

## Color matrix element corrections for parton showers

---

Simon Plätzer,<sup>a</sup> Malin Sjödal<sup>b</sup> and Johan Thorén<sup>b</sup>

<sup>a</sup>Particle Physics, Faculty of Physics, University of Vienna,  
Boltzmannngasse 5, 1030 Vienna, Austria

<sup>b</sup>Department of Astronomy and Theoretical Physics,  
Lund University, Sölvegatan 14A, 223 62 Lund, Sweden

E-mail: [simon.plaetzer@univie.ac.at](mailto:simon.plaetzer@univie.ac.at), [malin.sjodahl@thep.lu.se](mailto:malin.sjodahl@thep.lu.se),  
[johan.thoren@thep.lu.se](mailto:johan.thoren@thep.lu.se)

**ABSTRACT:** We investigate the effects of keeping the full color structure for parton emissions in parton showers for both LEP and LHC. This is done within the Herwig 7 dipole shower, and includes gluon emission, gluon splitting, initial state branching processes, as well as hadronization. The subleading  $N_c$  terms are included as color matrix element corrections to the splitting kernels by evolving an amplitude-level density operator and correcting the radiation pattern for each parton multiplicity, up to a fixed number of full color emissions, after which a standard leading color shower takes over. Our results are compared to data for a wide range of LEP and LHC observables and show that the subleading  $N_c$  corrections tend to be small for most observables probing hard, perturbative dynamics, for both LEP and LHC. However, for some of these observables they exceed 10%. On soft physics we find signs of significantly larger effects.

**KEYWORDS:** QCD Phenomenology

ARXIV EPRINT: [1808.00332](https://arxiv.org/abs/1808.00332)

---

**Contents**

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Essence of dipole factorization and dipole shower evolution</b>	<b>2</b>
<b>3</b>	<b>Color matrix element corrections</b>	<b>4</b>
<b>4</b>	<b>Color structure treatment</b>	<b>5</b>
<b>5</b>	<b>The weighted Sudakov algorithm</b>	<b>7</b>
<b>6</b>	<b>Evolution of the density operator</b>	<b>8</b>
<b>7</b>	<b>Results</b>	<b>10</b>
7.1	Outline of the simulation	10
7.2	LEP — final state radiation	13
7.2.1	Parton level analyses	13
7.2.2	Hadron level	14
7.3	LHC — coherent initial and final state radiation	14
7.3.1	Parton level analyses	14
7.3.2	Hadron level analyses	15
<b>8</b>	<b>Conclusion and outlook</b>	<b>18</b>
<b>A</b>	<b>Proof of the modified weighted veto algorithm</b>	<b>19</b>
<b>B</b>	<b>Choices for the modified veto algorithm</b>	<b>23</b>

---

**1 Introduction**

At present time event generators [1–4] have developed into indispensable tools for understanding collider phenomenology. At the same time, the high energy available at the LHC has significantly opened up the perturbative phase space available for radiation. This increases the demand of resummation, performed either analytically or, as in the case of parton showers, numerically e.g. via the Sudakov veto algorithm [5, 6].

From a QCD perspective, the high number of colored partons due to the large perturbative phase space, as well as due to the fact that the initial state partons carry color, calls for a better understanding of subleading  $N_c$  effects.

This is also in demand to ease the cancellation of infrared singularities in the matching and merging of parton showers with NLO calculations [7–9], which often fully include

subleading  $N_c$  effects. It should also be stressed that we expect such corrections to account for subleading  $N_c$ , leading-logarithmic problems which can arise in dipole-type parton shower algorithms due to an ambiguous definition of an emitter's color charge [10].

To achieve greater accuracy and better understanding of parton shower uncertainties, it is therefore time to include a better description of subleading color contributions in parton showers, similar to how parton showers recently have been improved by matrix element merging at leading order [11–16], and matching at next-to-leading order [7, 17–21].

The first steps in this direction has already been performed by some of the authors in the case of an  $e^+e^-$ -collider [22]. Others have pursued another road, keeping only a subset of the color suppressed terms [23, 24], and detailed studies have been carried out towards systematically expanding virtual and real effects in shower-type evolution algorithms [25, 26].

In the present paper we extend the color matrix element corrections, first implemented in [22], by including initial state hadrons as well as  $g \rightarrow q\bar{q}$ -splittings, subsequent leading color showering and hadronization. This is done within the Herwig 7.1 [27] implementation of the dipole shower algorithm [28], giving us a full-fledged general purpose event generator which can be used for studying color matrix element corrections to any process occurring at the LHC and other colliders, in practice up to a limited number of colored partons, restricted by the fast growing complexity in color space, however still reaching down to relatively soft emissions.

Our method is based on dipole factorization [29, 30], which we outline in section 2. The complication brought about by the color matrix element corrections is discussed in section 3, and implementation details, involving evolution of the color structure treatment, the weighted Sudakov algorithm and the density operator, are discussed in section 4, 5 and 6 respectively. Results for various processes, including initial and final state radiation, standard QCD observables, heavy quark production and  $Z$  plus jets are discussed in section 7, and concluding remarks are made in section 8.

## 2 Essence of dipole factorization and dipole shower evolution

This paper is based on dipole factorization, stating that whenever the next gluon to be emitted from an  $n$ -parton configuration becomes either soft or collinear to one of the existing partons, the squared amplitude for the  $n+1$ -parton case can be approximated with

$$|\mathcal{M}_{n+1}(\dots, p_i, \dots, p_j, \dots, p_k, \dots)|^2 \approx \sum_{k \neq i, j} \frac{1}{2p_i \cdot p_j} \langle \mathcal{M}_n(\dots, p_{\tilde{i}j}, \dots, p_{\tilde{k}}, \dots) | \mathbf{V}_{ij,k}(p_i, p_j, p_k) | \mathcal{M}_n(\dots, p_{\tilde{i}j}, \dots, p_{\tilde{k}}, \dots) \rangle, \quad (2.1)$$

in terms of the old amplitude  $|\mathcal{M}_n\rangle$ . In the above, an emitter  $\tilde{i}j \rightarrow i, j$  whereas a spectator  $k \rightarrow \tilde{k}$  absorbs the longitudinal recoil, such that all partons, before and after emission, stay

on-shell. For final state radiation we use the standard Sudakov decomposition,

$$p_i = zp_{\tilde{i}j} + \frac{p_{\perp}^2}{zs_{ijk}}p_{\tilde{k}} + k_{\perp} \quad (2.2)$$

$$p_j = (1-z)p_{\tilde{i}j} + \frac{p_{\perp}^2}{(1-z)s_{ijk}}p_{\tilde{k}} - k_{\perp} \quad (2.3)$$

$$p_k = \left(1 - \frac{p_{\perp}^2}{z(1-z)s_{ijk}}\right)p_{\tilde{k}}, \quad (2.4)$$

with  $p_{\tilde{i}j}^2 = p_{\tilde{k}}^2 = 0$ , a spacelike transverse momentum  $k_{\perp}$  with  $k_{\perp}^2 = -p_{\perp}^2$  and  $k_{\perp} \cdot p_{\tilde{i}j} = k_{\perp} \cdot p_{\tilde{k}} = 0$ . The cases of initial state emitter or spectator are discussed in [28].

Note that the sum in eq. (2.1) only runs over  $\tilde{i}j \neq \tilde{k}$ . This is possible since the collinear singularity corresponding to the square of the diagram where parton  $\tilde{i}j$  is the emitter, has been rewritten as a sum of interferences between that diagram and every other diagram using color conservation

$$\mathbf{T}_{\tilde{i}j}^2 = - \sum_{k \neq \tilde{i}j} \mathbf{T}_{\tilde{i}j} \cdot \mathbf{T}_k. \quad (2.5)$$

The splitting kernels are given in terms of the standard dipole splitting kernels as

$$\mathbf{V}_{ij,k}(p_i, p_j, p_k) = -8\pi\alpha_s V_{ij,k}(p_i, p_j, p_k) \frac{\mathbf{T}_{\tilde{i}j} \cdot \mathbf{T}_k}{\mathbf{T}_{\tilde{i}j}^2}, \quad (2.6)$$

where  $\mathbf{T}_{\tilde{i}j} \cdot \mathbf{T}_k$  describes the color space effect of exchanging a gluon between parton  $\tilde{i}j$  and parton  $\tilde{k}$ , and thus contains an implicit sum over gluon indices. In the massless case, the final-final dipole splitting kernels are given by

$$\begin{aligned} V_{qg,k}(p_i, p_j, p_k) &= C_F \left( \frac{2(1-z)}{(1-z)^2 + p_{\perp}^2/s_{ijk}} - (1+z) \right) \\ V_{gg,k}(p_i, p_j, p_k) &= 2C_A \left( \frac{1-z}{(1-z)^2 + p_{\perp}^2/s_{ijk}} + \frac{z}{z^2 + p_{\perp}^2/s_{ijk}} - 2 + z(1-z) \right). \end{aligned} \quad (2.7)$$

Note, however, that dipole factorization is valid also for massive particles in the quasi-collinear limit, and our implementation is general, using the massive dipole splitting kernels and kinematics available in Herwig, as detailed in [31]. For the dipole configurations involving initial state partons, we use the corresponding expressions as given in [30, 32]. In eq. (2.7) the factors of  $C_F = T_R(N_c^2 - 1)/N_c$  and  $C_A = 2T_R N_c$ , where  $T_R$  is defined by  $\text{tr}(t^a t^b) = T_R \delta^{ab}$ , explains the inclusion of  $\mathbf{T}_{\tilde{i}j}^2$  in eq. (2.6); in order to use the standard definition of eq. (2.7),  $\mathbf{T}_{\tilde{i}j}^2$  is introduced in the denominator of eq. (2.6).

In the large  $N_c$  limit the color correlator becomes

$$- \frac{\mathbf{T}_{\tilde{i}j} \cdot \mathbf{T}_k}{\mathbf{T}_{\tilde{i}j}^2} \rightarrow \frac{1}{1 + \delta_{\tilde{i}j}} \delta(\tilde{i}j, k \text{ color connected}), \quad (2.8)$$

where  $\delta_{\tilde{i}j} \equiv 1$  if  $\tilde{i}j$  is a gluon, and zero otherwise. In this way, the factor  $C_F$  for gluon emission off a quark is reproduced in the large  $N_c$  limit by coherent emission from the quark and its color-connected partner, and the factor  $C_A$  is reproduced for gluon splitting

by the sum of the coherent emissions from the gluon and its two color-connected partners, hence the factor 1/2 for gluons in eq. (2.8).

We remark that the leading  $N_c$  version of the above describes how to get a leading  $N_c$  correct emission pattern. It does not address the issue of assigning a leading  $N_c$  color flow from which subsequent emission can be performed. The default strategy is to use a color flow where the radiated gluon is inserted between the emitter and the spectator. However, with a gluon splitting kernel which is symmetric in  $z \leftrightarrow 1 - z$ , the question of which gluon is the radiator and which is radiated arises. In the limit where  $z$  is small, it is rather the parton with momentum fraction  $z$  which is soft and thus should be seen as radiated and therefore be inserted between the emitter (with the large momentum fraction  $1 - z$ ) and the spectator. For this reason, to mimic the swap of color, we swap momenta of the emitter and the emitted parton with probability  $1 - z$ , such that the soft gluon always tends to be inserted between the harder parton and the spectator in the color structure, guaranteeing that we get the correct soft limit. We remark, however that the probability  $1 - z$  is a choice, and that any function, having the same limits when  $z \rightarrow 1$  and  $z \rightarrow 0$ , would have been a valid choice. Alternatively, the splitting kernels could have been redefined to only contain the  $1 - z$  singularity (see [33, 34] for comparison).

As we want to describe LHC collisions, the color matrix element corrections have to be applied also to initial state hadrons, meaning that we have to deal with initial-initial emissions as well as initial-final and final-initial emissions. The initial state emission cases are treated in a standard backward evolution scheme, meaning that if the emitter is an initial state parton, the backward evolution is done by folding in the parton distribution functions (PDFs), using appropriate splitting kernels, and colors are updated as if the resulting (low energy) parton was emitted. To be more precise, denoting the emitter participating in the hard process by  $i$ , the radiated parton with  $j$ , and the (initial) parton going into the PDF by  $\tilde{i}j$ , the used splitting kernel is  $P_{\tilde{i}j \rightarrow ij}$ , and the emission probability is evolved using PDF ratios [35], as described in section 6.5 in [2]. The color structure of the full color shower, on the other hand, is, as discussed in section 6, treated as if  $i \rightarrow \tilde{i}j, j$ .

In the shower algorithm presented here, we also extend our analysis from [22] by including  $g \rightarrow q\bar{q}$  splitting. The description of how this splitting fits into the dipole formalism is given in section 6.

### 3 Color matrix element corrections

We like to stress that this paper deals with color matrix element corrections to parton showers. We thus correct each emission with the full color correlations, keeping all soft and collinear contributions to the emission, i.e., we use the right “antenna pattern”. In this sense, we do more than the standard leading  $N_c$  showers — where only the leading  $N_c$ -terms, and the color suppressed term in  $C_F = T_R(N_c^2 - 1)/N_c$ , are kept — but less than a full matrix element correction.

We thus start from some color amplitude,  $|\mathcal{M}_n\rangle$ , for  $n$  colored partons, decomposed in any arbitrary color basis, with basis vectors  $|\alpha_n\rangle$

$$|\mathcal{M}_n\rangle = \sum_{\alpha=1}^{d_n} c_{n,\alpha} |\alpha_n\rangle \quad \leftrightarrow \quad \mathcal{M}_n = (c_{n,1}, \dots, c_{n,d_n})^T . \tag{3.1}$$

Squaring the amplitude in order to calculate a cross section gives

$$|\mathcal{M}_n|^2 = \mathcal{M}_n^\dagger S_n \mathcal{M}_n = \text{Tr} \left( S_n \times \mathcal{M}_n \mathcal{M}_n^\dagger \right) \quad (3.2)$$

with  $S_n$  being the scalar product matrix,  $(S_n)_{\alpha\beta} = \langle \alpha_n | \beta_n \rangle$ , for color basis vectors  $|\alpha_n\rangle$  and  $|\beta_n\rangle$ . Upon emission from the  $\tilde{i}j, \tilde{k}$ -pair, the relevant color factor changes to

$$\langle \mathcal{M}_n | \mathbf{T}_{\tilde{i}j} \cdot \mathbf{T}_{\tilde{k}} | \mathcal{M}_n \rangle = \text{Tr} \left( S_{n+1} \times T_{\tilde{k},n} \mathcal{M}_n \mathcal{M}_n^\dagger T_{\tilde{i}j,n}^\dagger \right) \quad (3.3)$$

in terms of matrix representations,  $T_{\tilde{i}j}, T_{\tilde{k}} \in \mathbb{C}^{d_{n+1}, d_n}$ , of  $\mathbf{T}_{\tilde{i}j}, \mathbf{T}_{\tilde{k}}$ .

While this improves the radiation pattern to fixed order, we should remark that it is not the full story of color suppressed terms in the context of parton showers. To fully include all subleading  $N_c$  terms in the soft and collinear limits, virtual color rearranging terms associated with the same singularity structure should also be kept. To accomplish this, a full resummation of virtual exchanges is needed. Unfortunately, within the current event generator structure these contributions cannot be included, and we postpone their inclusion for future work. Instead we present a fully functional subleading  $N_c$  dipole shower, building on the algorithm presented in [22], but including  $g \rightarrow q\bar{q}$  splitting, hadronization, full mass dependence and initial state hadrons, meaning final-final, initial-final, final-initial and initial-initial dipoles.

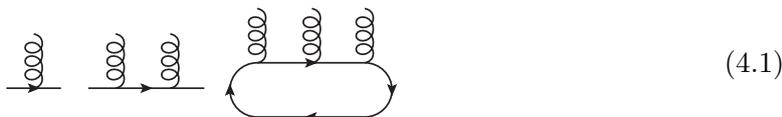
The rewriting of the collinear singularities in terms of dipole splitting kernels eqs. (2.6)–(2.7), using color conservation, eq. (2.5), means that the color flow will be replaced by every other color flow, except the flow associated with the color structure of the collinear singularity. In order to be able to continue the full  $N_c$  matrix element corrected parton shower with a leading  $N_c$  parton shower (as well as with subsequent hadronization), we do, however, keep the color structure associated with emission from the emitter, as in the standard event record, and we use that for the subsequent leading  $N_c$  shower and hadronization. Not having one color structure to start from when it comes to hadronization would imply considering a new approach to a hadronization model, which is much beyond the scope of the current paper. On the other hand, it is still interesting to add hadronization in a standard way to enable comparison to data.

## 4 Color structure treatment

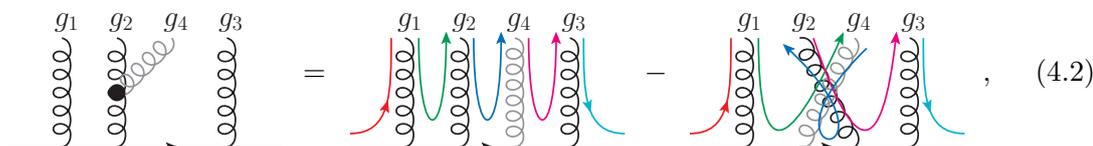
So far, the treatment of color structure in this paper has been completely basis independent. Any complete spanning set of relevant color structures, such as trace bases [22, 36–44], multiplet bases [45–51], or color flow bases [52–54] would do, as long as the matrices  $T_{\tilde{i}j,n}$  for gluon emission, the matrices  $t_{\tilde{i}j,n}$  describing gluon splitting and the scalar product matrices  $S_n$  could be calculated. This basis freedom is maintained within the Matchbox module of Herwig, which can be used to interface to any implementation of color structure, such as CVolver [25] or ColorFull [44].

For our simulations we use trace bases and the ColorFull implementation [44]. In the trace bases, the color structure is expressed in terms of open and closed quark-lines, of

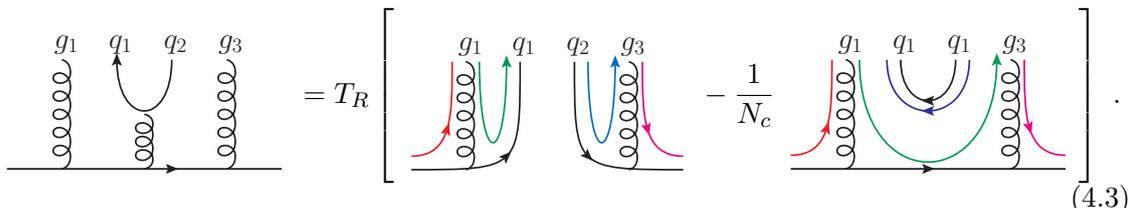
the form



for all possible quark and gluon permutations and all possible number of traces [42, 44]. These bases have several advantages. The color structure can trivially be translated into the leading  $N_c$  color flow, which is needed for subsequent leading  $N_c$  emission and standard hadronization. The processes of gluon emission, gluon splitting, and gluon exchange can all be very easily described, giving respectively at most two, two and four new basis vectors [39, 41, 42]. For example, considering gluon emission off an open quark-line we have the two trace basis vectors



where we have also illustrated the different color flows in the  $N_c \rightarrow \infty$  limit.<sup>1</sup> If we instead consider gluon splitting, we have



One of the disadvantages of trace basis lies in their overcompleteness, meaning that hence — strictly speaking — they are actually not bases, but rather spanning sets.

For a small set of partons, up to approximately five  $q\bar{q}$ -pairs and gluons, the overcompleteness is fairly moderate, but for more partons it becomes significant [49, 51]. The number of basis vectors scales as a factorial in the trace basis case, with roughly  $(N_g + N_{q\bar{q}})!/e$  basis vectors [49], whereas the number of basis vectors for finite  $N_c$  scale only as an exponential. On the other hand, basis vectors which are not needed for a given process can often easily be identified and crossed out for trace bases, in particular, at tree-level, the number of required basis vectors scales approximately as  $(N_g + N_{q\bar{q}} - 1)!$ . For example, in eq. (4.1), only color structures where all gluons are attached to the open quark-lines contribute.

The overcompleteness, and the fact that the scalar product matrices are dense, i.e., the scalar product between most pairs of basis vectors does not vanish, is the Achilles' heel of the trace bases. Instead of vanishing for two different basis vectors, the scalar product is suppressed by one or more powers of  $1/N_c$ . Thus, when  $N_c \rightarrow \infty$ , the bases are orthogonal,

<sup>1</sup>The relative sign on the right hand side is a matter of convention, and must be matched with the sign of the kinematics structure. Here we apply ColorFull's convention of introducing a minus sign when the emitted gluon is inserted before the emitter on the quark-line.

and the color structure can be replaced by color flows, as in eq. (4.2) and eq. (4.3). Since we keep the full color structure, it is the calculation of scalar products that limits the number of subleading  $N_c$  emissions that we can keep. At about  $N_g + N_{q\bar{q}} = 6$ , the color structure calculations start to take up significant time, and going beyond  $N_g + N_{q\bar{q}} = 7$  is very time consuming within our current setup.

This unfavorable scaling behavior could be circumvented by using orthogonal multiplet bases [49, 51], in which case the calculation of the radiation matrices  $T_{\tilde{i}j,n}$  would be more time consuming, however not to the same degree [50]. Alternatively color structure could be sampled over. This is a road which we have attempted to pursue. Within our current framework, using the weighted Sudakov veto algorithm from [55] (modified as described in section 5), it is, however, impractical to go beyond 3 subleading  $N_c$  emissions at LEP, or two subleading  $N_c$  emissions at LHC, due to the very poor statistical convergence from large weight fluctuations. Therefore, within our current implementation, sampling has proven disadvantageous.

## 5 The weighted Sudakov algorithm

The shower described in this paper treats up to  $N_{\max}$  emissions with the full color correlations. This corrects the emissions to appear first in the  $p_{\perp}$ -ordered evolution, down to smaller scales, and then the leading color shower handles the subsequent lower  $p_{\perp}$  emissions. The radiation pattern we aim at describing in the full color shower is

$$\begin{aligned}
 dP_{ij,k}(p_{\perp}^2, z; p_{\tilde{i}j}, p_{\tilde{k}}) &= \frac{\alpha_s}{2\pi} \frac{dp_{\perp}^2}{p_{\perp}^2} dz \mathcal{J}(p_{\perp}^2, z; p_{\tilde{i}j}, p_{\tilde{k}}) V_{ij,k}(p_{\perp}^2, z; p_{\tilde{i}j}, p_{\tilde{k}}) \\
 &\times \frac{-1}{\mathbf{T}_{\tilde{i}j}^2} \frac{\langle \mathcal{M}_n | \mathbf{T}_{\tilde{i}j} \cdot \mathbf{T}_{\tilde{k}} | \mathcal{M}_n \rangle}{|\mathcal{M}_n|^2}, \tag{5.1}
 \end{aligned}$$

where the factor after the multiplication sign is the color matrix element correction, which we will denote by  $\omega_{\tilde{i}j\tilde{k}}^n$ . For the cases when the color matrix element correction is negative, the weighted Sudakov veto algorithm from [55] is used, with one modification: in the competition version of the algorithm from [55], the weight for an emission gets a contribution from each veto and accept step, for all trial emissions of all competing pairs. The algorithm will, however, produce the same distribution if the total weight only receives contributions from the accept step of the winning emission and veto steps at scales larger than the winning scale. Discarding all weight contributions below some scale (in this case the scale of the winning emission) makes the algorithm generate another distribution below that scale, but the radiation pattern of the losing trial emissions below the winning scale cannot affect the final distribution, since these emissions are discarded anyway. We therefore choose to discard the weights below the winning scale. This modification significantly improves the convergence. The proof that this modified algorithm generates the same distribution as the algorithm of [55] can be found in appendix A.

Two choices in the algorithm from [55] are: the acceptance probability (denoted  $\epsilon$  in [55]) and the overestimate proposal distribution (denoted  $R$ ). These free choices were

used to improve the convergence of the algorithm, the details of the choices can be found in appendix B.

In total, to reproduce the radiation pattern in eq. (5.1), the shower steps given below are repeated until  $N_{\text{max}}$  emissions have been corrected or no emission is found above the cut-off scale  $\mu$ .

1. The starting scale  $Q_{\perp}$  is given by the hard scale. For the first emission, this is taken to be the  $Z$  mass for LEP and the average transverse momentum of hard jets in the final state for LHC. For subsequent emissions,  $Q_{\perp}$  is given by the scale of the previous emission.
2. All processes for all pairs of partons compete with each other and a winning hardest scale is chosen. For each dipole,  $i\tilde{j}, \tilde{k}$ , candidate emissions,  $i\tilde{j}, \tilde{k} \rightarrow i, j, k$  at scales  $p_{\perp, i\tilde{j}, k}$ , are chosen according to the Sudakov form factor

$$-\ln \Delta_{ij,k}(p_{\perp, ij,k}^2 | Q_{\perp}^2) = \frac{\alpha_s}{2\pi} \int_{p_{\perp, ij,k}^2}^{Q_{\perp}^2} \frac{dq_{\perp}^2}{q_{\perp}^2} \int_{z_-(q_{\perp}^2)}^{z_+(q_{\perp}^2)} dz \mathcal{P}_{ij,k}(q_{\perp}^2, z; p_{i\tilde{j}}, p_{\tilde{k}}), \quad (5.2)$$

where  $\mathcal{P}_{ij,k}$ , in accordance with eq. (5.1), is

$$\mathcal{P}_{ij,k}(p_{\perp}^2, z; p_{i\tilde{j}}, p_{\tilde{k}}) = \mathcal{J}(p_{\perp}^2, z; p_{i\tilde{j}}, p_{\tilde{k}}) V_{ij,k}(p_{\perp}^2, z; p_{i\tilde{j}}, p_{\tilde{k}}) \times \frac{-1}{\mathbf{T}_{i\tilde{j}}^2} \frac{\langle \mathcal{M}_n | \mathbf{T}_{i\tilde{j}} \cdot \mathbf{T}_k | \mathcal{M}_n \rangle}{|\mathcal{M}_n|^2} \quad (5.3)$$

and  $z_{\pm}(p_{\perp}^2)$  follow from the phase space boundaries at fixed transverse momentum. If the color matrix element correction is positive, the standard Sudakov veto algorithm is used (resulting in that the trial emission always contributes a factor 1 to the event weight) and if it is negative, the modified weighted veto algorithm is used (where the weight is, in general, multiplied by the weight in eq. (B.4)). The winning emission defines the details of the kinematics and the recoil is absorbed by the spectator  $\tilde{k}$  of the winning dipole, such that all partons are on-shell after the emission.

3. If no scale above the cut-off  $\mu$  was found, the shower terminates.
4. If this is the emission  $N_{\text{max}}$ , the leading  $N_c$  shower will continue showering the event, otherwise the density operator is updated as will be described in section 6.

If  $N_{\text{max}}$  emissions have been corrected, the leading  $N_c$  shower continues with the color structure given by the large  $N_c$  flow associated with emissions from the selected emitters, as discussed in section 3. The leading  $N_c$  shower then continues until reaching the cut-off scale  $\mu$ . Finally, the event may, or may not, be hadronized. If the hadronization is performed it starts from the leading  $N_c$  color flow.

## 6 Evolution of the density operator

The parton shower starts from the hard matrix element  $|\mathcal{M}_n\rangle$ . However, after emission, the resulting ‘‘dipole’’ color structure from eq. (3.3) cannot, within our framework, be cast

into the form of some new amplitude  $|\mathcal{M}_{n+1}\rangle$ . Instead we see from eq. (3.3) that the relevant  $n + 1$ -parton quantity, corresponding to  $M_n \equiv \mathcal{M}_n \mathcal{M}_n^\dagger$  for emission from  $\tilde{i}, \tilde{k}$  is  $\sim T_{\tilde{k},n} \mathcal{M}_n \mathcal{M}_n^\dagger T_{\tilde{i},n}^\dagger$ . For gluon emission (final or initial), keeping all contributions to the emission probability, and using the dipole factorization eq. (2.1) with the splitting kernels eq. (2.6), we see that if we define

$$M_{n+1} = - \sum_{i \neq j} \sum_{k \neq i,j} \frac{4\pi\alpha_s V_{ij,k}(p_i, p_j, p_k)}{p_i \cdot p_j \mathbf{T}_{ij}^2} T_{\tilde{k},n} M_n T_{\tilde{i},n}^\dagger, \quad (6.1)$$

the matrix element square for  $n + 1$  particles can be written analogously to eq. (3.2) as

$$|\mathcal{M}_{n+1}|^2 = \text{Tr}(S_{n+1} \times M_{n+1}). \quad (6.2)$$

Thus, when a phase space point has been selected for gluon emission,  $M_n$  could be updated according to eq. (6.1). Note, however, that the overall normalization of  $M_{n+1}$  is irrelevant, since when used in the  $n + 2$ -version of eq. (5.3) to calculate the emission of  $n + 2$  partons,  $M_{n+1}$  enters in both the numerator and denominator. Thus we could ignore any constant factor. In fact, for technical reasons, we only keep the eikonal parts,  $\sim p_i \cdot p_k / (p_i \cdot p_j p_k \cdot p_j)$ , of eq. (6.1). Clearly the dropped hard collinear pieces should not alter the subsequent emission of soft wide-angle radiation.

For the case of  $g \rightarrow q\bar{q}$ , there is no interference between various possible emitters, and the amplitude is symmetric in all final state gluons, meaning that  $M_n$  can be updated using only one term

$$M_{n+1} = t_{\tilde{i},n} M_n t_{\tilde{j},n}^\dagger, \quad (6.3)$$

where  $t_{\tilde{i},n}$  represents the color space map corresponding to  $t_{q\bar{q}}^g$ , i.e., the matrix where element  $\alpha\beta$  is the transition from basis vector  $\beta$  in the initial (smaller) basis, to basis vector  $\alpha$  in the final (larger) basis, where the difference between the color structures is that the gluon  $\tilde{i}$  has been contracted and replaced by the  $q\bar{q}$ -pair  $i, j$ , giving one or two new basis vectors. The possible recoil partners used to set up the gluon splitting into quarks are picked using eq. (5.1) with the  $g \rightarrow q\bar{q}$  splitting kernel, but with color matrix element corrections *as for gluon emission*. While this choice is ad-hoc in this case, it has the advantage of nicely fitting into the dipole picture. In the collinear limit, where the splitting becomes relevant, we can use color conservation eq. (2.5), to observe that the sum over the kernels using different spectators is indeed collapsing to the expected collinear splitting function.

Note that the same update of  $M_{n+1}$ , eq. (6.3), and the same recoil strategy, is applied irrespectively of if the splitting gluon is final or initial. The only difference is thus, as for the gluon emission case, the standard convolution with the PDFs for initial parton splitting rates. If we have an initial state quark (antiquark) which evolves backwards into a gluon going into the parton distribution function and an antiquark (quark) which is radiated, the shower has no interference with other diagrams, and the density matrix can be updated according to

$$M_{n+1} = T_{\tilde{i},n} M_n T_{\tilde{j},n}^\dagger, \quad (6.4)$$

without an irrelevant, overall factor. In the standard dipole large  $N_c$  shower, the momentum recoil is absorbed by the color-connected partner, implying that the emission can be accounted for within the dipole shower formalism.

A comment on the update of the density matrix is in order here. In this paper we keep the full color structure, of all possible emitter-spectator pairs which could have contributed to an emission. An alternative strategy is to pairwise sample over emitters and spectators, corresponding to keeping only one term in the double sum in eq. (6.1). While this samples from all color structures, we like to remark that it actually corresponds to a slightly different approximation. In our current implementation, after defining the kinematics of the new emission from the winning pair, *all* pairs contribute to the color structure of the next emission, and the weight of their color structure, is given by eq. (6.1), implicitly multiplying the same Sudakov factor, the Sudakov factor of the winning pair. In a sampling procedure, the terms in the eq. (6.1) would end up in different events, and each term would be associated with the Sudakov factor *of the pair emitting* in that winning phase space point. The difference in a sufficiently large Monte Carlo sample, thus lies within the Sudakov factor, coming with different scales for different pairs. We note however, that the two approaches should agree in the soft limit, and we have also checked that the numerical difference is small.

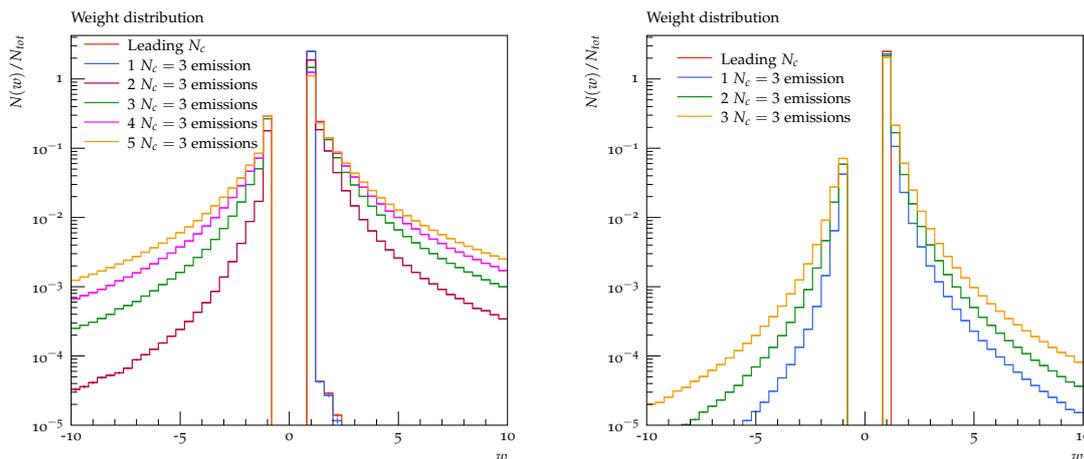
## 7 Results

### 7.1 Outline of the simulation

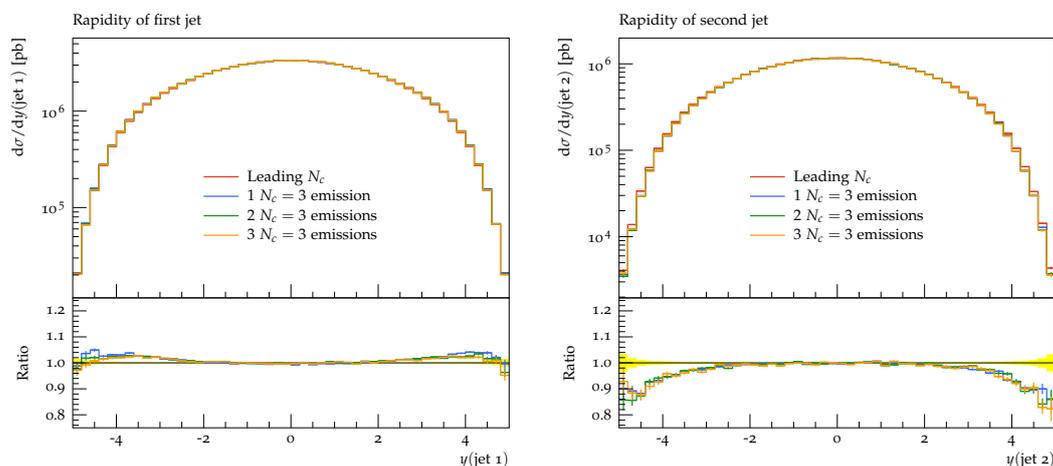
In this section we outline the simulation, and consider various shower distributions with the aim of understanding and validating the shower evolution.

We use the Herwig 7.1 dipole shower, with settings according to the 7.1.3 release [27], with the modified weighted Sudakov veto algorithm outlined in section 5, and with color matrix element corrections as described in section 3, starting from lowest order  $2 \rightarrow 2$  processes. The LHC generation  $p_\perp$ -cut is by default put to 30 GeV, and the default jet analysis veto is  $p_\perp = 50$  GeV. The energy is 13 TeV for LHC and 91.2 GeV for LEP unless stated otherwise. If jet clustering is required, our default choice is the anti- $k_T$  algorithm, as provided by the fastjet package [56], with  $R = 0.4$  at LHC. At LEP, no generation cuts are applied, and both at LEP and LHC we use the original Rivet [57] analysis published along with the data in comparison to data.

To ensure statistical convergence, we start with investigating the weight distribution. In figure 1 we show the weight distribution for up to five/three subleading  $N_c$  emissions at LEP/LHC respectively corresponding to up to six/seven  $q\bar{q}$ -pairs plus gluons in the color bases. We note that although the weight distributions get broader with the number of emissions, it stays sufficiently narrow to ensure convergence for the considered number of subleading emissions. The number of colored partons, which we can practically include, is therefore limited by the evaluation of scalar products in color space, as described in section 4. (We will find, however, that we are able to keep sufficiently many full color emissions for standard hard observables to converge.)



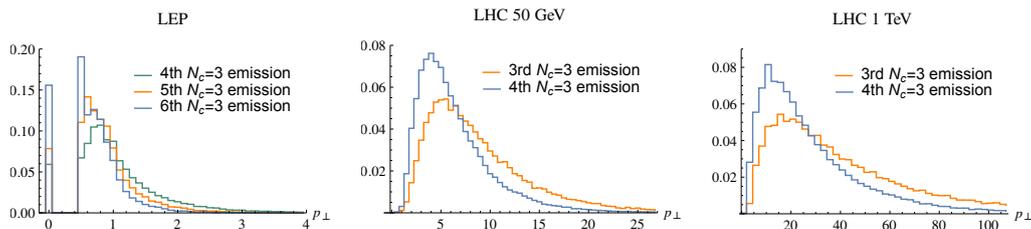
**Figure 1.** Weight distribution for  $e^+e^-$  (left) and  $pp$  collisions (right) depending on the number of  $N_c = 3$  emissions allowed. All generated events are used in these plots, i.e., no further analysis cut is applied.



**Figure 2.** Rapidity distribution of the first and second jet as zero, one, two and three  $N_c = 3$  emissions are kept.

Somewhat against intuition, we see a broader weight distribution for LEP events than for LHC events, despite the fact that we tend to have more colored partons at the LHC. This can be attributed to the fact that the corrections often tend to be negative at LEP (starting from  $e^+e^- \rightarrow q\bar{q}$ ), due to the negative contribution from coherent emission from the  $q\bar{q}$ -pair. In line with this, we also note that if we separately study  $qq \rightarrow qq$ ,  $qg \rightarrow qg$  and  $gg \rightarrow gg$ , we find the largest weight variations for  $qq \rightarrow qq$ , another case where we can expect large negative corrections from  $q\bar{q}$ -pairs.

We next turn to the convergence of standard observables with respect to the number of subleading  $N_c$  corrected emissions. As an example, we consider the rapidities of the hardest LHC jets in figure 2. As can be seen, the curves converge as more subleading emissions are added, and in this respect we see the same pattern for all standard hard LHC observables; they all converge when up to three subleading emissions are added, i.e., starting with a



**Figure 3.** Transverse momentum distribution in the dipole frame of the last  $N_c = 3$  emission (orange) and the first non-corrected emission (blue). For LEP the second-to-last  $N_c = 3$  emission is also shown (green). For LHC we use a 50 GeV (middle) and a 1 TeV (right) cut on the hardest parton. The shower cut-off at LEP is 0.45 GeV, and events which reach this scale are shown in the 0-bin above.

$2 \rightarrow 2$  topology and adding three subleading emissions (followed by leading  $N_c$  showering) gives results very similar to adding just two subleading  $N_c$  emissions (followed by leading  $N_c$  showering). LEP observables show a similar convergence pattern. Only when explicitly considering very many jets, does the convergence fail. This strongly suggests that for standard hard observables, subleading  $N_c$  corrections can be well approximated by color correcting the first few emissions.

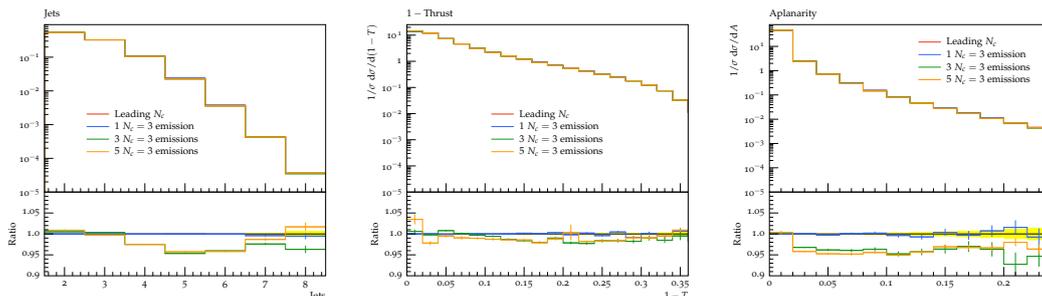
The convergence can also be underpinned by studying the evolution scale at which the  $N_c = 3$  parton shower terminates, and further evolution only is given by the leading  $N_c$  shower. This is investigated in figure 3 where the transverse momentum, as measured in the frame of the emitting dipole, is shown in orange for the last full color corrected emission, i.e., while keeping up to three subleading emissions at the LHC and up to five subleading emissions at LEP. For comparison we also show the distribution of the first emission not to be corrected.

We find that the typical scale of the last color corrected emission is about 1 GeV at LEP and about 5–10 GeV at LHC, using a 50 GeV cut. Increasing the cut to 1 TeV gives a much harder last subleading  $N_c$  corrected emission, as expected.

The convergence of observables is also in line with results from a very recent paper on subleading  $N_c$  corrections at LEP [58], where the authors claim to observe good convergence while keeping subleading corrections down to a variable cut-off at around 3 GeV.

In [58] a Monte Carlo sampling over color is advocated, and the point is made that this avoids the factorial scaling in color space. In view of figure 3, we note, however, that we typically go further down in  $p_\perp$  despite keeping the full color structure, although clearly there is no scale down to which we guarantee that we always keep the  $N_c = 3$  color structure. In fact, even if we limit ourselves to four subleading  $N_c$  emissions, for which the time penalty due to color structure treatment is negligible, we also tend to go well below 3 GeV. Therefore, in the case of hard observables at LEP, color structure sampling seems to be of no benefit for the convergence of observables.

We have also performed a number of standard shower variation checks, including shower scale variation and infrared cutoff variation. While varying the shower scale, we find effects very similar to the leading  $N_c$  case. Increasing the infrared cutoff from 1 to



**Figure 4.** Parton level plots for fraction of LEP-events containing  $n$  jets with  $E > 5$  GeV (left), thrust (middle) and aplanarity (right), the jets have been clustered with the anti- $k_{\perp}$ -type generalized  $e^+e^-$  clustering algorithm with  $R = 0.7$ . For hadronized events the effect on thrust vanishes, whereas the effects on number of jets and aplanarity are somewhat reduced. Note that the case of one full  $N_c$  emission should agree with the leading  $N_c$  shower, as is seen.

2 GeV likewise results in differences well comparable to the leading  $N_c$  shower, and so does adding multiple interactions.

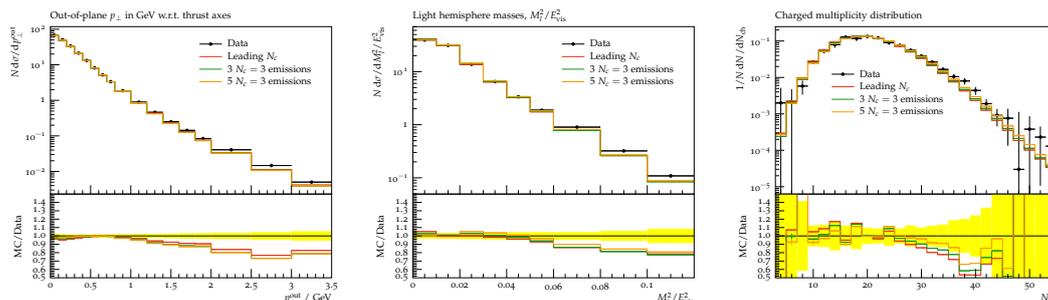
Finally, we have checked what happens if the shower is turned off completely after one to three subleading emissions. Here, we find large differences for the LHC, even if we keep up to three subleading  $N_c$  corrected emissions. We therefore conclude that it is essential to keep showering beyond three color corrected emissions, but it is — for standard hard QCD observables, or likely any observable which is mostly sensitive to the hardest jets — not important to keep color correcting the subsequent, softer and softer, emissions. For observables depending on soft physics, the situation may be different, as indicated below.

## 7.2 LEP — final state radiation

### 7.2.1 Parton level analyses

We first recapitulate the exercise from [22] and run an  $e^+e^-$  simulation at  $\sqrt{s} = 91.2$  GeV, without hadronization, but this time including subsequent leading  $N_c$  showering beyond the (up to) five subleading  $N_c$  emissions, as well as  $g \rightarrow q\bar{q}$  splittings.<sup>2</sup> Again we find that the corrections to most LEP observables are small. As examples of observables which show some effect, we show the fraction of events containing  $n$  jets with  $E > 5$  GeV, the thrust distribution, and the aplanarity in figure 4, in all cases showing corrections below 10%. On the other hand, effects can be significant in tailored situations. For example, considering the average transverse momentum with respect to a thrust axis defined by the three hardest partons, we find corrections above 10%. We caution, however, that without modification this is not an observable.

<sup>2</sup>The shower has also changed in several other respects compared to [22], the momentum fraction integration boundaries have changed, the leading  $N_c$  assignment has been updated as described in section 2 (having a large effect on predictions, similar to in [34]), the running of  $\alpha_s$  is different, and we use the Sudakov veto algorithm from section 5. For all these reasons, a direct comparison is not possible.



**Figure 5.** Out-of-plane  $p_{\perp}$  w.r.t. the thrust and thrust major axes (left), light hemisphere mass (middle) and fraction of events containing  $N_{\text{ch}}$  charged particles, using to data and Rivet analyses from [59, 60].

## 7.2.2 Hadron level

We have studied hadronized LEP events for a large class of observables from [59, 60]. For planarity, sphericity, oblateness, and in- and out-of-plane  $p_{\perp}$  w.r.t. sphericity axis, we find small differences of a few percent or less, whereas C-parameter shows some effects of 5–10% in the low C-parameter region. As examples, in figure 5, we show the total out-of-plane  $p_{\perp}$  (w.r.t. the plane defined by the thrust and thrust major axes) and the light hemisphere mass in comparison to data from [59].

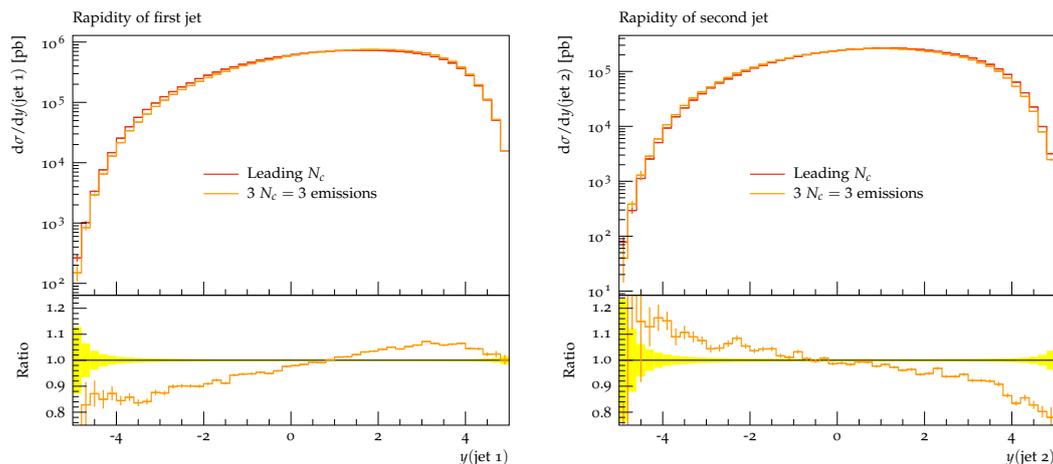
In general the deviation of simulated results compared to data from [59–61] is clearly dominated by other factors, and the overall description of data does not change visibly. Turning, on the other hand, to observables sensitive to soft physics, we find larger differences. As an example we show the charged multiplicity distribution in figure 5 (right), compared to data from [60], but we likewise see large effects for Durham jet resolution variables in the soft region, on the color singlet cluster masses for the Herwig hadronization model [2] and on individual hadron multiplicities.

While it is tempting to interpret the charged particle multiplicity plot in figure 5, as improved data description, we remark that the hadronization model in use, is the standard Herwig cluster hadronization model [2], with a leading  $N_c$  color flow, as described in section 3. We therefore caution that the differences should only be seen as an indication of effects on soft physics, from the altered particle kinematics entering the hadronization. While this can be interpreted as a need to retune the full color parton shower, retuning shower parameters is beyond the scope of the present paper.

## 7.3 LHC — coherent initial and final state radiation

### 7.3.1 Parton level analyses

We now turn to the LHC and start with reconsidering figure 2. Considering the leading two jets, with our standard  $p_{\perp} > 50 \text{ GeV}$  analysis cut, we see that they tend to have slightly different rapidity distributions, with the second jet being more central. For most other standard observables we find small differences, of a few percent or less. Nevertheless it is illustrative to separately consider scattering involving different partons. Doing so we find that for  $qq \rightarrow qq$  and  $gg \rightarrow gg$ , the subleading  $N_c$  corrections are small for all studied



**Figure 6.** Rapidity distribution of the hardest and second hardest jet while considering only  $qg \rightarrow qg$  scattering.

observables, whereas for  $qg \rightarrow qg$ , they can be more sizable. In figure 6 we therefore revisit the rapidity distributions of the two leading jets, and find corrections going in opposite directions for the two jets. Since  $qg$ -induced scattering contributes with a large fraction of the cross section for the applied cuts, and since LHC data contains  $qg \rightarrow qg$  and  $gg \rightarrow gg$  (along with all other processes), it can be concluded that significant cancellation of subleading  $N_c$  corrections is present at LHC.

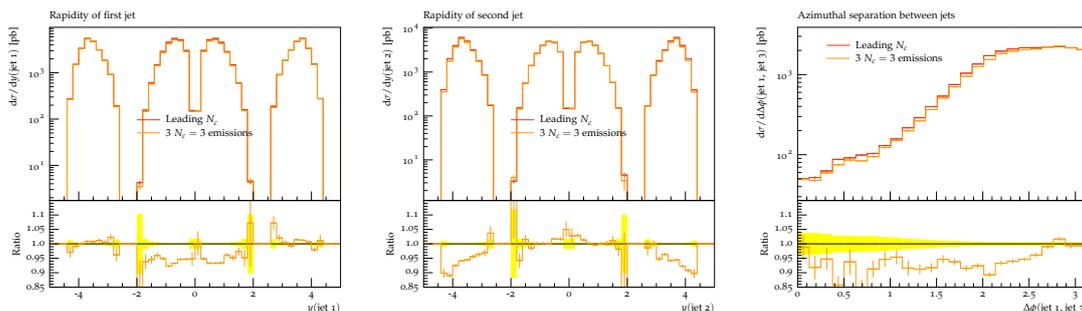
Using the cuts of [62],

$$\begin{aligned}
 400 \text{ GeV} < M_{12} < 600 \text{ GeV}, \\
 3.8 < |y_1 + y_2| < 5.2, \\
 1.5 < |y_2 - y_1| < 3.5,
 \end{aligned}
 \tag{7.1}$$

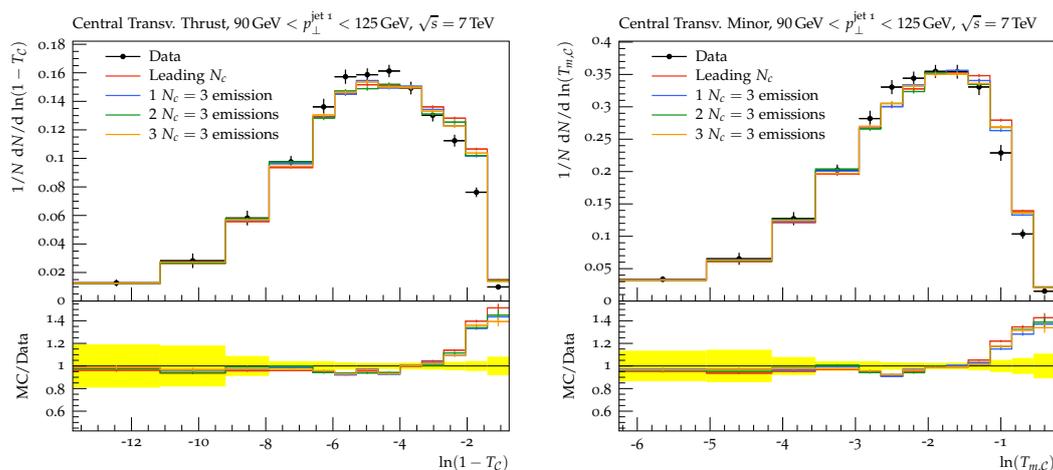
where  $y_1$  and  $y_2$  are the rapidities of the hardest and second hardest jet and  $M_{12}$  is the invariant mass of the two hardest jets, events dominated by the hard process  $qg \rightarrow qg$  can be statistically enhanced. These cuts select events with one of the two hardest jets being forward (in either direction) and one central. With these cuts we find differences of 5 – 10% for the rapidity distributions of the hardest three jets, as illustrated in figure 7 for the hardest jet. From figure 7, we see that in the  $N_c = 3$  shower, the hardest jet tends to be central less often as compared to the leading  $N_c$  shower. The rapidity distribution of the second hardest jet shows that it is forward less often. There are also 5 – 10% differences in  $\Delta\phi_{ij} = \phi_i - \phi_j$ ,  $\Delta\eta_{ij} = \eta_i - \eta_j$  and  $\Delta R_{ij} = \sqrt{\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2}$ , for  $i = 1, 2, j = 3$ . As an example  $\Delta\phi_{13}$  is also shown in figure 7. In general, with these cuts subleading  $N_c$  effects show sizable corrections for many standard QCD observables.

### 7.3.2 Hadron level analyses

We now turn our attention to hadronized events and to comparisons with LHC data. We have compared the subleading  $N_c$  corrected parton shower to experimental data for a wide range of QCD observables, using data from [63–69]. First, in figure 8, we consider the event



**Figure 7.** Rapidity distribution of the hardest jet (left), second hardest jet (middle) and separation in  $\phi$  of the hardest and third hardest jets (right). Our standard analysis cut of  $p_{\perp} > 50$  GeV is used.



**Figure 8.** Distribution of  $\log_e(1 - T_C)$  (left) and  $T_{m,c}$  (right) for  $\sqrt{s} = 7$  GeV. Data and Rivet analysis are taken from [63].

shape central transverse thrust, defined as [70]

$$T_C = \max_{\hat{n}_T} \frac{\sum_i |\vec{p}_{\perp,i} \cdot \hat{n}_T|}{\sum_i p_{\perp,i}}, \quad (7.2)$$

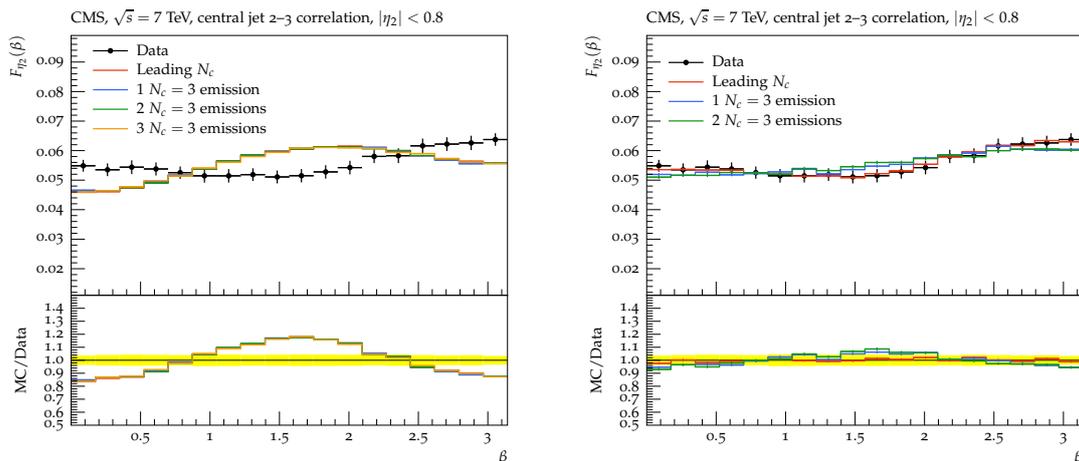
where  $p_{\perp,i}$  is the transverse momentum of the central jet  $i$ , having pseudorapidity  $\eta < 1.3$ , and  $\hat{n}_T$  is the direction perpendicular to the beam axis, which maximizes the sum. We also consider central thrust minor, defined again in terms of central jets with  $\eta < 1.3$ ,

$$T_{m,C} = \frac{\sum_i |\vec{p}_{\perp,i} \times \hat{n}_T|}{\sum_i p_{\perp,i}}, \quad (7.3)$$

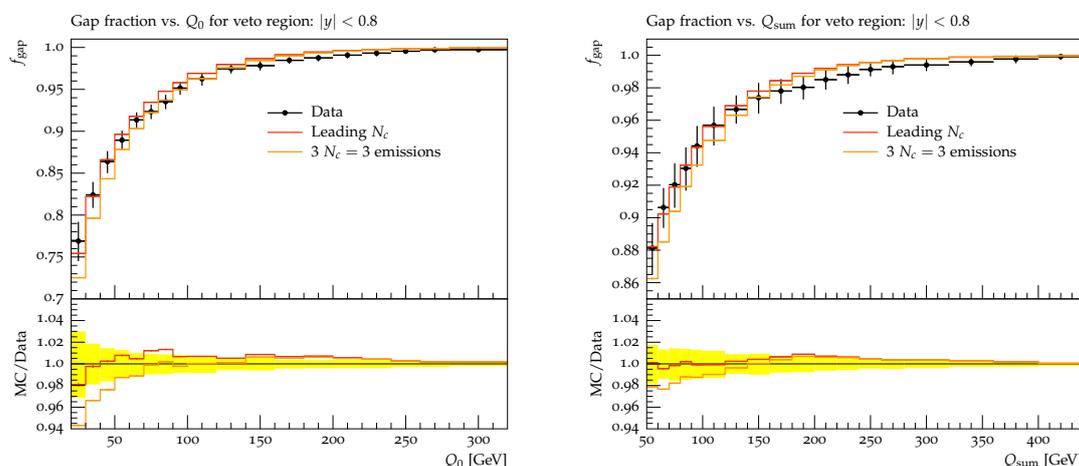
and compare to data from [63]. As can be seen we find relatively small corrections, at the 5% level or below.

A comparison to the jet shapes and jet masses for high  $p_{\perp}$  jets, from [64], shows yet smaller subleading  $N_c$  corrections, typically below a few percent.

We have also compared data to the so-called color coherence effects from [67]. In figure 9 we show the distribution of the  $\beta$ -angle, defined in terms of the pseudorapidities



**Figure 9.** The angle  $\beta$ , defined as in eq. (7.4) using (left) an underlying  $2 \rightarrow 2$  hard process and (right) an underlying  $2 \rightarrow 3$  hard process. Data and Rivet analysis are taken from [67].



**Figure 10.** Fraction of events having no additional jet with  $p_{\perp}$  above  $Q_0$  within a rapidity interval  $|y| < 0.8$  (left) and fraction of events where the scalar sum of transverse momenta within  $|y| < 0.8$  does not exceed  $Q_{\text{sum}}$  (right) for  $t\bar{t}$  events at  $\sqrt{s} = 7$  TeV. Data and Rivet analysis are taken from [65].

$\eta_2$  and  $\eta_3$  and the azimuthal angles  $\phi_2$  and  $\phi_3$ , as

$$\tan \beta = \frac{|\phi_3 - \phi_2|}{\text{sign}(\eta_2)(\eta_3 - \eta_2)}. \tag{7.4}$$

Experimental data is first compared to shower predictions using a  $2 \rightarrow 2$  hard matrix element, and then using a  $2 \rightarrow 3$ . We clearly see that the use of a  $2 \rightarrow 3$  hard matrix element significantly improves the description of data, whereas adding subleading  $N_c$  showering to the  $2 \rightarrow 2$  process changes the distribution compared to the leading  $N_c$  shower very marginally. This casts doubt upon the description of this observable as probing color coherence, and rather illustrates its dependence on the hard matrix element (or alternatively on other details of the shower algorithm).

We have also investigated the effects from subleading color contributions on top-pair production, and compared to data from [65, 66]. Here we find that the jet shapes from [66] are essentially unaltered, whereas the measurement of the additional jet activity in  $t\bar{t}$  events [65] show intriguing effects, displayed in figure 10. In particular we note that the data description improves in the region of a modest ratio between  $Q_0/Q_{\text{sum}}$  and the hard scale. For gap observables, like in figure 10, with a large scale hierarchy, we remark that we can expect effects from resummation of virtual gluons, which we do not include in this paper. These subleading  $N_c$  corrections may be sizable [26, 71–73].

As a clean test of initial state radiation, we have also compared the  $N_c = 3$  shower to event shapes in leptonic  $Z$  decay events, [69]. Here, as expected, having fewer colored particles in the hard process, we find very small corrections, at the percent level or below.

## 8 Conclusion and outlook

In this paper we have investigated the effect of keeping the full color structure in parton shower emissions in realistic simulations of LHC and LEP events. This is pursued within the dipole shower of the Herwig 7.1 framework [27] as color matrix element corrections.

The  $N_c = 3$  color shower corrects the first few (five for LEP and three for LHC) emissions using the full  $N_c = 3$  emission pattern. All hard observables we have studied, with the exception of observables explicitly considering very many jets, have converged with respect to keeping additional subleading  $N_c$  color corrected emissions. The convergence can also be underpinned by noting that the last emission to be corrected, corresponds to a low evolution scale, in comparison to the studied observables, see figure 3. In particular for LEP, we typically go down to evolution scales around one GeV.

Our results show that, for most QCD variables at the LHC, the subleading  $N_c$  effects are, similarly to at LEP, of the order of a few percent. At the LHC we can, however, see larger effects, 10 – 20%, for the tails of the rapidity distribution of the second hardest jet, figure 2, as well as for more tailored situations, cf. figure 7.

In a gedanken experiment, where quarks and gluons are collided, larger differences can be found, indicating that cancellation of subleading  $N_c$  effects are present at the LHC. To capture this situation, we consider LHC events while requiring that of the two most energetic jets, one is central and one is forward. In this case we find differences of 5 – 10% also for standard QCD observables. These differences are not only in the tails of the distributions, but over the whole range of several observables.

Turning to soft observables the situation is different. In many cases, including jet resolution variables at low scales, charged particle multiplicities (figure 5), individual hadron multiplicities and the number of very soft jets at LEP, we find large effects, of several 10%. While we cannot expect to make accurate predictions for any of these cases, due to sensitivity to hadronization, multiple interaction and resummation effects, it should be stressed that subleading  $N_c$  effects can be expected to play an important role for the final state of the shower, entering the hadronization. An immediate extension of this work is therefore retuning of the  $N_c = 3$  shower. Indeed, an improved description to (most) observables cannot be expected until retuning is performed.

Another natural next step is to include virtual corrections. Virtual gluon exchanges would allow rearrangement of the color structure without emission. This would be expected to have an effect on gap fraction observables, such as in figure 10, and we therefore caution that our conclusions regarding the magnitude of the subleading  $N_c$  corrections may not be applicable in these cases.

More precisely, virtual gluon exchanges should be the mechanism underlying (perturbative) color reconnection effects. In the longer perspective, it would be desirable to update the standard hadronization model to encompass a subleading  $N_c$  shower.

For these reasons, we like to stress that this work should be considered as the *start* of subleading  $N_c$  corrections at the LHC, not the *end*. Indeed much work remains to be done.

**Note added.** While this work has been finalized, a similar approach has been reported in [58]. In [58] color matrix element corrections, as well as the weighted Sudakov algorithm for final state evolution is used, but a sampling of color structure is advocated.

## Acknowledgments

We thank Johannes Bellm and Mike Seymour for discussions on color flows in dipole showers and Torbjörn Sjöstrand for comments on the manuscript. Johan Thorén wishes to thank Universität Wien for their hospitality. This work was supported by the Swedish Research Council (contract numbers 2012-02744 and 2016-05996), and in part by the MC-netITN3 H2020 Marie Curie Initial Training Network, contract number 722104, as well as the European Union’s Horizon 2020 research and innovation programme (grant agreement No 668679), and the COST action (“Unraveling new physics at the LHC through the precision frontier”) No. CA16201.

## A Proof of the modified weighted veto algorithm

In this appendix we will recapitulate the weighted Sudakov veto algorithm of [55], and prove that our modified algorithm generates the same probability distribution for competing processes.

The algorithm of [55] generates splittings at the scale  $q$  with  $d$  additional splitting variables  $x$  from a splitting kernel  $P(q, x)$  (which can be negative), according to the distribution

$$dS_P(\mu, x_\mu|q, x|Q) = dqd^d x \left[ \Delta_P(\mu|Q)\delta(q - \mu)\delta^{(d)}(x - x_\mu) + P(q, x)\Delta_P(q|Q)\theta(Q - q)\theta(q - \mu) \right], \quad (\text{A.1})$$

where  $Q$  is the starting scale,  $\mu$  is the cutoff scale,  $x_\mu$  is a parameter point associated with the cutoff,  $\theta$  is the Heaviside step function, where we use the convention  $\theta(0) = 1$ , and  $\Delta_P(q|Q)$  is the Sudakov form factor,

$$\Delta_P(q|Q) = \exp \left( - \int_q^Q dk \int d^d z P(k, z) \right). \quad (\text{A.2})$$

We note that the distribution  $S_P$  is normalized to unity, in the sense that

$$\int_{\mu}^Q \int dS_P(\mu, x_{\mu}|q', x'|Q) = \Delta_P(\mu|Q) + 1 - \Delta_P(\mu|Q) = 1, \quad (\text{A.3})$$

where the integrations are over  $q'$  and  $x'$ . As in [55] we use an integrable splitting kernel  $R(q, x)$  for which the primitive function can be inverted. This function is required to be positive everywhere, but actually does not need to be an overestimate of  $P(q, x)$ . We also define an acceptance probability  $\epsilon(q, x)$ , such that

$$0 < \epsilon(q, x) \leq 1. \quad (\text{A.4})$$

Now we add the modification to the algorithm, instead of a single weight that is updated upon each trial emission, we keep track of the scale and weight of every veto and accept step. The algorithm starts at a scale  $q_0 = Q$  and a step counter  $i = 0$ . The algorithm is as follows, for every competing splitting kernel:

1.  $i$  is incremented by one.
2. A trial splitting is generated, at the scale  $q_i$  and with splitting variables  $x$ , from  $S_R(\mu, x_{\mu}|q_i, x|q_{i-1})$ .
3. If  $q_i = \mu$ , the algorithm terminates without an emission, and returns the scales  $q_i$  and weights  $\omega_i$  for all  $i$ , that may have been acquired in the steps below.
4. The trial splitting is accepted with the probability  $\epsilon(q_i, x)$ . If accepted, the weight

$$\omega_i = \frac{1}{\epsilon(q_i, x)} \frac{P(q_i, x)}{R(q_i, x)}, \quad (\text{A.5})$$

is defined and the scales  $q_i$  and weights  $\omega_i$  for all  $i$ , along with the splitting variables  $x$ , are returned.

5. Otherwise, the weight

$$\omega_i = \frac{1}{1 - \epsilon(q_i, x)} \left( 1 - \frac{P(q_i, x)}{R(q_i, x)} \right), \quad (\text{A.6})$$

is stored and the algorithm repeats, starting at step 1.

In the algorithm of [55] the weight associated with this trial splitting is

$$\omega = \prod_{i=1}^{n_{\text{steps}}} \omega_i, \quad (\text{A.7})$$

where  $n_{\text{steps}}$  is the number of veto and accept steps. Our modification will use the weight

$$\omega = \prod_{i=1}^{n_{\text{steps}}} \omega_i \theta(q_i - q_w), \quad (\text{A.8})$$

where  $q_w$  is the scale of the winning splitting, i.e. we ignore weights below the winning scale.

Now we consider the competition algorithm. The standard proof of the competition algorithm does not depend on how the probability distribution,  $S_P$ , is generated, nor does it depend on the sign of the splitting kernels, hence the standard competition proof applies for the weighted Sudakov veto algorithm. To prove that the weight eq. (A.8) generates the same distribution as the weight eq. (A.7) when using the competition algorithm, we will consider the probability density for a specific scale  $q$  and splitting variables  $x$ , from the splitting kernel  $P_\alpha$ , competing with some other splitting kernels  $P_\beta$ ,  $\beta \neq \alpha$ . For a splitting kernel  $P_\alpha$ , the probability density for the algorithm defined by steps 1 to 5 above, starting at scale  $q_0$  and splitting variables  $x_0$ , to traverse a sequence of scales and splitting variables of length  $n$ , before emitting at the scale  $q$  with the splitting variables  $x$  (using the short-hand notation  $\epsilon_{\alpha,i} = \epsilon_\alpha(q_i, x_i)$ ) is

$$\begin{aligned}
 dS_{R_\alpha, \bar{\epsilon}_\alpha}^{(n)}(\mu, x_\mu; q, x | q_n, x_n | \dots | q_0, x_0) \\
 = \left[ \Delta_{R_\alpha}(\mu | q_0) \delta(q - \mu) \delta^{(d)}(x - x_\mu) \right. \\
 \left. + \epsilon_\alpha(q, x) R_\alpha(q, x) \Delta_{R_\alpha}(q | q_0) \theta(q_n - q) \theta(q - \mu) \right] dq d^d x \\
 \times \prod_{i=1}^n R_\alpha(q_i, x_i) (1 - \epsilon_{\alpha,i}) \theta(q_{i-1} - q_i) \theta(q_i - \mu) dq_i d^d x_i,
 \end{aligned} \tag{A.9}$$

where  $R_\alpha$  is the ‘‘nice’’ splitting kernel corresponding to the splitting kernel  $P_\alpha$ . The weight associated with this is

$$\begin{aligned}
 \omega_{P_\alpha, R_\alpha, \bar{\epsilon}_\alpha}^{(n)}(\mu; q, x | q_n, x_n | \dots | q_0, x_0) = \left( \prod_{i=1}^n \omega_i \left[ \begin{array}{c} 1 \\ \theta(q_i - q_w) \end{array} \right] \right) \\
 \times \left[ \begin{array}{c} 1 \\ \theta(q - q_w) \end{array} \right] \begin{cases} \frac{1}{\epsilon_\alpha(q, x)} \frac{P_\alpha(q, x)}{R_\alpha(q, x)} & \text{if } q > \mu \\ 1 & \text{if } q = \mu, \end{cases}
 \end{aligned} \tag{A.10}$$

where the upper element of the square brackets correspond to the weight from [55] and the lower element corresponds to the modified weight, and  $\omega_i$  is a function of  $q_i$  and  $x_i$  as seen from eq. (A.5) and eq. (A.6). The total probability density for the splitting variables  $q, x$  from the splitting kernel  $P_\alpha$ , competing with the kernels  $P_\beta$ , is

$$\begin{aligned}
 \sum_{n=0}^{\infty} \int_{q_1, x_1, \dots, q_n, x_n} dS_{R_\alpha, \bar{\epsilon}_\alpha}^{(n)}(\mu, x_\mu; q, x | q_n, x_n | \dots | q_0, x_0) \omega_{P_\alpha, R_\alpha, \bar{\epsilon}_\alpha}^{(n)}(\mu; q, x | q_n, x_n | \dots | q_0, x_0) \\
 \times \prod_{\beta \neq \alpha} \int_{q^{(\beta)}, x^{(\beta)}} \sum_{n_\beta=0}^{\infty} \int_{q_1^{(\beta)}, x_1^{(\beta)}, \dots, q_{n_\beta}^{(\beta)}, x_{n_\beta}^{(\beta)}} dS_{R_\beta, \bar{\epsilon}_\beta}^{(n_\beta)}(\mu, x_\mu; q^{(\beta)}, x^{(\beta)} | q_{n_\beta}^{(\beta)}, x_{n_\beta}^{(\beta)} | \dots | q_0, x_0) \\
 \times \omega_{P_\beta, R_\beta, \bar{\epsilon}_\beta}^{(n_\beta)}(\mu; q^{(\beta)}, x^{(\beta)} | q_{n_\beta}^{(\beta)}, x_{n_\beta}^{(\beta)} | \dots | q_0, x_0) \theta(q - q^{(\beta)}),
 \end{aligned} \tag{A.11}$$

where the product is over every competing kernel and the  $\theta$ -function is to enforce that the

winning scale is  $q$  from the splitting kernel  $P_\alpha$ . In the first line of eq. (A.11) we have

$$\begin{aligned}
 dS_{R_\alpha, \bar{\epsilon}_\alpha}^{(n)}(\mu, x_\mu; q, x|q_n, x_n | \dots |q_0, x_0) \omega_{P_\alpha, R_\alpha, \bar{\epsilon}_\alpha}^{(n)}(\mu; q, x|q_n, x_n | \dots |q_0, x_0) \\
 = \left[ \Delta_{R_\alpha}(\mu|q_0) \delta(q - \mu) \delta^{(d)}(x - x_\mu) + P_\alpha(q, x) \Delta_{R_\alpha}(q|q_0) \theta(q_n - q) \theta(q - \mu) \right] dq d^d x \\
 \times \prod_{i=1}^n (R_\alpha(q_i, x_i) - P_\alpha(q_i, x_i)) \theta(q_{i-1} - q_i) \theta(q_i - \mu) dq_i d^d x_i. \tag{A.12}
 \end{aligned}$$

Using this and extracting all of the terms of eq. (A.11) that depend on the intermediate scales,  $q_1, \dots, q_n$ , we have

$$\begin{aligned}
 \sum_{n=0}^{\infty} \int_q^{q_0} dq_1 \int d^d x_1 \int_q^{q_1} dq_2 \int d^d x_2 \dots \int_q^{q_{n-1}} dq_n \int d^d x_n \prod_{i=1}^n (R_\alpha(q_i, x_i) - P_\alpha(q_i, x_i)) \\
 = \sum_{n=0}^{\infty} \frac{1}{n!} \left[ \int_q^{q_0} dk \int d^d z (R_\alpha(k, z) - P_\alpha(k, z)) \right]^n \\
 = \exp \left( \int_q^{q_0} dk \int d^d z (R_\alpha(k, z) - P_\alpha(k, z)) \right), \tag{A.13}
 \end{aligned}$$

where the first equality uses that the ordered integrals can be exchanged for unordered integrals divided by  $1/n!$ . Hence,  $\Delta_{R_\alpha}$  is canceled in eq. (A.12) and replaced with  $\Delta_{P_\alpha}$ . The first line in eq. (A.11) is then exactly  $dS_{P_\alpha}(\mu, x_\mu|q, x|q_0)$ . If we consider one of the factors in the product over  $\beta$  in eq. (A.11), noting that the integration, using the  $\theta$ -function, has the limits  $q^{(\beta)} = \mu$  and  $q^{(\beta)} = q$ , we have

$$\begin{aligned}
 \int_{q^{(\beta)}=\mu}^{q^{(\beta)}=q} \int_{x^{(\beta)}} \sum_{n_\beta=0}^{\infty} \int_{q_1^{(\beta)}, x_1^{(\beta)}, \dots, q_{n_\beta}^{(\beta)}, x_{n_\beta}^{(\beta)}} dS_{R_\beta, \bar{\epsilon}_\beta}^{(n_\beta)}(\mu, x_\mu; q^{(\beta)}, x^{(\beta)}|q_{n_\beta}^{(\beta)}, x_{n_\beta}^{(\beta)} | \dots |q_0, x_0) \\
 \times \omega_{P_\beta, R_\beta, \bar{\epsilon}_\beta}^{(n_\beta)}(\mu; q^{(\beta)}, x^{(\beta)}|q_{n_\beta}^{(\beta)}, x_{n_\beta}^{(\beta)} | \dots |q_0, x_0). \tag{A.14}
 \end{aligned}$$

If the weight eq. (A.7) is used, the argument in eq. (A.13) can be used to show that eq. (A.14) is

$$\begin{aligned}
 \int_{q^{(\beta)}=\mu}^{q^{(\beta)}=q} \int_{x^{(\beta)}} dS_{P_\beta}(\mu, x_\mu; q^{(\beta)}, x^{(\beta)}|q_0) = \Delta_{P_\beta}(\mu|q_0) + \Delta_{P_\beta}(q|q_0) - \Delta_{P_\beta}(\mu|q_0) \\
 = \Delta_{P_\beta}(q|q_0), \tag{A.15}
 \end{aligned}$$

where the first term comes from the first term in eq. (A.1) and the other two terms come from the second term in eq. (A.1). If the modified weight, eq. (A.8), is used, the veto algorithm no longer generates the splitting function  $P_\beta$  for all scales. The modified weights will generate emissions according to the splitting kernel

$$P'_\beta(q, x) = \begin{cases} \epsilon_\beta(q, x) R_\beta(q, x) & \text{for } q < q_w, \\ P_\beta(q, x) & \text{for } q > q_w, \end{cases} \tag{A.16}$$

where the behavior below  $q_w$  can be seen from eq. (A.9). Above  $q_w$  it generates  $P_\beta$  as the weights are identical. Hence, if using the modified weights, eq. (A.14) becomes

$$\Delta_{P'_\beta}(\mu|q_0) + \Delta_{P'_\beta}(q|q_0) - \Delta_{P'_\beta}(\mu|q_0) = \Delta_{P'_\beta}(q|q_0) = \Delta_{P_\beta}(q|q_0), \tag{A.17}$$

where the second equality uses eq. (A.16). We now see that eq. (A.15) and eq. (A.17) are equal, so the modified weights give the correct Sudakov form factor and the distribution in eq. (A.11) will be the same.

To conclude, the physical interpretation of the argument in this appendix, is that emissions below the winning scale do not affect the physics above it. Hence, we are free to use the weights in eq. (A.8), which offer a better convergence for the algorithm.

## B Choices for the modified veto algorithm

This appendix contains the details for the choice of acceptance probability and overestimate kernel for the modified weighted Sudakov veto algorithm. The arguments of the splitting kernels,  $p_{\perp}^2$ ,  $z$ ,  $p_{\tilde{j}}$  and  $p_{\tilde{k}}$ , have been suppressed for clarity in the following equations.

We use the modified version of the weighted sudakov veto algorithm from [55], as described in section 5. The splitting kernel we aim at generating,  $\mathcal{P}_{ij,k}$ , is given in eq. (5.3). Defining  $\mathcal{P}_{ij,k}^{N_c \rightarrow \infty}$  to be the leading color limit of  $\mathcal{P}_{ij,k}$ , assuming  $\tilde{j}$  and  $\tilde{k}$  are color connected (cf. eq. (2.8)), and letting  $\mathcal{R}_{ij,k}^{N_c \rightarrow \infty}$  be the standard overestimate used in Herwig, as obtained from the ExSample [6] adapted grid proposal, we can write

$$\mathcal{P}_{ij,k} = \text{sign}(\omega_{\tilde{j}\tilde{k}}^n) |\omega_{\tilde{j}\tilde{k}}^n| (1 + \delta_{\tilde{j}}) \mathcal{P}_{ij,k}^{N_c \rightarrow \infty} . \tag{B.1}$$

Our choice for the overestimate kernel is  $\mathcal{R}_{ij,k} = |\omega_{\tilde{j}\tilde{k}}^n| (1 + \delta_{\tilde{j}}) \mathcal{R}_{ij,k}^{N_c \rightarrow \infty}$  and our choice for the acceptance probability is

$$\epsilon = \frac{|\mathcal{P}_{ij,k}|}{\mathcal{R}_{ij,k}} = \frac{\mathcal{P}_{ij,k}^{N_c \rightarrow \infty}}{\mathcal{R}_{ij,k}^{N_c \rightarrow \infty}} , \tag{B.2}$$

i.e. we use the same acceptance probability as for the leading  $N_c$  shower (if  $\tilde{j}$ ,  $\tilde{k}$  were color connected). The acceptance weight is then

$$\text{sign}(\omega_{\tilde{j}\tilde{k}}^n) = \begin{cases} 1 & \text{for } \omega_{\tilde{j}\tilde{k}}^n > 0 \\ -1 & \text{for } \omega_{\tilde{j}\tilde{k}}^n < 0 \end{cases} , \tag{B.3}$$

and the veto weight is

$$\frac{\mathcal{R}_{ij,k}^{N_c \rightarrow \infty} - \text{sign}(\omega_{\tilde{j}\tilde{k}}^n) \mathcal{P}_{ij,k}^{N_c \rightarrow \infty}}{\mathcal{R}_{ij,k}^{N_c \rightarrow \infty} - \mathcal{P}_{ij,k}^{N_c \rightarrow \infty}} = \begin{cases} 1 & \text{for } \omega_{\tilde{j}\tilde{k}}^n > 0 \\ \frac{\mathcal{R}_{ij,k}^{N_c \rightarrow \infty} + \mathcal{P}_{ij,k}^{N_c \rightarrow \infty}}{\mathcal{R}_{ij,k}^{N_c \rightarrow \infty} - \mathcal{P}_{ij,k}^{N_c \rightarrow \infty}} & \text{for } \omega_{\tilde{j}\tilde{k}}^n < 0 \end{cases} . \tag{B.4}$$

From eq. (B.3) and eq. (B.4) we see that trial emissions corresponding to positive  $\omega_{\tilde{j}\tilde{k}}^n$  never change the event weight.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] 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](#)].
- [2] M. Bahr et al., *HERWIG++ physics and manual*, *Eur. Phys. J. C* **58** (2008) 639 [[arXiv:0803.0883](#)] [[INSPIRE](#)].
- [3] J. Bellm et al., *HERWIG 7.0/HERWIG++ 3.0 release note*, *Eur. Phys. J. C* **76** (2016) 196 [[arXiv:1512.01178](#)] [[INSPIRE](#)].
- [4] T. Gleisberg et al., *SHERPA 1.alpha: a proof of concept version*, *JHEP* **02** (2004) 056 [[hep-ph/0311263](#)] [[INSPIRE](#)].
- [5] S. Platzer and M. Sjödal, *The Sudakov veto algorithm reloaded*, *Eur. Phys. J. Plus* **127** (2012) 26 [[arXiv:1108.6180](#)] [[INSPIRE](#)].
- [6] S. Plätzer, *ExSample: a library for sampling Sudakov-type distributions*, *Eur. Phys. J. C* **72** (2012) 1929 [[arXiv:1108.6182](#)] [[INSPIRE](#)].
- [7] S. Plätzer and S. Gieseke, *Dipole showers and automated NLO matching in HERWIG++*, *Eur. Phys. J. C* **72** (2012) 2187 [[arXiv:1109.6256](#)] [[INSPIRE](#)].
- [8] S. Plätzer, *Controlling inclusive cross sections in parton shower + matrix element merging*, *JHEP* **08** (2013) 114 [[arXiv:1211.5467](#)] [[INSPIRE](#)].
- [9] J. Bellm, S. Gieseke and S. Plätzer, *Merging NLO Multi-jet Calculations with Improved Unitarization*, *Eur. Phys. J. C* **78** (2018) 244 [[arXiv:1705.06700](#)] [[INSPIRE](#)].
- [10] M. Dasgupta et al., *Logarithmic accuracy of parton showers: a fixed-order study*, *JHEP* **09** (2018) 033 [[arXiv:1805.09327](#)] [[INSPIRE](#)].
- [11] L. Lönnblad, *Correcting the color dipole cascade model with fixed order matrix elements*, *JHEP* **05** (2002) 046 [[hep-ph/0112284](#)] [[INSPIRE](#)].
- [12] F. Krauss, *Matrix elements and parton showers in hadronic interactions*, *JHEP* **08** (2002) 015 [[hep-ph/0205283](#)] [[INSPIRE](#)].
- [13] S. Hoeche et al., *Matching parton showers and matrix elements*, [hep-ph/0602031](#) [[INSPIRE](#)].
- [14] N. Lavesson and L. Lönnblad, *Merging parton showers and matrix elements: back to basics*, *JHEP* **04** (2008) 085 [[arXiv:0712.2966](#)] [[INSPIRE](#)].
- [15] S. Hoeche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053 [[arXiv:0903.1219](#)] [[INSPIRE](#)].
- [16] K. Hamilton, P. Richardson and J. Tully, *A modified CKKW matrix element merging approach to angular-ordered parton showers*, *JHEP* **11** (2009) 038 [[arXiv:0905.3072](#)] [[INSPIRE](#)].
- [17] M. Dobbs, *Phase space veto method for next-to-leading order event generators in hadronic collisions*, *Phys. Rev. D* **65** (2002) 094011 [[hep-ph/0111234](#)] [[INSPIRE](#)].
- [18] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029 [[hep-ph/0204244](#)] [[INSPIRE](#)].
- [19] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040 [[hep-ph/0409146](#)] [[INSPIRE](#)].
- [20] Z. Nagy and D.E. Soper, *Matching parton showers to NLO computations*, *JHEP* **10** (2005) 024 [[hep-ph/0503053](#)] [[INSPIRE](#)].

- [21] S. Frixione, F. Stoeckli, P. Torrielli and B.R. Webber, *NLO QCD corrections in HERWIG++ with MC@NLO*, *JHEP* **01** (2011) 053 [[arXiv:1010.0568](#)] [[INSPIRE](#)].
- [22] S. Platzer and M. Sjö Dahl, *Subleading  $N_c$  improved parton showers*, *JHEP* **07** (2012) 042 [[arXiv:1201.0260](#)] [[INSPIRE](#)].
- [23] Z. Nagy and D.E. Soper, *Parton shower evolution with subleading color*, *JHEP* **06** (2012) 044 [[arXiv:1202.4496](#)] [[INSPIRE](#)].
- [24] Z. Nagy and D.E. Soper, *Effects of subleading color in a parton shower*, *JHEP* **07** (2015) 119 [[arXiv:1501.00778](#)] [[INSPIRE](#)].
- [25] S. Plätzer, *Summing large- $N$  towers in colour flow evolution*, *Eur. Phys. J. C* **74** (2014) 2907 [[arXiv:1312.2448](#)] [[INSPIRE](#)].
- [26] R. Ángeles Martínez et al., *Soft gluon evolution and non-global logarithms*, *JHEP* **05** (2018) 044 [[arXiv:1802.08531](#)] [[INSPIRE](#)].
- [27] J. Bellm et al., *HERWIG 7.1 release note*, [arXiv:1705.06919](#) [[INSPIRE](#)].
- [28] S. Plätzer and S. Gieseke, *Coherent parton showers with local recoils*, *JHEP* **01** (2011) 024 [[arXiv:0909.5593](#)] [[INSPIRE](#)].
- [29] S. Catani and M.H. Seymour, *The dipole formalism for the calculation of QCD jet cross-sections at next-to-leading order*, *Phys. Lett. B* **378** (1996) 287 [[hep-ph/9602277](#)] [[INSPIRE](#)].
- [30] S. Catani and M.H. Seymour, *A general algorithm for calculating jet cross-sections in NLO QCD*, *Nucl. Phys. B* **485** (1997) 291 [*Erratum ibid.* **B 510** (1998) 503] [[hep-ph/9605323](#)] [[INSPIRE](#)].
- [31] K. Cormier et al., *Parton shower and matching uncertainties in top quark pair production with Herwig 7*, [arXiv:1810.06493](#).
- [32] S. Catani, S. Dittmaier, M.H. Seymour and Z. Trócsányi, *The dipole formalism for next-to-leading order QCD calculations with massive partons*, *Nucl. Phys. B* **627** (2002) 189 [[hep-ph/0201036](#)] [[INSPIRE](#)].
- [33] G. Gustafson and U. Pettersson, *Dipole formulation of QCD cascades*, *Nucl. Phys. B* **306** (1988) 746 [[INSPIRE](#)].
- [34] J. Bellm, *Colour rearrangement for dipole showers*, *Eur. Phys. J. C* **78** (2018) 601 [[arXiv:1801.06113](#)] [[INSPIRE](#)].
- [35] T. Sjöstrand, *A model for initial state parton showers*, *Phys. Lett. B* **157** (1985) 321.
- [36] J.E. Paton and H.-M. Chan, *Generalized Veneziano model with isospin*, *Nucl. Phys. B* **10** (1969) 516 [[INSPIRE](#)].
- [37] F.A. Berends and W. Giele, *The six gluon process as an example of Weyl-Van der Waerden spinor calculus*, *Nucl. Phys. B* **294** (1987) 700 [[INSPIRE](#)].
- [38] M.L. Mangano, S.J. Parke and Z. Xu, *Duality and multi-gluon scattering*, *Nucl. Phys. B* **298** (1988) 653 [[INSPIRE](#)].
- [39] M.L. Mangano, *The color structure of gluon emission*, *Nucl. Phys. B* **309** (1988) 461 [[INSPIRE](#)].
- [40] D.A. Kosower, *Color factorization for fermionic amplitudes*, *Nucl. Phys. B* **315** (1989) 391 [[INSPIRE](#)].

- [41] Z. Nagy and D.E. Soper, *Parton showers with quantum interference*, *JHEP* **09** (2007) 114 [[arXiv:0706.0017](#)] [[INSPIRE](#)].
- [42] M. Sjödalh, *Color structure for soft gluon resummation: A General recipe*, *JHEP* **09** (2009) 087 [[arXiv:0906.1121](#)] [[INSPIRE](#)].
- [43] J. Alwall et al., *MadGraph 5: going beyond*, *JHEP* **06** (2011) 128 [[arXiv:1106.0522](#)] [[INSPIRE](#)].
- [44] M. Sjödalh, *ColorFull — A C++ library for calculations in  $SU(N_c)$  color space*, *Eur. Phys. J. C* **75** (2015) 236 [[arXiv:1412.3967](#)] [[INSPIRE](#)].
- [45] A. Kyrieleis and M.H. Seymour, *The colour evolution of the process  $qq \rightarrow qgg$* , *JHEP* **01** (2006) 085 [[hep-ph/0510089](#)] [[INSPIRE](#)].
- [46] Yu.L. Dokshitzer and G. Marchesini, *Soft gluons at large angles in hadron collisions*, *JHEP* **01** (2006) 007 [[hep-ph/0509078](#)] [[INSPIRE](#)].
- [47] M. Sjödalh, *Color evolution of  $2 \rightarrow 3$  processes*, *JHEP* **12** (2008) 083 [[arXiv:0807.0555](#)] [[INSPIRE](#)].
- [48] M. Beneke, P. Falgari and C. Schwinn, *Soft radiation in heavy-particle pair production: all-order colour structure and two-loop anomalous dimension*, *Nucl. Phys. B* **828** (2010) 69 [[arXiv:0907.1443](#)] [[INSPIRE](#)].
- [49] S. Keppeler and M. Sjödalh, *Orthogonal multiplet bases in  $SU(N_c)$  color space*, *JHEP* **09** (2012) 124 [[arXiv:1207.0609](#)] [[INSPIRE](#)].
- [50] Y.J. Du, M. Sjödalh and J. Thorén, *Recursion in multiplet bases for tree-level MHV gluon amplitudes*, *JHEP* **05** (2015) 119 [[arXiv:1503.00530](#)] [[INSPIRE](#)].
- [51] M. Sjödalh and J. Thorén, *Decomposing color structure into multiplet bases*, *JHEP* **09** (2015) 055 [[arXiv:1507.03814](#)] [[INSPIRE](#)].
- [52] G. 't Hooft, *A planar diagram theory for strong interactions*, *Nucl. Phys. B* **72** (1974) 461 [[INSPIRE](#)].
- [53] A. Kanaki and C.G. Papadopoulos, *HELAC-PHEGAS: automatic computation of helicity amplitudes and cross-sections*, [hep-ph/0012004](#) [[INSPIRE](#)].
- [54] F. Maltoni, K. Paul, T. Stelzer and S. Willenbrock, *Color flow decomposition of QCD amplitudes*, *Phys. Rev. D* **67** (2003) 014026 [[hep-ph/0209271](#)] [[INSPIRE](#)].
- [55] J. Bellm et al., *Reweighting parton showers*, *Phys. Rev. D* **94** (2016) 034028 [[arXiv:1605.08256](#)] [[INSPIRE](#)].
- [56] M. Cacciari, G.P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) 063 [[arXiv:0802.1189](#)] [[INSPIRE](#)].
- [57] A. Buckley et al., *Rivet user manual*, *Comput. Phys. Commun.* **184** (2013) 2803 [[arXiv:1003.0694](#)] [[INSPIRE](#)].
- [58] J. Isaacson and S. Prestel, *On stochastically sampling color configurations*, [arXiv:1806.10102](#) [[INSPIRE](#)].
- [59] DELPHI collaboration, P. Abreu et al., *Tuning and test of fragmentation models based on identified particles and precision event shape data*, *Z. Phys. C* **73** (1996) 11 [[INSPIRE](#)].
- [60] ALEPH collaboration, R. Barate et al., *Studies of quantum chromodynamics with the ALEPH detector*, *Phys. Rept.* **294** (1998) 1 [[INSPIRE](#)].

- [61] ALEPH collaboration, A. Heister et al., *Studies of QCD at  $e^+e^-$  centre-of-mass energies between 91 GeV and 209 GeV*, *Eur. Phys. J. C* **35** (2004) 457 [[INSPIRE](#)].
- [62] S. Gieseke, S. Plätzer and R. Podskubka, *From jet topology to jet tagging*, in preparation.
- [63] CMS collaboration, *First measurement of hadronic event shapes in pp collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Lett. B* **699** (2011) 48 [[arXiv:1102.0068](#)] [[INSPIRE](#)].
- [64] ATLAS collaboration, *ATLAS measurements of the properties of jets for boosted particle searches*, *Phys. Rev. D* **86** (2012) 072006 [[arXiv:1206.5369](#)] [[INSPIRE](#)].
- [65] ATLAS collaboration, *Measurement of  $t\bar{t}$  production with a veto on additional central jet activity in pp collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **72** (2012) 2043 [[arXiv:1203.5015](#)] [[INSPIRE](#)].
- [66] ATLAS collaboration, *Measurement of jet shapes in top-quark pair events at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **73** (2013) 2676 [[arXiv:1307.5749](#)] [[INSPIRE](#)].
- [67] CMS collaboration, *Probing color coherence effects in pp collisions at  $\sqrt{s} = 7$  TeV*, *Eur. Phys. J. C* **74** (2014) 2901 [[arXiv:1311.5815](#)] [[INSPIRE](#)].
- [68] ATLAS collaboration, *Measurements of jet vetoes and azimuthal decorrelations in dijet events produced in pp collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **74** (2014) 3117 [[arXiv:1407.5756](#)] [[INSPIRE](#)].
- [69] ATLAS collaboration, *Measurement of event-shape observables in  $Z \rightarrow \ell^+\ell^-$  events in pp collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector at the LHC*, *Eur. Phys. J. C* **76** (2016) 375 [[arXiv:1602.08980](#)] [[INSPIRE](#)].
- [70] A. Banfi, G.P. Salam and G. Zanderighi, *Phenomenology of event shapes at hadron colliders*, *JHEP* **06** (2010) 038 [[arXiv:1001.4082](#)] [[INSPIRE](#)].
- [71] J.R. Forshaw and M. Sjö Dahl, *Soft gluons in Higgs plus two jet production*, *JHEP* **09** (2007) 119 [[arXiv:0705.1504](#)] [[INSPIRE](#)].
- [72] B.E. Cox, J.R. Forshaw and A.D. Pilkington, *Extracting Higgs boson couplings using a jet veto*, *Phys. Lett. B* **696** (2011) 87 [[arXiv:1006.0986](#)] [[INSPIRE](#)].
- [73] R.M. Duran Delgado, J.R. Forshaw, S. Marzani and M.H. Seymour, *The dijet cross section with a jet veto*, *JHEP* **08** (2011) 157 [[arXiv:1107.2084](#)] [[INSPIRE](#)].