with total charge equal to zero as a function of minimum rapidity of particles in the third jet.

therefore be strongly suppressed, below $1/N_C^2$. This aspect is not included in our model, since the history of the final state particles is not considered. The model may therefore overestimate the CR effect on three-jet events.

An additional consequence is that the jet will also have a total electric charge of zero, if all particles from the fragmentation fall within the jet. The best observable uses both of these ideas. Firstly a rapidity gap is required in order to select events with minimum radiation between the jets. This also enhances the probability that all the particles from the fragmentation falls within the jet. Secondly the jet is required to have a total charge of zero. This observable was first proposed during the LEP runs [117] and successively followed up by several of the experiments [35, 37, 40]. The studies were limited to excluding only the most extreme CR models, with no conclusions drawn between more moderate CR models and no CR; the data was located in between the two predictions.

Rather than comparing our model directly on the LEP data, we took a slightly simpler approach by only considering the difference between no CR and the new CR model in the relevant observables. The difference is found to be negligible on those observables, cf. the “No CR” and “New CR” histograms in figure 16. Thus we do not expect that the new CR model could be ruled out by these LEP constrains. We note that the small difference between with and without CR can be understood, by remembering the large focus on junction structures in the new CR model. Junction structures do not produce colour-singlet jets in the same manner as ordinary strings, and thus are not sensitive to this observable in the same way as ordinary strings reconnections. It is possible to consider a more extreme version of the new CR model where all dipoles are allowed to connect with each other, i.e. effectively replacing $1/N_C^2$ by unity! This is illustrated by the “max CR” histogram in figure 16. For this unphysically extreme case the difference between the two models

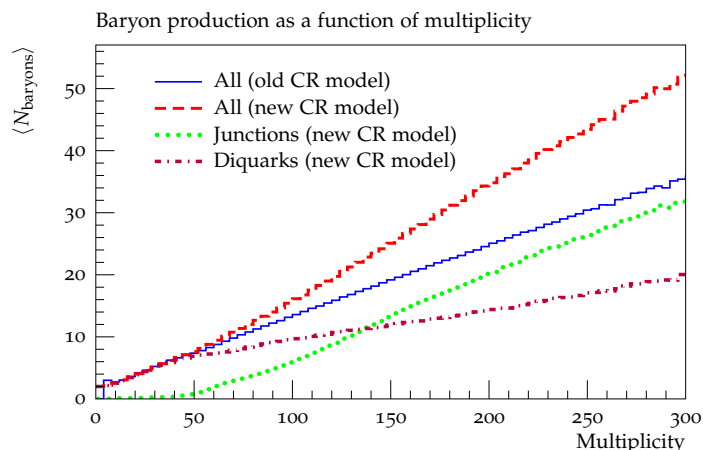


Figure 17. Average baryon multiplicity as a function of hadron multiplicity (generator-level, including all particles, hadronic decays switched off).

becomes so large it most likely would have been ruled out by the experiments. However such an extreme case would also be eliminated by just considering LHC measurements (e.g. $\langle p_{\perp} \rangle$ vs n_{ch}).

3.4 Suggestions for new observables

As discussed in section 3.2, the new CR model is able to reach agreement with some key observables that have otherwise proved difficult for the string model (as implemented in PYTHIA) in the pp environment. In particular, subleading dipole connections that minimise the string-length measure can account for the rise in the $\langle p_{\perp} \rangle (n_{\text{ch}})$ distribution (a feature also present in earlier CR models in PYTHIA, though without the connection with subleading colour), and subleading string-junction connections can account for the observed increase in e.g. the Λ/K ratio between ee and pp collisions.

The price is (a few) new free parameters governing the CR modelling, so the question naturally arises to what extent this type of model can be distinguished clearly from other other phenomenological modelling attempts to describe the same data. All models we are aware of that simultaneously aim to describe both the LEP data and the LHC data (or ee and pp data more generally) rely on the higher colour/energy densities present in pp collisions to provide the extra baryons. The multiplicity scaling of the baryon production is therefore expected to be higher than the linear scaling of the diquark model. This is also what we observe, cf. figure 17. For low multiplicity, both of the CR models agree with each other, however the increase happens faster for the new model. This shape difference in the scaling with particle multiplicity could provide an additional probe to test the new model.

One also notices that there is a significant difference between the baryon production of the old model and baryon production from diquarks in the new model. This is somewhat surprising since the hadronisation model is essentially left untouched. The explanation for this is two-fold: 1) the new CR model produces a different mass spectrum of strings (with generally lower invariant masses), and 2) low-mass strings and junction structures produce fewer additional diquarks.

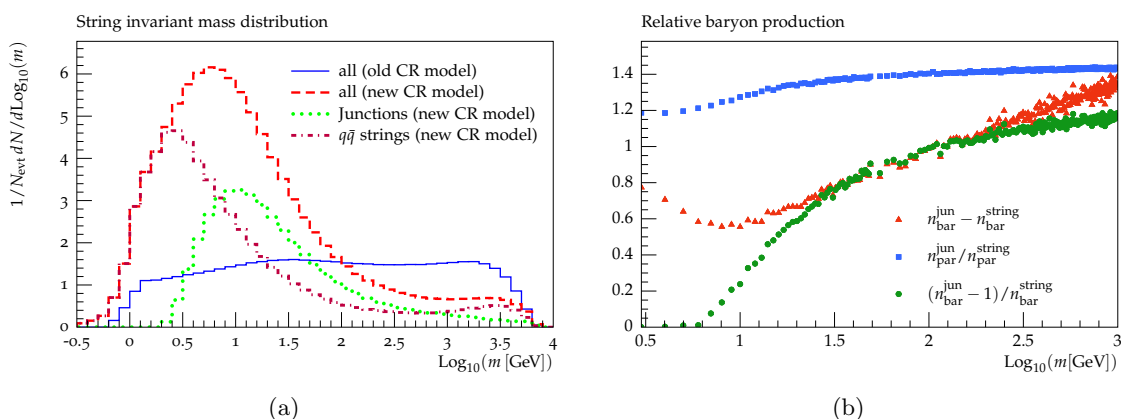


Figure 18. (a): invariant mass distribution of string systems (note logarithmic x axis). (b): the production of baryons from respectively a $q\bar{q}$ system (“string”) and a qqq system (“junction”) with the same total invariant mass. The red triangles show the difference of total primary-baryon multiplicities, the blue squares show the ratio of total primary-hadron multiplicities (mesons+baryons), and the green circles show the ratio for the primary-baryons multiplicities, subtracting off the extra baryon that the junction topology always produces.

The first point is illustrated in figure 18(a), which shows the invariant-mass distribution of strings in the new and old model. In the old model, the distribution is essentially flat, and includes a significant plateau towards very large invariant masses, whereas the distribution is strongly peaked at small invariant masses in the new model. The differences arise both from the junction cleanup procedure (by which longer strings can be split by insertion of an additional quark-antiquark pair), and from the minimisation of the λ -measure. The old model also minimised the λ -measure, however this is achieved by combining strings, giving fewer but higher-mass string systems than before the CR.

Due to energy-momentum conservation (and a greater relative importance of the quark endpoints), low-invariant-mass strings produce fewer baryons. Despite the fact that each junction system produces at least one baryon, we therefore note that this does not automatically lead to an increase in the total number of produced baryons. Since the invariant mass of a qqq junction system is distributed on three string pieces, whereas that of a $q\bar{q}$ system is carried by a single string, diquarks, which are relatively heavy, can actually be quite strongly phase-space suppressed in junction topologies, especially at low invariant masses of the string system (where the majority of the junction topologies lie cf. figure 18(a)). In addition the diquarks need to be pair produced to conserve baryon number, and the current implementation requires the pair to be on the same junction leg, leading to an even larger phase-space suppression. This effect is illustrated in figure 18(b), where for instance a junction system with $E_{\text{CM}} = 10$ GeV (in the peak of the mass distribution) has a five times lower probability to produce an (extra) diquark pair compared to a 10-GeV dipole string. At fixed multiplicity this effect is hidden in the tuned parameters, but can be observed by the different scaling.

Considering baryon production in more detail, the relative yield of different types of baryons is highly revealing. A collection of such yields for the different models are listed

in table 1. For most of the baryons, the new CR model predicts about 20 – 50% above the old model. This is in agreement with result for Λ production shown earlier (figure 12). There are however also some clear order-of-magnitude differences, for charm and bottom baryons. One example is Σ_c^0 production,⁷ for which the new CR model predicts a rate more than a factor of 20 above that of the old model!

To understand how such large differences can occur, we need to recall how baryons are produced in ordinary string fragmentation. Since no charm is produced in string break-ups, the only way to produce a Σ_c^0 is to produce a dd -diquark and combine it with c quark from the shower. However, since dd diquarks must have spin 1 (due to symmetry), their production is heavily suppressed relative to ud ones which can also exist in the far lighter spin-0 state. Enter junctions, for which there is a priori no specific penalty associated with having two legs end on same-flavour quarks as compared to different flavours. Up to some combinatorics and symmetry factors the production of ud and dd is therefore expected to be the same for junction systems, in good agreement with the observed predictions of the new CR model. As such the production of baryons like Σ_c^0 theoretically provides an excellent probe to study the relative importance between diquark- and junction-driven baryon production.

Similar results are also observed for Σ_c^{++} as well as for the production of the analogous b -baryons. And with the large b -physics programme at the LHC, we believe this could be an interesting study, both for its physics value as well as a possible source of background for other measurements. The effect can also be seen for Ω^- , however not as clearly as for charm and bottom baryons. Moreover the presence of additional suppression of strange diquarks in the string fragmentation model makes Ω^- connection more complicated. We note that it can be an important validation channel however.

Due to the majority of baryons being produced by the junction mechanism in the new CR model, the baryon yields also provides a clear probe to test the spin structure of the diquarks formed from the junctions. The large difference in yield between Σ_c^{*+} and Σ_c^0 is due to the choice of spin suppression mentioned earlier in the tuning section. An actual measurement could be directly applied as a constraint for this variable. At least with the development of the present model, we now have a vehicle that allows to explore this type of phenomenology and interpret the findings.

We should note that the baryons considered above are excited states that rapidly decay through the emission of a pion, e.g. $\Sigma_c^0 \rightarrow \Lambda_c^+ \pi^-$. As such they may be quite challenging to observe experimentally. It is therefore not a given that it will be easy to utilise measurements of these yields. But it does provide a theoretical motivation for studying the production and measurement of heavy baryons.

Another special class of baryons is the double (or triple) heavy baryons containing at least two c or b quarks. These baryons can not be formed in ordinary string fragmentation and is therefore almost non-existent in the old CR scheme. The only production mechanism is via the junctions from the beams (which also means that for p^+p^+ collisions no double-heavy antibaryons are predicted). This is also observed in table 1, where only a single

⁷The Σ_c^0 is a cdd state with spin $S = 1/2$, mass ~ 2.5 GeV and PDG code 4112 [26].

Particle	New CR model ($N_{\text{par}}/N_{\text{events}}$)			Old CR model $N_{\text{par}}/N_{\text{events}}$ (all)
	string	junction	all	
π^+	$2.5 \cdot 10^1$	0	$2.5 \cdot 10^1$	$2.4 \cdot 10^1$
p	2.5	1.4	3.8	3.2
n	2.4	1.3	3.7	3.2
Δ^{++}	$6.1 \cdot 10^{-1}$	$4.5 \cdot 10^{-1}$	1.1	$8.9 \cdot 10^{-1}$
Δ^+	$6.0 \cdot 10^{-1}$	$4.0 \cdot 10^{-1}$	1.0	$8.6 \cdot 10^{-1}$
Δ^0	$5.5 \cdot 10^{-1}$	$4.0 \cdot 10^{-1}$	$9.4 \cdot 10^{-1}$	$7.9 \cdot 10^{-1}$
Δ^-	$4.7 \cdot 10^{-1}$	$4.4 \cdot 10^{-1}$	$9.1 \cdot 10^{-1}$	$7.1 \cdot 10^{-1}$
K^+	5.2	0	5.2	5.1
Λ	$4.7 \cdot 10^{-1}$	$3.9 \cdot 10^{-1}$	$8.6 \cdot 10^{-1}$	$6.5 \cdot 10^{-1}$
Σ^+	$3.4 \cdot 10^{-1}$	$4.2 \cdot 10^{-1}$	$7.6 \cdot 10^{-1}$	$5.1 \cdot 10^{-1}$
Σ^0	$3.5 \cdot 10^{-1}$	$4.5 \cdot 10^{-1}$	$7.9 \cdot 10^{-1}$	$5.1 \cdot 10^{-1}$
Σ^-	$3.2 \cdot 10^{-1}$	$4.2 \cdot 10^{-1}$	$7.4 \cdot 10^{-1}$	$4.9 \cdot 10^{-1}$
Σ^{*+}	$9.6 \cdot 10^{-2}$	$8.9 \cdot 10^{-2}$	$1.9 \cdot 10^{-1}$	$1.5 \cdot 10^{-1}$
Σ^{*0}	$9.2 \cdot 10^{-2}$	$7.7 \cdot 10^{-2}$	$1.7 \cdot 10^{-1}$	$1.4 \cdot 10^{-1}$
Σ^{*-}	$8.3 \cdot 10^{-2}$	$8.7 \cdot 10^{-2}$	$1.7 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$
Ξ^-	$6.9 \cdot 10^{-2}$	$1.1 \cdot 10^{-1}$	$1.8 \cdot 10^{-1}$	$1.1 \cdot 10^{-1}$
Ω^-	$2.0 \cdot 10^{-3}$	$1.3 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$3.9 \cdot 10^{-3}$
D^+	$5.3 \cdot 10^{-2}$	0	$5.3 \cdot 10^{-2}$	$6.5 \cdot 10^{-2}$
Λ_c^+	$4.0 \cdot 10^{-3}$	$7.9 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	$6.6 \cdot 10^{-3}$
Σ_c^{++}	$2.7 \cdot 10^{-4}$	$1.3 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	$5.4 \cdot 10^{-4}$
Σ_c^+	$2.5 \cdot 10^{-4}$	$1.5 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$5.2 \cdot 10^{-4}$
Σ_c^0	$2.5 \cdot 10^{-4}$	$1.3 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	$5.1 \cdot 10^{-4}$
Σ_c^{*++}	$5.1 \cdot 10^{-4}$	$1.7 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$	$9.5 \cdot 10^{-4}$
Σ_c^{*+}	$4.9 \cdot 10^{-4}$	$1.9 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$9.4 \cdot 10^{-4}$
Σ_c^{*0}	$4.8 \cdot 10^{-4}$	$1.7 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$	$9.1 \cdot 10^{-4}$
$ccq^{(*)}$	0	$2.1 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$	$1.0 \cdot 10^{-7}$
B^+	$1.6 \cdot 10^{-3}$	0	$1.6 \cdot 10^{-3}$	$2.3 \cdot 10^{-3}$
Λ_b^0	$1.9 \cdot 10^{-4}$	$6.3 \cdot 10^{-4}$	$8.2 \cdot 10^{-4}$	$3.9 \cdot 10^{-4}$
Σ_b^+	$1.1 \cdot 10^{-5}$	$9.3 \cdot 10^{-4}$	$9.5 \cdot 10^{-4}$	$3.1 \cdot 10^{-5}$
Σ_b^0	$1.2 \cdot 10^{-5}$	$1.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$3.7 \cdot 10^{-5}$
Σ_b^-	$1.1 \cdot 10^{-5}$	$9.3 \cdot 10^{-4}$	$9.4 \cdot 10^{-4}$	$3.2 \cdot 10^{-5}$
Σ_b^{*+}	$1.1 \cdot 10^{-5}$	$9.3 \cdot 10^{-4}$	$9.5 \cdot 10^{-4}$	$3.1 \cdot 10^{-5}$
Σ_b^{*0}	$1.2 \cdot 10^{-5}$	$1.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$3.7 \cdot 10^{-5}$
Σ_b^{*-}	$1.1 \cdot 10^{-5}$	$9.3 \cdot 10^{-4}$	$9.4 \cdot 10^{-4}$	$3.2 \cdot 10^{-5}$
$bcq^{(*)}$	0	$1.8 \cdot 10^{-5}$	$1.8 \cdot 10^{-5}$	0
$bbq^{(*)}$	0	$1.1 \cdot 10^{-6}$	$1.1 \cdot 10^{-6}$	0

(*): Double heavy baryons where the last q can be any quark.

Table 1. Primary particle and antiparticle production of identified hadrons. Ten million ND events were simulated and all particles and antiparticles were counted. Hadron decays were switched off to only look at the primary production.

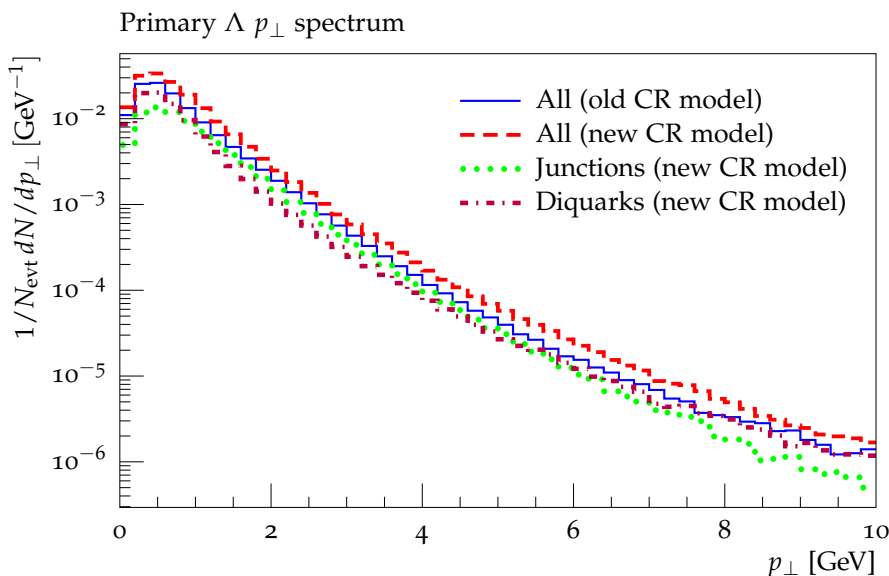


Figure 19. The Λ p_{\perp} -distribution separated by production mechanism. Only ND events were included and hadron decays were switched off.

double-heavy baryon is produced in the 10 million events. With the large amount of junctions, the new CR model provides a natural production mechanism for double-heavy baryons, and as such the expected amount is also significantly higher than for the old CR model. The effect of massive quarks in the λ -measure is not well understood, however, and possible other production mechanism might also contribute, thus the estimate is most likely rather crude. Irrespectively, a measurement of double-heavy baryons probes a region of hadronisation that the current models do not describe. And it could possible also shed some light on whether the junction mechanism might be a reasonable production mechanism.

So far, we only considered total particle yields; more knowledge is available by studying more differential distributions. A natural next extension is the transverse momentum distributions. As the junctions are formed by minimising the λ -measure, the particles defining the junction may be expected to preferentially be moving in the same direction and thereby create a boosted baryon. This in turn leads to an expected increase in transverse momentum for such junction baryons. This is also observed in the low- p_{\perp} region (below roughly $p_{\perp} \sim 4$ GeV), where the particle production peak is higher for junction baryons (figure 19). In the region of very high p_{\perp} (above roughly $p_{\perp} \sim 4$ GeV) the particle production is dominated by jets, for which the hard high- p_{\perp} partons are more important than the overall boost. In addition, the perturbative gluons associated with the jet already provides a low λ -measure and as such limited CR is expected inside the jet regions. This leads to the high- p_{\perp} region being occupied predominantly by baryons produced in ordinary (diquark) string-breaks.

Transverse momentum spectra have already been measured for some of the more common baryon species and a comparison with the Λ p_{\perp} spectrum measured by CMS is given in in figure 20. Sadly, the improvement is far less satisfactory here. The new CR model (as well as the old model) overshoots the production in the very low p_{\perp} and the high p_{\perp}

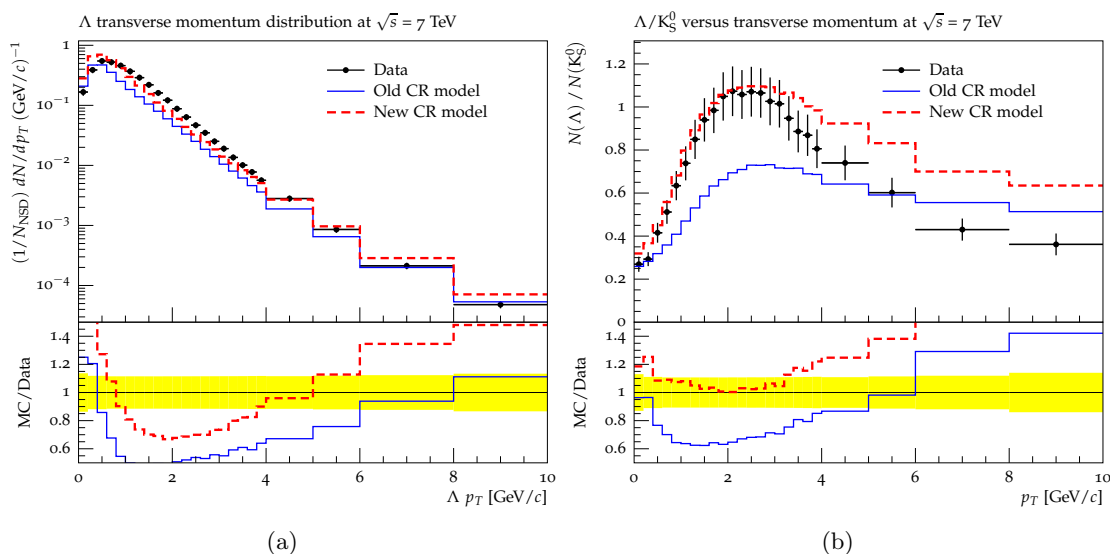


Figure 20. The (a) Λ p_{\perp} -distribution and (b) the Λ/K_s^0 p_{\perp} -distribution as measured by the CMS experiment [114]. All PYTHIA simulations were NSD with a lifetime cut-off ($\tau_{\text{max}} = 10$ mm/c) and a rapidity cut on 2 ($|y| < 2$). The yellow error band represents the experimental 1σ deviation.

region, whereas too few Λ baryons are predicted in between. Thus the Λ baryons from junctions tend to fall in the right region, however the effect is not large enough. An interesting observation is that the ratio Λ/K_s is now well described in the low p_{\perp} region. This shows that the problem with the p_{\perp} distribution is not specific to baryons but is more generic. The discrepancy between data and the model for large p_{\perp} still exists, however the baryon production in this region is primarily from diquark string breaks in jets and as such is not really unique for the new CR model. It may point to a revision needed of the spectrum of hard (leading?) baryon production in jets, which may not be unique to the pp environment, see [34].

The problem in the low p_{\perp} domain is a common theme for all heavier hadrons (i.e. anything but pions) and would be interesting to explore further. (E.g., a measurement of ρ spectra could reveal whether it depends on the presence of strange quarks.) The PYTHIA models predict a p_{\perp} -distribution that peaks at lower values than what is actually observed. To study this in more detail, one can calculate the average p_{\perp} for the identified hadrons and plot it a function of their mass, as done e.g. by the STAR collaboration for pp collisions at $E_{\text{CM}} = 200$ GeV [118]. In purely longitudinal string fragmentation the expected result is a roughly flat curve, since no correlation between the mass of the particle and p_{\perp} is present. The flat prediction is altered when hadron decays and jet physics are included, leading to the curve seen in figure 21. The prediction is also altered if the string is boosted (e.g., by partonic string endpoints), the boost is transferred to the final particles and for the same boost velocity a heavy particle will gain more p_{\perp} than a light one. This effect can be enhanced by CR, since minimisation of the λ -measure prefers reconnections among partons moving in the same direction, thus creating boosted strings [75]. CR is therefore expected to give a sharper rise of the $\langle p_{\perp} \rangle$ vs mass distribution. Unfortunately, we do not observe

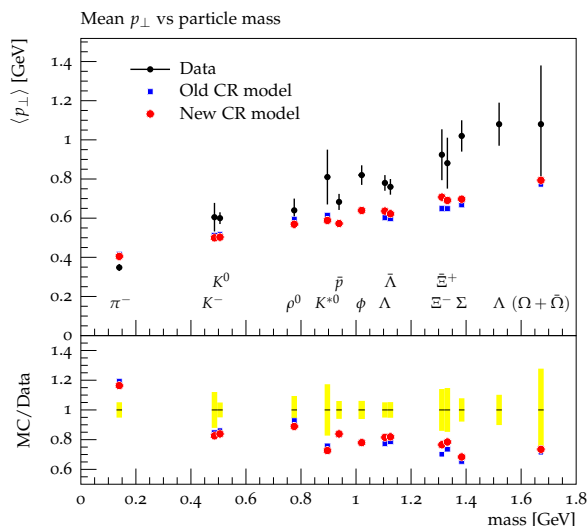


Figure 21. Average p_{\perp} versus hadron mass at $E_{\text{CM}} = 200$ GeV, compared with STAR data [118]. The yellow error band represents the experimental 1σ deviation.

this expected effect at any significant level (figure 21). To be candid, it is disappointing that the new model does not appear to address this problem at all. At the very least, it leaves room open for criticism and possibly additional new physics. Of special interest in this context are possible collective phenomena, such as (gas-like) hadron reinteractions or (hydro-like) flow, either of which could provide a (weak or strong, respectively) velocity-equalising component, and at least the latter has been applied successfully in the context of the EPOS model [72]. So-called “colour ropes” (strings carrying more than one unit of charge and hence having a higher tension) can also generate harder momentum spectra, while remaining within a string context, as was recently explored in [77]. These effects are generally not expected to be present at the relatively low energy density in pp collisions at a centre-of-mass energy of 200 GeV, however. It would therefore be of great interest to redo the measurement in detail at LHC energies, for high- and low-multiplicity samples and/or in the underlying event, to study whether the slope is steeper at the higher energy densities.

Finally, we emphasise that the rapidity gap survival is heavily affected by the choice of CR model [68, 84]. The explanation is similar to that of the three-jet LEP measurement, where the reconnected colour-singlet jet produces a rapidity gap to the other jets. The old CR model combines different strings into a single larger string, and thereby covers the same rapidity span, essentially adding kinks to an already existing string topology. Instead the new model can produce the colour-singlet union of particles in the same rapidity region. The new model is therefore expected to produce more rapidity gaps compared with the old model, which is also what we observe in figure 22. The new model predictions are significantly above the data in the mid-range rapidity region for low p_{\perp} cut-offs. (For higher p_{\perp} cuts the effect vanishes, due to the partonic description being more influential on the rapidity gap survival.) We therefore emphasise that the new CR scenario should in principle be accompanied by a retuning of diffractive parameters. This is not straightforward however, and involves not only the shown rapidity-gap survival distributions, but several

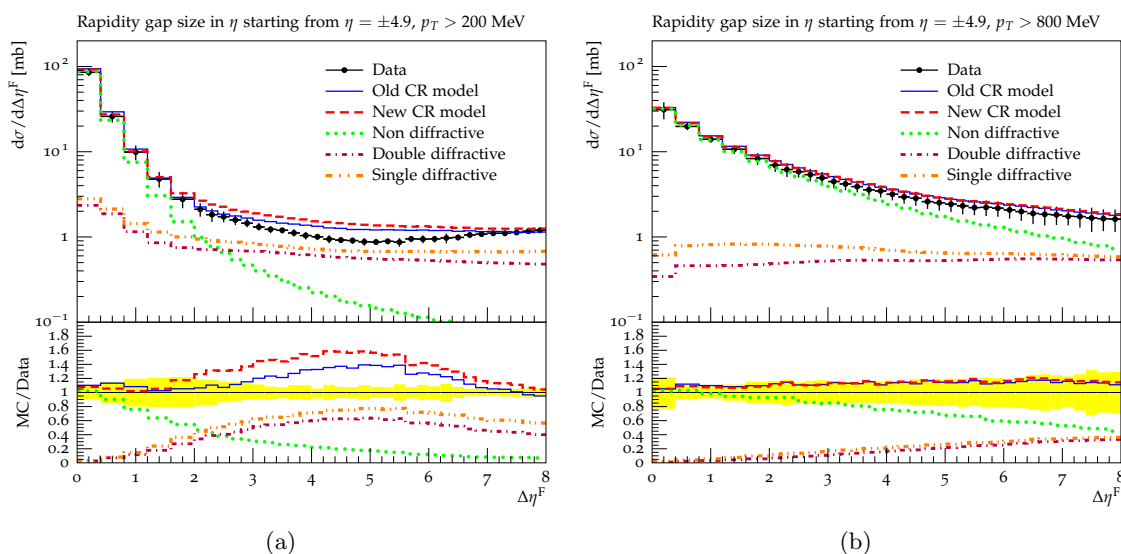


Figure 22. The rapidity gap survival for a low p_{\perp} cut (a) and a high p_{\perp} cut (b). The different components are also shown for the new CR model (DD/SD/ND). The gathered data is from the ATLAS experiment [119]. The yellow error band represents the experimental 1σ deviation.

other measurements at different energies. It was therefore deemed beyond the scope of this study to perform a retuning of the diffractive components and we limit ourselves to pointing out the interplay. A future dedicated study of this aspect could also well incorporate a study of the interference between the new BR model and diffractive events.

4 Application to top mass measurements at hadron colliders

Colour reconnections contribute one of the dominant uncertainties on current experimental top-mass extractions in hadronic channels (see e.g. the mini-review in [65]), and their size was recently reexamined in the context of several simplified CR schemes [71]. It is thus interesting to consider our new CR model in the same framework. We follow ref. [71], to which we refer for details on selection of the events and top mass reconstruction procedure.

Briefly stated, the idea is to select semi-leptonic top events, using the charged lepton and escaping neutrino mainly for tagging and then reconstruct the top mass from the hadronic decay. A mass window around the W mass is required and the raw top mass is extracted by fitting the invariant mass distribution of the three jets with a skewed Gaussian distribution, which fits the distributions better than a standard Gaussian [71]. For the models/tunes considered in this work (details below), the resulting shift on the calibrated top mass is below 200 MeV, which is comparable with the current level of CR uncertainty on the measurements [58, 59, 63, 64] and far below some of the “devil’s-advocate” toy models considered in [71]. Characteristic for those models is that they allow some fraction of reconnections where the λ measure is increased, whereas all the models considered here involve a minimisation of λ , one way or another.

Model	\hat{m}_{top} [GeV]	$\Delta\hat{m}_{\text{top}}$ [GeV]	$\Delta\hat{m}_{\text{top}}^{\text{rescaled}}$ [GeV]
no CR	169.57 ± 0.06	0	0
default ERD	169.26 ± 0.06	-0.36 ± 0.09	-0.04 ± 0.10
default	168.95 ± 0.06	-0.67 ± 0.09	$+0.17 \pm 0.10$
new model ERD	169.18 ± 0.06	-0.45 ± 0.09	$+0.03 \pm 0.10$
new model	168.97 ± 0.06	-0.66 ± 0.09	$+0.14 \pm 0.11$
max CR ERD	169.41 ± 0.06	-0.22 ± 0.09	-0.05 ± 0.10
no CR (tuned)	168.99 ± 0.06	-0.63 ± 0.09	-0.06 ± 0.10
max CR (tuned)	170.28 ± 0.07	$+0.66 \pm 0.09$	$+0.06 \pm 0.11$

Table 2. Values of m_{top} as predicted by the different CR models. The rescaled top mass is obtained by $\hat{m}_{\text{top}}^{\text{rescaled}} = \frac{80.385}{\hat{m}_W} \hat{m}_{\text{top}}$.

We compare six models: no CR, the existing (default) PYTHIA 8 CR model with and without early resonance decays (ERD), the new baseline CR model with and without ERD, and the new CR with maximal CR and ERD. Since the largest effect is expected for ERD, the maximal CR scenario is only considered with ERD switched on. Two versions of the “no CR” scenario were considered, one in which CR was simply switched off without any further changes (resulting in significant increase in central hadron multiplicity) and the other a semi-tuned version, where the activity ($\sum E_{\perp}$) in the central region ($|\eta| < 3$) for ND events was retuned. A similar approach is used for the maximal CR models, where again a non-tuned version and retuned version are considered. It should be noted that neither of the retuned models provide a good description of all MB data (for instance $\langle p_{\perp} \rangle (n_{\text{charged}})$ is described by neither tune). Since none of the models considered exhibit any top-specific behaviour, no additional retuning to top events was needed. The results are collected in table 2.

The first observation is the large difference between the no CR model and the non ERD models for $\Delta\hat{m}_{\text{top}}$ (~ 650 MeV). Since no CR is performed for the top decay products, not much of difference was expected, and any difference observed has to reside entirely in the underlying event (UE). The CR models considered lower the total string length, and thereby the activity in the UE, which directly influences the non-scaled top mass. Since the “no CR” model uses the same tune parameters as the CR models, it has a too high activity in the UE, leading to the negative mass shifts seen. After retuning to a similar activity in the central region, the “no CR (tuned)” model agrees with the default CR scenario within the statistical uncertainty. This emphasises the importance of using consistent UE tunes for this type of exercise, though we also note that after recalibration by the hadronic W mass, the rescaled top mass $\hat{m}_{\text{top}}^{\text{rescaled}}$ is remarkably stable.

For the ERD models the above shift in UE is still present, but the UE is now also allowed to reconnect with the top decay products. Therefore, an UE parton in close proximity to a jet from the top decay will have a large probability to be reconnected with the jet. This will result in narrower jets, leading to less of the energy falling outside the jet cone, and thereby a larger top mass. This is in agreement with the simulations, where the ERD results are above the non-ERD results. But the effect is smaller than that of the UE activity, thus the overall shift is still negative compared to the non-tuned no CR model.

Both of the above effects are magnified in the maximal CR scenario. We remind the reader that in this scenario, the $1/N_C^2$ suppression of subleading connections is switched off, hence this should be considered an unphysical extreme variation. Coincidentally, however, the two effects end up approximately canceling. A similar retuning of the central activity as above, “max CR (tuned)”, shows a significant increase in the top mass shift of more than one GeV with respect to the tuned no CR model, though again, the W mass calibration removes most of it.

The fact that the rescaled top mass ($\hat{m}_{\text{top}}^{\text{rescaled}}$) is less sensitive to CR is due to a cancellation between CR in the W mass and the top mass. This is in perfect agreement with the simulations, where the deviation for all rescaled masses are below their respective non-rescaled deviations. On the other hand, this means that any interpretation of where the variations between the models arise becomes extremely difficult. We will therefore refrain from attempting this, and instead purely discuss the numerical values in term of the uncertainty on the top mass.

For the rescaled top mass, the differences between the models stay below 200 MeV. This is slightly less than what was observed earlier even for identical CR models (default) [71]. The variations in the results can be attributed to a new tune combined with a change in the PS for $t\bar{t}$ events. We regard the smaller differences as somewhat coincidental however, and further work is needed to genuinely improve our understanding of CR effects in the top mass measurement. What we can say at least is that the results from the new model lie within those from the default CR model, and therefore do not generate a need for larger uncertainties. Even the maximal CR, which is our attempt to mimic the very large shifts seen for models that ignore MB/UE constraints, does not change this picture. Instead a pattern emerges, namely that whenever the minimisation of the λ measure is used as a guideline for the CR, the shifts stay below 300 MeV (taking the models studied in [71] into account as well). The reason for this is two-fold: firstly the coherence of the PS ensures that the jet structure is not too significantly altered; secondly, the alterations are realised in a systematic fashion leading to a similar shift in both the top and the W mass, implying that the hadronic W mass recalibration is highly robust. An increased understanding of this interplay could potentially lower the uncertainty even further.

Since these models are mainly constrained by measurements, further gains can also be achieved by improving and extending the programme of measurements sensitive to CR effects. A few new observables targeting top events specifically were already suggested in [71]. In the context of top mass uncertainties, such observables are of course especially relevant, as this is the closest to in-situ constraints as can be obtained, mimising the “extrapolation” that the model has to cover between the constraint and measurement environments.

In order to establish whether the small effects on m_t predicted by λ -minimising models are indeed conservative or not, it would be of crucial importance to test these models as directly as possible in a variety of environments, top included. The fact that our new CR models do not yet give good descriptions of identified-particle p_{\perp} spectra should, in this context, be seen as a warning that there can be additional non-perturbative uncertainties left unaccounted for, possibly of a dynamical origin.

5 Summary and outlook

The question “*between which partons do confining potentials arise?*” is a fundamental one in non-perturbative QCD, which any attempt at modelling the process of hadronisation must address. In the leading-colour approximation and neglecting beam-remnant correlations, this is relatively simple: there is a one-to-one mapping between perturbative QCD dipoles and string pieces / clusters. In this paper, we have attempted to take a first step beyond leading colour, by including a randomisation over the set of possible subleading-colour topologies, with probabilities chosen according to a simplified version of the $SU(3)$ colour algebra.

We present the argument that while the LC approximation may be quite good in the environment of e^+e^- collisions (more specifically in the absence of multi-parton interactions), we expect very significant deviations from it in pp collisions, where the survival of the strict LC topology should be heavily suppressed. Although the probability for a subleading colour connection to be possible between any given pair of (uncorrelated) partons is only roughly $1/N_C^2$, it becomes increasingly unlikely *not* to have any such connections as the number of uncorrelated partons increases, as e.g. in the case when considering MPI.

This implies that a complex multi-parton system will in general have several different string/cluster configurations open to it, at the time of hadronisation; the LC one is only one among many possibilities. We invoke the string-length λ measure to choose which one is preferred, so far via a simple winner-takes-all algorithm that does not purport to always find the global minimum. Nonetheless, we believe that this model represents a significant step in the right direction, allowing us to probe for the first time the effects of subleading colour on hadronisation in a way that may be said to be systematic and consistent with (a simplified version of) QCD.

One noteworthy new aspect of our work is the use of string junctions to represent antisymmetric colour combinations, such as two colours combining to form an overall anticolour. This provides a new source of baryon production, with properties qualitatively different from the standard diquark scenario. We have shown that this aspect allows to reconcile measured baryon/meson ratios with the string model in both pp and ee collisions simultaneously. However, we caution that the shapes of the p_\perp spectra are still not well described. We had anticipated that the preference of reconnections to produce boosted string pieces should lead to an enhancement of the $\langle p_\perp \rangle$ especially for heavier hadrons, but the magnitude of this effect observed in our model is still far too low to explain the data.

We emphasised that there is a conceptual difference between colour *connections* and colour *reconnections*. The former is related to colour-space *ambiguities*, such as the unknown colour correlations between different MPI initiators or the subleading-colour connections explored in this work. Colour *reconnections* are related to *dynamical* reconfigurations of the colour/string space, via perturbative gluon exchanges or non-perturbative string interactions; i.e., they involve *momentum* exchange as well. We did not explore effects of the latter type directly in this work, though we note that the fairly realistic string-interaction scenarios constructed by Khoze and Sjöstrand in the context of $e^+e^- \rightarrow W^+W^-$ studies [43, 66, 67] also feature string-length minimisation; hence it is possible that our tuned

parameters effectively attempt to cover both types. If so, the fact that the momentum spectra remain discrepant may point to the need for dynamical CR.

Finally, we presented a few suggestions for additional observables, the measurement of which would give further insight and possibly help to distinguish both physical and unphysical CR models, as well as other ideas such as models based on colour ropes [76, 77], hydrodynamics [72], or (non-hydro) hadron rescattering. We ended by considering the simplified top-mass analysis of [71] and conclude that the models presented here lead to shifts in the top mass of order 200 MeV, which is within the current level of non-perturbative uncertainties on the measurements.

Acknowledgements

Many thanks to T. Sjöstrand for his valuable comments on both our physics and our code. JRC thanks the CERN theory unit for hospitality during the main part of this study. Work supported in part by the Swedish Research Council, contract number 621-2013-4287, in part by the MCnetITN FP7 Marie Curie Initial Training Network, contract PITN-GA-2012-315877, and in part by the Australian Research Council, contract FT130100744.

A Model parameters

A complete list of all the parameters that differ from the Monash tune for the three different models are listed in the table below.

Parameter	Monash	Mode 0	Mode 2	Mode 3
StringPT:sigma	= 0.335	= 0.335	= 0.335	= 0.335
StringZ:aLund	= 0.68	= 0.36	= 0.36	= 0.36
StringZ:bLund	= 0.98	= 0.56	= 0.56	= 0.56
StringFlav:probQQtoQ	= 0.081	= 0.078	= 0.078	= 0.078
StringFlav:ProbStoUD	= 0.217	= 0.2	= 0.2	= 0.2
StringFlav:probQQ1toQQ0join	= 0.5, 0.7, 0.9, 1.0	= 0.0275, 0.0275, 0.0275, 0.0275	= 0.0275, 0.0275, 0.0275, 0.0275	= 0.0275, 0.0275, 0.0275, 0.0275
MultiPartonInteractions:pT0Ref	= 2.28	= 2.12	= 2.15	= 2.05
BeamRemnants:remnantMode	= 0	= 1	= 1	= 1
BeamRemnants:saturation	—	= 5	= 5	= 5
ColourReconnection:mode	= 0	= 1	= 1	= 1
ColourReconnection:allowDoubleJunRem	= on	= off	= off	= off
ColourReconnection:m0	—	= 2.9	= 0.3	= 0.3
ColourReconnection:allowJunctions	—	= on	= on	= on
ColourReconnection:junctionCorrection	—	= 1.43	= 1.20	= 1.15
ColourReconnection:timeDilationMode	—	= 0	= 2	= 3
ColourReconnection:timeDilationPar	—	—	= 0.18	= 0.073

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] G. Dissertori, I. Knowles and M. Schmelling, *Quantum chromodynamics — High energy experiments and theory*, Oxford Science Publications, Oxford U.K. (2003).
- [2] P. Skands, *Introduction to QCD*, lectures given at *TASI2012*, June 4–29, Boulder, U.S.A. (2012), [arXiv:1207.2389](https://arxiv.org/abs/1207.2389).
- [3] A. Buckley et al., *General-purpose event generators for LHC physics*, *Phys. Rept.* **504** (2011) 145 [[arXiv:1101.2599](https://arxiv.org/abs/1101.2599)] [[INSPIRE](#)].
- [4] R.D. Ball et al., *Parton distribution benchmarking with LHC data*, *JHEP* **04** (2013) 125 [[arXiv:1211.5142](https://arxiv.org/abs/1211.5142)] [[INSPIRE](#)].
- [5] E. Boos et al., *Generic user process interface for event generators*, [hep-ph/0109068](https://arxiv.org/abs/hep-ph/0109068).
- [6] J. Alwall et al., *A standard format for Les Houches event files*, *Comput. Phys. Commun.* **176** (2007) 300 [[hep-ph/0609017](https://arxiv.org/abs/hep-ph/0609017)] [[INSPIRE](#)].
- [7] A. van Hameren, C.G. Papadopoulos and R. Pittau, *Automated one-loop calculations: a proof of concept*, *JHEP* **09** (2009) 106 [[arXiv:0903.4665](https://arxiv.org/abs/0903.4665)] [[INSPIRE](#)].
- [8] F. Cascioli, P. Maierhofer and S. Pozzorini, *Scattering amplitudes with open loops*, *Phys. Rev. Lett.* **108** (2012) 111601 [[arXiv:1111.5206](https://arxiv.org/abs/1111.5206)] [[INSPIRE](#)].
- [9] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: going beyond*, *JHEP* **06** (2011) 128 [[arXiv:1106.0522](https://arxiv.org/abs/1106.0522)] [[INSPIRE](#)].
- [10] S. Alioli et al., *Update of the binoth Les Houches accord for a standard interface between Monte Carlo tools and one-loop programs*, *Comput. Phys. Commun.* **185** (2014) 560 [[arXiv:1308.3462](https://arxiv.org/abs/1308.3462)] [[INSPIRE](#)].
- [11] G. Cullen et al., *GoSam-2.0: a tool for automated one-loop calculations within the Standard Model and beyond*, *Eur. Phys. J. C* **74** (2014) 3001 [[arXiv:1404.7096](https://arxiv.org/abs/1404.7096)] [[INSPIRE](#)].
- [12] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [[arXiv:1405.0301](https://arxiv.org/abs/1405.0301)] [[INSPIRE](#)].
- [13] G. Gustafson and U. Pettersson, *Dipole formulation of QCD cascades*, *Nucl. Phys. B* **306** (1988) 746 [[INSPIRE](#)].
- [14] T. Sjöstrand and P.Z. Skands, *Transverse-momentum-ordered showers and interleaved multiple interactions*, *Eur. Phys. J. C* **39** (2005) 129 [[hep-ph/0408302](https://arxiv.org/abs/hep-ph/0408302)] [[INSPIRE](#)].
- [15] Z. Nagy and D.E. Soper, *Matching parton showers to NLO computations*, *JHEP* **10** (2005) 024 [[hep-ph/0503053](https://arxiv.org/abs/hep-ph/0503053)] [[INSPIRE](#)].
- [16] W.T. Giele, D.A. Kosower and P.Z. Skands, *A simple shower and matching algorithm*, *Phys. Rev. D* **78** (2008) 014026 [[arXiv:0707.3652](https://arxiv.org/abs/0707.3652)] [[INSPIRE](#)].
- [17] M. Dinsdale, M. Ternick and S. Weinzierl, *Parton showers from the dipole formalism*, *Phys. Rev. D* **76** (2007) 094003 [[arXiv:0709.1026](https://arxiv.org/abs/0709.1026)] [[INSPIRE](#)].

- [18] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **03** (2008) 038 [[arXiv:0709.1027](#)] [[INSPIRE](#)].
- [19] S. Platzer and S. Gieseke, *Coherent parton showers with local recoils*, *JHEP* **01** (2011) 024 [[arXiv:0909.5593](#)] [[INSPIRE](#)].
- [20] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: Multi-scale Improved NLO*, *JHEP* **10** (2012) 155 [[arXiv:1206.3572](#)] [[INSPIRE](#)].
- [21] L. Lönnblad and S. Prestel, *Merging multi-leg NLO matrix elements with parton showers*, *JHEP* **03** (2013) 166 [[arXiv:1211.7278](#)] [[INSPIRE](#)].
- [22] L. Hartgring, E. Laenen and P. Skands, *Antenna showers with one-loop matrix elements*, *JHEP* **10** (2013) 127 [[arXiv:1303.4974](#)] [[INSPIRE](#)].
- [23] B. Andersson, *The Lund model*, Cambridge University Press, Cambridge U.K. (1997).
- [24] B. Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand, *Parton fragmentation and string dynamics*, *Phys. Rept.* **97** (1983) 31 [[INSPIRE](#)].
- [25] G.S. Bali and K. Schilling, *Static quark-anti-quark potential: scaling behavior and finite size effects in SU(3) lattice gauge theory*, *Phys. Rev. D* **46** (1992) 2636 [[INSPIRE](#)].
- [26] PARTICLE DATA GROUP collaboration, J. Beringer et al., *Review of particle physics*, *Phys. Rev. D* **86** (2012) 010001 [[INSPIRE](#)].
- [27] G. Gustafson, *Dual description of a confined color field*, *Phys. Lett. B* **175** (1986) 453 [[INSPIRE](#)].
- [28] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026 [[hep-ph/0603175](#)] [[INSPIRE](#)].
- [29] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [[arXiv:0710.3820](#)] [[INSPIRE](#)].
- [30] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [[arXiv:1410.3012](#)] [[INSPIRE](#)].
- [31] A. Buckley, H. Hoeth, H. Lacker, H. Schulz and J.E. von Seggern, *Systematic event generator tuning for the LHC*, *Eur. Phys. J. C* **65** (2010) 331 [[arXiv:0907.2973](#)] [[INSPIRE](#)].
- [32] N. Firdous and G. Rudolph, *Tuning of PYTHIA6 to minimum bias data*, *EPJ Web Conf.* **60** (2013) 20056.
- [33] N. Fischer, S. Gieseke, S. Plätzer and P. Skands, *Revisiting radiation patterns in e^+e^- collisions*, *Eur. Phys. J. C* **74** (2014) 2831 [[arXiv:1402.3186](#)] [[INSPIRE](#)].
- [34] P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 tune*, *Eur. Phys. J. C* **74** (2014) 3024 [[arXiv:1404.5630](#)] [[INSPIRE](#)].
- [35] OPAL collaboration, G. Abbiendi et al., *Tests of models of color reconnection and a search for glueballs using gluon jets with a rapidity gap*, *Eur. Phys. J. C* **35** (2004) 293 [[hep-ex/0306021](#)] [[INSPIRE](#)].
- [36] L3 collaboration, P. Achard et al., *Search for color reconnection effects in $e^+e^- \rightarrow W^+W^- \rightarrow$ hadrons through particle flow studies at LEP*, *Phys. Lett. B* **561** (2003) 202 [[hep-ex/0303042](#)] [[INSPIRE](#)].
- [37] L3 collaboration, P. Achard et al., *Search for color singlet and color reconnection effects in hadronic Z decays at LEP*, *Phys. Lett. B* **581** (2004) 19 [[hep-ex/0312026](#)] [[INSPIRE](#)].

- [38] DELPHI collaboration, M. Siebel, *A Study of the charge of leading hadrons in gluon and quark fragmentation*, [hep-ex/0505080](#).
- [39] OPAL collaboration, G. Abbiendi et al., *Colour reconnection in $e^+e^- \rightarrow W^+W^-$ at $\sqrt{s} = 189\text{ GeV} - 209\text{ GeV}$* , *Eur. Phys. J. C* **45** (2006) 291 [[hep-ex/0508062](#)] [[INSPIRE](#)].
- [40] ALEPH collaboration, S. Schael et al., *Test of colour reconnection models using three-jet events in hadronic Z decays*, *Eur. Phys. J. C* **48** (2006) 685 [[hep-ex/0604042](#)] [[INSPIRE](#)].
- [41] DELPHI collaboration, J. Abdallah et al., *Investigation of colour reconnection in WW events with the DELPHI detector at LEP-2*, *Eur. Phys. J. C* **51** (2007) 249 [[arXiv:0704.0597](#)] [[INSPIRE](#)].
- [42] ALEPH collaboration, S. Schael et al., *Measurement of the W boson mass and width in e^+e^- collisions at LEP*, *Eur. Phys. J. C* **47** (2006) 309 [[hep-ex/0605011](#)] [[INSPIRE](#)].
- [43] T. Sjöstrand and V.A. Khoze, *On color rearrangement in hadronic $W^+ W^-$ events*, *Z. Phys. C* **62** (1994) 281 [[hep-ph/9310242](#)] [[INSPIRE](#)].
- [44] ALEPH, DELPHI, L3, OPAL, LEP ELECTROWEAK collaboration, S. Schael et al., *Electroweak measurements in electron-positron collisions at W-boson-pair energies at LEP*, *Phys. Rept.* **532** (2013) 119 [[arXiv:1302.3415](#)] [[INSPIRE](#)].
- [45] G. Marchesini and B.R. Webber, *Simulation of QCD jets including soft gluon interference*, *Nucl. Phys. B* **238** (1984) 1 [[INSPIRE](#)].
- [46] T. Sjöstrand and M. van Zijl, *A multiple interaction model for the event structure in hadron collisions*, *Phys. Rev. D* **36** (1987) 2019 [[INSPIRE](#)].
- [47] S. Gieseke, C. Rohr and A. Siodmok, *Colour reconnections in HERWIG++*, *Eur. Phys. J. C* **72** (2012) 2225 [[arXiv:1206.0041](#)] [[INSPIRE](#)].
- [48] 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 [[INSPIRE](#)].
- [49] UA1 collaboration, C. Albajar et al., *A study of the general characteristics of $p\bar{p}$ collisions at $\sqrt{s} = 0.2\text{ TeV}$ to 0.9 TeV* , *Nucl. Phys. B* **335** (1990) 261 [[INSPIRE](#)].
- [50] CDF collaboration, D. Acosta et al., *Soft and hard interactions in $p\bar{p}$ collisions at $\sqrt{s} = 1800\text{ GeV}$ and 630 GeV* , *Phys. Rev. D* **65** (2002) 072005 [[INSPIRE](#)].
- [51] CDF collaboration, T. Aaltonen et al., *Measurement of particle production and inclusive differential cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\text{ TeV}$* , *Phys. Rev. D* **79** (2009) 112005 [*Erratum ibid.* **D 82** (2010) 119903] [[arXiv:0904.1098](#)] [[INSPIRE](#)].
- [52] CMS collaboration, *Charged particle multiplicities in pp interactions at $\sqrt{s} = 0.9, 2.36$ and 7 TeV* , *JHEP* **01** (2011) 079 [[arXiv:1011.5531](#)] [[INSPIRE](#)].
- [53] ATLAS collaboration, *Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC*, *New J. Phys.* **13** (2011) 053033 [[arXiv:1012.5104](#)] [[INSPIRE](#)].
- [54] ALICE collaboration, *Multiplicity dependence of the average transverse momentum in pp, p-Pb and Pb-Pb collisions at the LHC*, *Phys. Lett. B* **727** (2013) 371 [[arXiv:1307.1094](#)] [[INSPIRE](#)].
- [55] A. Karneyeu, L. Mijovic, S. Prestel and P.Z. Skands, *MCPLLOTS: a particle physics resource based on volunteer computing*, *Eur. Phys. J. C* **74** (2014) 2714 [[arXiv:1306.3436](#)] [[INSPIRE](#)].

- [56] R. Corke and T. Sjöstrand, *Interleaved parton showers and tuning prospects*, *JHEP* **03** (2011) 032 [[arXiv:1011.1759](#)] [[INSPIRE](#)].
- [57] P.Z. Skands and D. Wicke, *Non-perturbative QCD effects and the top mass at the Tevatron*, *Eur. Phys. J. C* **52** (2007) 133 [[hep-ph/0703081](#)] [[INSPIRE](#)].
- [58] CDF collaboration, T. Aaltonen et al., *Precision top-quark mass measurements at CDF*, *Phys. Rev. Lett.* **109** (2012) 152003 [[arXiv:1207.6758](#)] [[INSPIRE](#)].
- [59] CMS collaboration, *Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **12** (2012) 105 [[arXiv:1209.2319](#)] [[INSPIRE](#)].
- [60] CDF, D0 collaboration, T.A. Aaltonen et al., *Combination of measurements of the top-quark pair production cross section from the Tevatron collider*, *Phys. Rev. D* **89** (2014) 072001 [[arXiv:1309.7570](#)] [[INSPIRE](#)].
- [61] CMS collaboration, *Measurement of the top-quark mass in all-jets $t\bar{t}$ events in pp collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J. C* **74** (2014) 2758 [[arXiv:1307.4617](#)] [[INSPIRE](#)].
- [62] ATLAS, CDF, CMS, D0 collaboration, *First combination of Tevatron and LHC measurements of the top-quark mass*, [arXiv:1403.4427](#) [[INSPIRE](#)].
- [63] D0 collaboration, V.M. Abazov et al., *Precision measurement of the top-quark mass in lepton+jets final states*, *Phys. Rev. Lett.* **113** (2014) 032002 [[arXiv:1405.1756](#)] [[INSPIRE](#)].
- [64] ATLAS collaboration, *Measurement of the top quark mass in the $t\bar{t} \rightarrow$ lepton + jets and $t\bar{t} \rightarrow$ dilepton channels using $\sqrt{s} = 7$ TeV ATLAS data*, *Eur. Phys. J. C* **75** (2015) 330 [[arXiv:1503.05427](#)] [[INSPIRE](#)].
- [65] A. Juste et al., *Determination of the top quark mass circa 2013: methods, subtleties, perspectives*, *Eur. Phys. J. C* **74** (2014) 3119 [[arXiv:1310.0799](#)] [[INSPIRE](#)].
- [66] V.A. Khoze and T. Sjöstrand, *Color correlations and multiplicities in top events*, *Phys. Lett. B* **328** (1994) 466 [[hep-ph/9403394](#)] [[INSPIRE](#)].
- [67] V.A. Khoze and T. Sjöstrand, *QCD interconnection studies at linear colliders*, *Eur. Phys. J. direct* **C 2** (2000) 1 [[hep-ph/9912297](#)] [[INSPIRE](#)].
- [68] J. Rathsmann, *A generalized area law for hadronic string re-interactions*, *Phys. Lett. B* **452** (1999) 364 [[hep-ph/9812423](#)] [[INSPIRE](#)].
- [69] M. Sandhoff and P.Z. Skands, *Colour annealing — A toy model of colour reconnections*, [hep-ph/0604120](#).
- [70] C. Buttar et al., *Les houches physics at TeV colliders 2005, standard model and Higgs working group: Summary report*, [hep-ph/0604120](#).
- [71] S. Argyropoulos and T. Sjöstrand, *Effects of color reconnection on $t\bar{t}$ final states at the LHC*, *JHEP* **11** (2014) 043 [[arXiv:1407.6653](#)] [[INSPIRE](#)].
- [72] T. Pierog, I. Karpenko, J.M. Katzy, E. Yatsenko and K. Werner, *EPOS LHC: test of collective hadronization with LHC data*, [arXiv:1306.0121](#) [[INSPIRE](#)].
- [73] CMS collaboration, *Observation of long-range near-side angular correlations in proton-proton collisions at the LHC*, *JHEP* **09** (2010) 091 [[arXiv:1009.4122](#)] [[INSPIRE](#)].
- [74] K. Werner, I. Karpenko and T. Pierog, *The ‘ridge’ in proton-proton scattering at 7 TeV*, *Phys. Rev. Lett.* **106** (2011) 122004 [[arXiv:1011.0375](#)] [[INSPIRE](#)].

- [75] A. Ortiz Velasquez et al., *Color reconnection and flowlike patterns in pp collisions*, *Phys. Rev. Lett.* **111** (2013) 042001 [[arXiv:1303.6326](#)] [[INSPIRE](#)].
- [76] B. Andersson and P.A. Henning, *On the dynamics of a color rope: the fragmentation of interacting strings and the longitudinal distributions*, *Nucl. Phys. B* **355** (1991) 82 [[INSPIRE](#)].
- [77] C. Bierlich, G. Gustafson, L. Lönnblad and A. Tarasov, *Effects of overlapping strings in pp collisions*, *JHEP* **03** (2015) 148 [[arXiv:1412.6259](#)] [[INSPIRE](#)].
- [78] R. Corke and T. Sjöstrand, *Multiparton interactions and rescattering*, *JHEP* **01** (2010) 035 [[arXiv:0911.1909](#)] [[INSPIRE](#)].
- [79] H. Fritzsche, *Producing heavy quark flavors in hadronic collisions: a test of quantum chromodynamics*, *Phys. Lett. B* **67** (1977) 217 [[INSPIRE](#)].
- [80] A. Ali, J.G. Korner, G. Kramer and J. Willrodt, *Nonleptonic weak decays of bottom mesons*, *Z. Phys. C* **1** (1979) 269 [[INSPIRE](#)].
- [81] H. Fritzsche, *How to discover the B mesons*, *Phys. Lett. B* **86** (1979) 343 [[INSPIRE](#)].
- [82] D. Eriksson, G. Ingelman and J. Rathsman, *Colour rearrangements in B-meson decays*, *Phys. Rev. D* **79** (2009) 014011 [[arXiv:0811.2651](#)] [[INSPIRE](#)].
- [83] W. Buchmüller and A. Hebecker, *A parton model for diffractive processes in deep inelastic scattering*, *Phys. Lett. B* **355** (1995) 573 [[hep-ph/9504374](#)] [[INSPIRE](#)].
- [84] A. Edin, G. Ingelman and J. Rathsman, *Soft color interactions as the origin of rapidity gaps in DIS*, *Phys. Lett. B* **366** (1996) 371 [[hep-ph/9508386](#)] [[INSPIRE](#)].
- [85] J.M. Butterworth, J.R. Forshaw and M.H. Seymour, *Multiparton interactions in photoproduction at HERA*, *Z. Phys. C* **72** (1996) 637 [[hep-ph/9601371](#)] [[INSPIRE](#)].
- [86] S. Gieseke, S. Plätzer, A. Siodmok, and C. Röhr, *Minimum bias and underlying event developments in Herwig++*, talk given at the *Physics at the LHC 2010*, [DESY-PROC-2010-01](#) (2010).
- [87] T. Sjöstrand and P.Z. Skands, *Multiple interactions and the structure of beam remnants*, *JHEP* **03** (2004) 053 [[hep-ph/0402078](#)] [[INSPIRE](#)].
- [88] CDF collaboration, F. Abe et al., *Transverse momentum distributions of charged particles produced in $\bar{p}p$ Interactions at $\sqrt{s} = 630$ GeV and 1800 GeV*, *Phys. Rev. Lett.* **61** (1988) 1819 [[INSPIRE](#)].
- [89] CDF collaboration, T. Affolder et al., *Charged jet evolution and the underlying event in $p\bar{p}$ collisions at 1.8 TeV*, *Phys. Rev. D* **65** (2002) 092002 [[INSPIRE](#)].
- [90] CDF collaboration, R.D. Field, *The underlying event in hard scattering processes*, *eConf C* **010630** (2001) P501 [[hep-ph/0201192](#)] [[INSPIRE](#)].
- [91] CDF collaboration, R. Field and R.C. Group, *PYTHIA tune A, HERWIG and JIMMY in Run 2 at CDF*, [hep-ph/0510198](#) [[INSPIRE](#)].
- [92] T. Sjöstrand, *Colour reconnection and its effects on precise measurements at the LHC*, [arXiv:1310.8073](#).
- [93] D. Wicke and P.Z. Skands, *Non-perturbative QCD effects and the top mass at the Tevatron*, *Nuovo Cim. B* **123** (2008) S1 [[arXiv:0807.3248](#)] [[INSPIRE](#)].
- [94] P.Z. Skands, *The Perugia tunes*, [arXiv:0905.3418](#).

- [95] P.Z. Skands, *Tuning Monte Carlo generators: the Perugia tunes*, *Phys. Rev. D* **82** (2010) 074018 [[arXiv:1005.3457](#)] [[INSPIRE](#)].
- [96] H. Schulz and P.Z. Skands, *Energy scaling of minimum-bias tunes*, *Eur. Phys. J. C* **71** (2011) 1644 [[arXiv:1103.3649](#)] [[INSPIRE](#)].
- [97] B.R. Webber, *A QCD model for jet fragmentation including soft gluon interference*, *Nucl. Phys. B* **238** (1984) 492 [[INSPIRE](#)].
- [98] E. Avsar, G. Gustafson and L. Lönnblad, *Small- x dipole evolution beyond the large- N_c limit*, *JHEP* **01** (2007) 012 [[hep-ph/0610157](#)] [[INSPIRE](#)].
- [99] E. Avsar, *On the dipole swing and the search for frame independence in the dipole model*, *JHEP* **11** (2007) 027 [[arXiv:0709.1371](#)] [[INSPIRE](#)].
- [100] T. Sjöstrand and P.Z. Skands, *Baryon number violation and string topologies*, *Nucl. Phys. B* **659** (2003) 243 [[hep-ph/0212264](#)] [[INSPIRE](#)].
- [101] J.R. Gaunt and W.J. Stirling, *Double parton distributions incorporating perturbative QCD evolution and momentum and quark number sum rules*, *JHEP* **03** (2010) 005 [[arXiv:0910.4347](#)] [[INSPIRE](#)].
- [102] C. Flensburg, G. Gustafson, L. Lönnblad and A. Ster, *Correlations in double parton distributions at small x* , *JHEP* **06** (2011) 066 [[arXiv:1103.4320](#)] [[INSPIRE](#)].
- [103] B. Blok, Yu. Dokshitzer, L. Frankfurt and M. Strikman, *p QCD physics of multiparton interactions*, *Eur. Phys. J. C* **72** (2012) 1963 [[arXiv:1106.5533](#)] [[INSPIRE](#)].
- [104] M. Diehl, D. Ostermeier and A. Schafer, *Elements of a theory for multiparton interactions in QCD*, *JHEP* **03** (2012) 089 [[arXiv:1111.0910](#)] [[INSPIRE](#)].
- [105] A.V. Manohar and W.J. Waalewijn, *A QCD analysis of double parton scattering: color correlations, interference effects and evolution*, *Phys. Rev. D* **85** (2012) 114009 [[arXiv:1202.3794](#)] [[INSPIRE](#)].
- [106] A.V. Manohar and W.J. Waalewijn, *What is double parton scattering?*, *Phys. Lett. B* **713** (2012) 196 [[arXiv:1202.5034](#)] [[INSPIRE](#)].
- [107] H.-M. Chang, A.V. Manohar and W.J. Waalewijn, *Double parton correlations in the bag model*, *Phys. Rev. D* **87** (2013) 034009 [[arXiv:1211.3132](#)] [[INSPIRE](#)].
- [108] B. Blok, Y. Dokshitzer, L. Frankfurt and M. Strikman, *Perturbative QCD correlations in multi-parton collisions*, *Eur. Phys. J. C* **74** (2014) 2926 [[arXiv:1306.3763](#)] [[INSPIRE](#)].
- [109] A.M. Snigirev, N.A. Snigireva and G.M. Zinovjev, *Perturbative and nonperturbative correlations in double parton distributions*, *Phys. Rev. D* **90** (2014) 014015 [[arXiv:1403.6947](#)] [[INSPIRE](#)].
- [110] S. Kirkpatrick, C. D. Gelatt, and M.P. Vecchi, *Optimization by simulated annealing*, *Science* **220** (1983) 671.
- [111] L3 collaboration, P. Achard et al., *Studies of hadronic event structure in e^+e^- annihilation from 30 GeV to 209 GeV with the L3 detector*, *Phys. Rept.* **399** (2004) 71 [[hep-ex/0406049](#)] [[INSPIRE](#)].
- [112] W.T. Giele, D.A. Kosower and P.Z. Skands, *Higher-order corrections to timelike jets*, *Phys. Rev. D* **84** (2011) 054003 [[arXiv:1102.2126](#)] [[INSPIRE](#)].

- [113] CMS collaboration, *Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. Lett.* **105** (2010) 022002 [[arXiv:1005.3299](#)] [[INSPIRE](#)].
- [114] CMS collaboration, *Strange particle production in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV*, *JHEP* **05** (2011) 064 [[arXiv:1102.4282](#)] [[INSPIRE](#)].
- [115] A. Buckley et al., *Rivet user manual*, *Comput. Phys. Commun.* **184** (2013) 2803 [[arXiv:1003.0694](#)] [[INSPIRE](#)].
- [116] TOTEM collaboration, G. Antchev et al., *Measurement of the forward charged particle pseudorapidity density in pp collisions at $\sqrt{s} = 7$ TeV with the TOTEM experiment*, *Europhys. Lett.* **98** (2012) 31002 [[arXiv:1205.4105](#)] [[INSPIRE](#)].
- [117] C. Friberg, G. Gustafson and J. Hakkinen, *Color connections in e^+e^- annihilation*, *Nucl. Phys.* **B 490** (1997) 289 [[hep-ph/9604347](#)] [[INSPIRE](#)].
- [118] STAR collaboration, B.I. Abelev et al., *Strange particle production in p+p collisions at $\sqrt{s} = 200$ GeV*, *Phys. Rev.* **C 75** (2007) 064901 [[nucl-ex/0607033](#)] [[INSPIRE](#)].
- [119] ATLAS collaboration, *Rapidity gap cross sections measured with the ATLAS detector in pp collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J.* **C 72** (2012) 1926 [[arXiv:1201.2808](#)] [[INSPIRE](#)].