Journal of High Energy Physics

GenEvA (I): a new framework for event generation

To cite this article: Christian W. Bauer *et al* JHEP12(2008)010

View the article online for updates and enhancements.

You may also like

- Finite-size effects and finite-size scaling in time evolution during a colorless confining phase transition
Salah Cherif, Madjid Lakhdar Hamou Ladrem, Zainab Zaki Mohammed Alfull et al.
- Effect of hadronic transport processes on HBT radii in the AMPT model
Yao Zhang
- The muon Smasher's guide
Hind Al Ali, Nima Arkani-Hamed, Ian Banta et al.

This content was downloaded from IP address 173.177.195.252 on 26/02/2026 at 17:11

Published by IOP Publishing for SISSA
Received: *September 6, 2008*
Accepted: *November 10, 2008*
Published: *December 2, 2008*

GenEvA (I): a new framework for event generation

Christian W. Bauer, Frank J. Tackmann and Jesse Thaler

Ernest Orlando Lawrence Berkeley National Laboratory, University of California, 1 Cyclotron Road, Berkeley, CA 94720, U.S.A.
E-mail: cwbauer@lbl.gov, ftackmann@lbl.gov, jthaler@jthaler.net

Abstract: We show how many contemporary issues in event generation can be recast in terms of partonic calculations with a matching scale. This framework is called GenEvA, and a key ingredient is a new notion of phase space which avoids the problem of phase space double-counting by construction and includes a built-in definition of a matching scale. This matching scale can be used to smoothly merge any partonic calculation with a parton shower. The best partonic calculation for a given region of phase space can be determined through physics considerations alone, independent of the algorithmic details of the merging. As an explicit example, we construct a positive-weight partonic calculation for $e^+e^- \to n$ jets at next-to-leading order (NLO) with leading-logarithmic (LL) resummation. We improve on the NLO/LL result by adding additional higher-multiplicity tree-level (LO) calculations to obtain a merged NLO/LO/LL result. These results are implemented using a new phase space generator introduced in a companion paper [1].

Keywords: Jets, Phenomenological Models, NLO Computations, QCD.

© SISSA 2008

JHEP12(2008)010

Contents

1. Introduction **2**

2. Overview of the GenEvA framework **4**
- 2.1 The challenge of QCD 4
- 2.2 Combining different QCD descriptions 6
- 2.3 A new approach to phase space 10
- 2.4 Variable resolution scale 12
- 2.5 Strategy for partonic calculations 14
- 2.6 Summary of GenEvA 16

3. GenEvA in a toy example **18**
- 3.1 Review of toy theory 18
- 3.2 Phase space with a matching scale 20
- 3.3 The master formula 23
- 3.4 Partonic matrix elements for the first emission 24
- 3.5 Multiple emissions 28
- 3.6 The importance of phase space projection 30

4. The first emission in QCD **32**
- 4.1 Three-body phase space 32
- 4.2 The master formula 35
- 4.3 Relevant QCD calculations 35
- 4.4 Partonic cross sections at LO 37
- 4.5 Partonic cross sections at NLO 40

5. NLO/LO/LL merging **42**
- 5.1 Generalized master formula 42
- 5.2 Sudakov improvements 43
- 5.3 Merging nested descriptions 45

6. Results **47**
- 6.1 Total cross section scale dependence 48
- 6.2 Interpolating kinematic extremes 50
- 6.3 More results from GenEvA best 53

7. Conclusions **56**

JHEP12(2008)010

1. Introduction

Despite facing a complicated detector environment, top quark [2] and W boson [3, 4] studies at the Tevatron have proved that it is nevertheless possible to make precision measurements at hadron colliders. With the upcoming Large Hadron Collider (LHC), a complete understanding of Standard Model backgrounds will be essential for discovering new physics at the energy frontier [5 – 7]. Therefore, precision theoretical calculations are needed to complement the increasingly sophisticated experimental techniques available at hadron colliders. At lepton colliders, data is often compared to theoretical predictions for inclusive quantities, but at hadron colliders, it is more typical for data to be compared to theoretical predictions for exclusive quantities in order to more readily apply experimental cuts that may not be well defined in an inclusive theoretical framework.

Monte Carlo programs have proved indispensable for making exclusive theoretical predictions. Together with parton distribution functions, hadronization models, and underlying event models, traditional event generators [8 – 20] agree remarkably well with Tevatron data over a wide variety of experimental observables [21, 22]. However, in anticipation of further theoretical progress in refining Standard Model (and Beyond the Standard Model) predictions, it is worthwhile to consider possible improvements to the traditional Monte Carlo approach. There has been much work in recent years on merging fixed-order matrix element calculations with parton showers [23 – 35]. Furthermore, there are several programs which implement higher-order calculations to produce inclusive cross sections [36 – 39], though these next-to-leading (NLO) order programs are not cast in the form of a Monte Carlo program that can generate fully-hadronized exclusive events. While several ideas exist in the literature of how to implement a combination of NLO results and parton showers [40 – 53], only a few publicly available programs currently exist [54, 55] that implement this properly.

At the level of partons, current event generators are based on two independent frameworks: fixed-order matrix element calculations and parton showers. While fixed-order calculations include quantum interference and can be systematically improved through perturbative loop calculations, parton showers are always necessary to generate additional QCD radiation in a Monte Carlo framework. This is because fixed-order calculations cannot handle the large number of final states typically present in high-energy collisions. Schematically, the way the fully differential hadronic cross section $\mathrm{d}\sigma$ is achieved in a traditional approach is through

$$\mathrm{d}\sigma^{\mathrm{trad.}} = \mathrm{MC}\left(|\mathcal{M}|^2\,\mathrm{d}\Phi\right)\,, \tag{1.1}$$

where $|\mathcal{M}|^2$ represents a fixed-order QCD calculation, $\mathrm{d}\Phi$ represents a fixed-multiplicity phase space algorithm, and MC represents the action of a showering/hadronization scheme. The main challenge of eq. (1.1) is that the actual partonic four-momenta are generated both through the action of $\mathrm{d}\Phi$ and MC, and in order to avoid phase space double-counting, either $|\mathcal{M}|^2$ has to be modified from the value calculated in QCD, or the action of MC has to modified to accommodate the fixed-order calculation, or both.

In this paper, we present a new Monte Carlo framework `GenEvA`— for **Gen**erate **Ev**ents **A**nalytically — that allows almost any parton-level calculation to be translated

JHEP12(2008)010

into hadron-level events in a generic way with only a minimal modification of field theoretic methods.[1] The key idea is to separate the physics considerations, which determine the appropriate distribution to use, from the algorithmic details, which define how phase space is generated and what models are used to perform additional showering and hadronization. As we will see, it is possible to separate the physics and algorithmic issues using a matching scale μ. Schematically, the fully differential hadronic cross section $\mathrm{d}\sigma$ in `GenEvA` is given by

$$\mathrm{d}\sigma^{\texttt{GenEvA}} = |\mathcal{M}(\mu)|^2 \, \mathrm{dMC}(\mu) \,. \tag{1.2}$$

The quantity $\mathrm{dMC}(\mu)$ represents both a phase space algorithm *and* a showering/hadronization scheme that starts generating radiation at the scale μ. The only information the "partonic calculation" $|\mathcal{M}(\mu)|^2$ needs to know about the specific implementation of phase space generation and showering is this matching scale μ. From the point of view of the partonic calculation, μ is an infrared scale and can be thought of as the scale at which the partonic calculation is interfaced with a parton shower. We will see that the details of this matching are independent of any specific parton shower algorithm. While the partonic calculation $|\mathcal{M}(\mu)|^2$ does not necessarily correspond to a traditional QCD amplitude,[2] we will show various ways that $|\mathcal{M}(\mu)|^2$ can be determined solely in the context of perturbative QCD, independent of the algorithmic details of $\mathrm{dMC}(\mu)$.

In this way, `GenEvA` allows the user to focus on determining the most accurate differential cross section in a given region of phase space, instead of trying to figure out an algorithm which not only covers phase space with the right distributions, but also interfaces cleanly with a showering/hadronization scheme. As a concrete example of the power of the `GenEvA` framework, we will present a conceptually simple implementation of two known methods to improve the accuracy of Monte Carlo: parton shower/matrix element merging (PS/ME) [23 – 35] and Monte Carlo at next-to-leading order (PS/NLO) [40 – 53]. In previous approaches, PS/ME merging and PS/NLO merging relied on different Monte Carlo algorithms, but in the context of `GenEvA`, they correspond just to different choices for $|\mathcal{M}(\mu)|^2$ and share the same algorithmic underpinnings.

In fact, `GenEvA` allows PS/ME merging and PS/NLO merging to be *combined* to create a positive-weight Monte Carlo sample that merges NLO information with higher-order tree-level (LO) matrix elements and leading-logarithmic (LL) resummation. The simplicity with which we achieve an NLO/LO/LL merged sample through a special choice for $|\mathcal{M}(\mu)|^2$ suggests obvious generalizations to including next-to-next-to-leading order (NNLO) or next-to-leading logarithmic (NLL) information in event generators, though we will not pursue those directions in the present work.

Our focus will be on getting the most accurate distribution possible using available theoretical tools, and not necessarily on generating these distributions efficiently. In a companion paper [1], we describe a $\mathrm{dMC}(\mu)$ generator based on parton shower reweighting [63],

[1] This framework is unrelated to the "Geneva" jet algorithm [56].

[2] In the context of soft-collinear effective theory (SCET) [57 – 60], $|\mathcal{M}(\mu)|^2$ corresponds to the square of a Wilson coefficient and $\mathrm{dMC}(\mu)$ corresponds to the matrix element of an SCET operator [61, 62]. This explains why there should be a cancellation of the μ dependence in $|\mathcal{M}(\mu)|^2 \mathrm{dMC}(\mu)$. While SCET offers formal definitions for what $|\mathcal{M}(\mu)|^2$, $\mathrm{dMC}(\mu)$, and μ are, these formal definitions are not necessarily required.

JHEP12(2008)010

and we will see that certain gains in efficiency come as a bonus for using the parton shower as a phase space generator. However, the GenEvA framework is more general than the specific algorithmic implementation in ref. [1]. Though we anticipate extending GenEvA to handle the full range of Standard Model processes at hadron colliders, we will focus here on the process $e^+e^- \to n$ jets, which is sufficiently complicated to highlight all of the novel techniques introduced by GenEvA. In the companion paper [1], we comment on additional technical issues that arise in trying to understand Tevatron or LHC physics.

In the next section, we review the challenges faced when constructing Monte Carlo tools and the insights offered by GenEvA. In section 3, we discuss the GenEvA framework in a simple toy model. We then transition to the more complicated case of QCD, where we discuss the first emission in detail in section 4 and extend the results to multiple emissions in section 5. After showing results from the GenEvA program in section 6, we conclude in section 7.

2. Overview of the GenEvA framework

2.1 The challenge of QCD

To build a perfect Monte Carlo program, one would need a complete description of the Standard Model that is valid over all of phase space. Unfortunately, no such description exists, mainly because we do not have a complete description of QCD that is valid for every energy scale and every kinematic configuration. Instead, we know various limits of QCD, various perturbative expansions of QCD, and various phenomenological models based on QCD. The best Standard Model Monte Carlo program we can hope to build is one that coherently combines as many descriptions of QCD as possible.

One crucial combination that has been the subject of many recent advances in Monte Carlo is the merging of fixed-order calculations with parton showers. Fixed-order calculations are reliable when pairs of partons are well separated in phase space, while parton showers are reliable in the soft-collinear limit. For this reason, a Monte Carlo program that combines both descriptions has a better chance to correctly describe experimental data over a broad range of observables and energy scales. The problem with the language of "fixed-order calculation" and "parton shower" is that these terms convolute calculational definitions with algorithmic ones, and the goal of the GenEvA framework is to isolate the calculational from the algorithmic challenges involved in combining different descriptions of QCD.

As shown in figure 1, a fixed-order calculation is based on a perturbative expansion of QCD to fixed order in α_s. To distribute events according to a fixed-order calculation, one usually uses a non-Markovian algorithm (such as an adaptive grid) to generate points in fixed-multiplicity n-body phase space $\mathrm{d}\Phi_n$. The parton shower, on the other hand, is defined in the soft-collinear limit of QCD, and the splitting functions and Sudakov factors are calculated in this limit. Parton showers are usually used in the context of a Markovian phase space algorithm that recursively generates all of variable-multiplicity phase space through a probabilistic mapping of $\mathrm{d}\Phi_n \to \mathrm{d}\Phi_{n+1}$.[3] Therefore, in trying to combine a fixed-order calculation with a parton shower, one is simultaneously trying to merge two

[3]More general parton showers such as in refs. [34, 35] can also have non-Markovian aspects.

JHEP12(2008)010

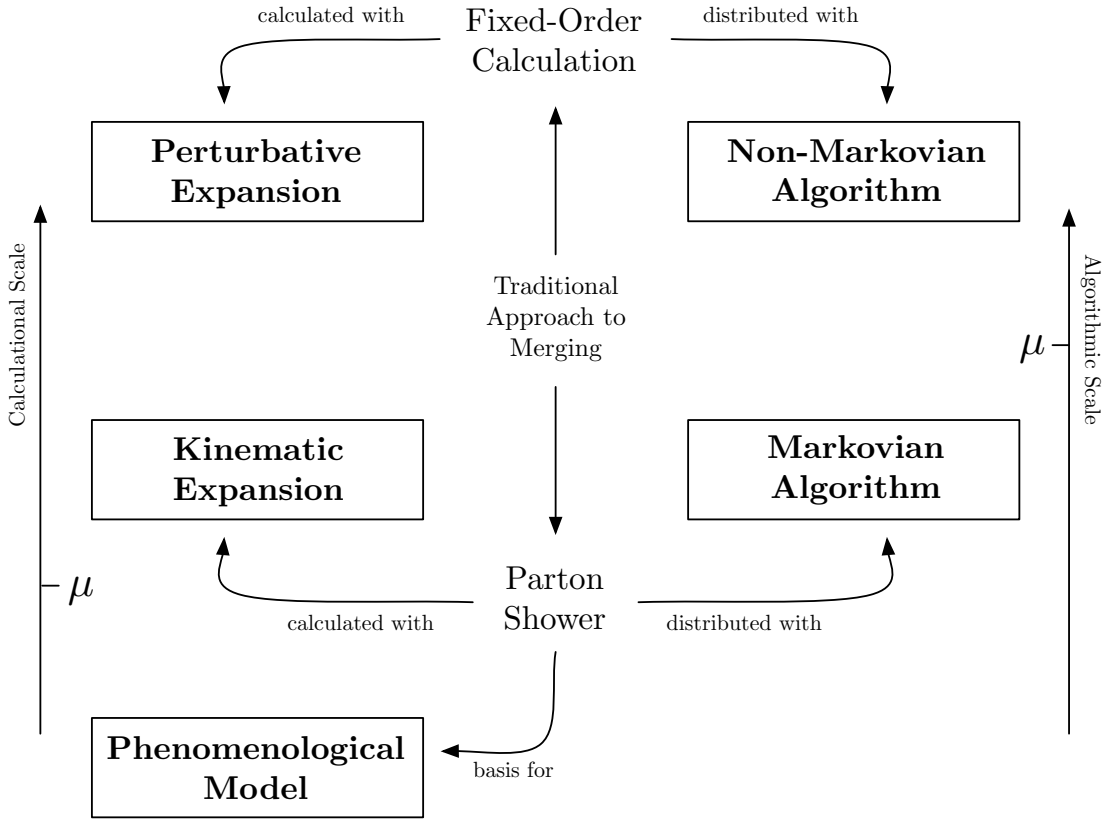


Figure 1: The traditional approach to merging fixed-order calculations with parton showers. Traditionally, fixed-order calculations tie a perturbative expansion to a non-Markovian fixed-multiplicity phase space algorithm. The parton shower ties a kinematic expansion to a Markovian variable-multiplicity phase space algorithm. Hence, in trying to merge fixed-order calculations with parton showers, one is led to simultaneously having to merge two different QCD expansions and two different algorithmic methods to generate phase space. On the algorithmic side, imposing a phase space cut μ to separate the two different algorithms would results in residual μ dependence and uncanceled infrared divergences as explained in the text. `GenEvA` is based on imposing a calculational μ cut that separates QCD calculations from QCD phenomenological models. This requires splitting the parton shower into a formal component and a phenomenological component.

different expansions of QCD and two different phase space algorithms. Indeed, currently there does not exist a solution for how to merge *generic* fixed-order calculations with parton showers, mainly because separate algorithmic merging procedures are currently necessary depending on whether one is considering tree-level calculations or one-loop calculations.

Naively, one might hope to define an algorithmic phase space cut μ to separate fixed-order calculations from parton showers as on the right side of figure 1. One could imagine using a fixed-order calculation when the invariant mass between two partons is greater than μ and using a parton shower when the invariant mass is less than μ. This *algorithmic* μ would immediately solve the problem of double-counting that arises when the parton shower acting on n-body phase space covers exactly the same phase space regions as an

JHEP12(2008)010

$(m > n)$-body generator. However, this approach is suspect for two reasons. First, the parton shower resums the leading Sudakov logarithms [64] in the problem, so if the parton shower is only used below μ, it will exhibit double-logarithmic $\alpha_s \log^2 \mu$ sensitivity to this arbitrary, unphysical μ scale. Second, infrared divergences in virtual diagrams contributing to the fixed-order results are canceled by collinear and/or soft real-emission diagrams, which are by definition contributing below the scale μ and are therefore contained in the parton shower picture. Thus, one has to somehow link the fixed-order calculation to the parton shower in order to properly cancel infrared divergences.

We argue that a better definition of μ is as a *calculational* scale that separates calculations performed in QCD from phenomenological models based on QCD as on the left side of figure 1. This requires cleanly separating two different uses of the parton shower. Above the scale μ, the parton shower corresponds directly to quantities that can be defined formally in the soft-collinear limit. In particular, the splitting functions [65 – 67] and Sudakov factors [64] can be systematically derived [61, 62] using soft-collinear effective theory (SCET) [57 – 60]. Below the scale μ, the parton shower should be regarded as simply a QCD-inspired phenomenological model. For example, a phenomenological shower can be used to extrapolate into regions of phase space away from the soft-collinear limit if high-multiplicity fixed-order matrix elements are unavailable, creating partonic final states that include the correct singularity and symmetry structure of QCD, but lack the full quantum interference. Similarly, a phenomenological shower can serve as the entry point to nonperturbative hadronization in a general fragmentation scheme, and in that context, the shower can and should be tuned to data to reproduce the measured fragmentation properties seen in experiments.

It is the fact that the parton shower has meaning above and below μ that can be used to mitigate the unphysical μ dependence, since it implies that the double-logarithmic dependence of the parton shower is identical to the double-logarithmic dependence of QCD. This guarantees that partonic calculations with leading-logarithmic improvements will have no leading μ dependence when interfaced with a parton shower at the scale μ. However, the challenge of defining μ in this way is that now it is not clear whether the issue of phase space double-counting will be solved, nor is it obvious that such a definition of μ will provide a method to combine generic fixed-order calculations with a parton shower, especially in the presence of infrared divergences. At this point, it is thus worthwhile to reconsider the various components of figure 1 to distill the essential challenges in combining fixed-order calculations with parton showers.

2.2 Combining different QCD descriptions

In trying to combine different descriptions of QCD into a coherent Monte Carlo framework, there are really three definitions of what is meant by "combine" as shown in figure 2: combining different formal expansions, combining formal calculations with phenomenological models, and combining different phase space algorithms. These definitions of "combine" are in one-to-one correspondence with three technical problems that arise when trying to merge fixed-order calculations with parton showers: regulating infrared divergences, cancel-

JHEP12(2008)010

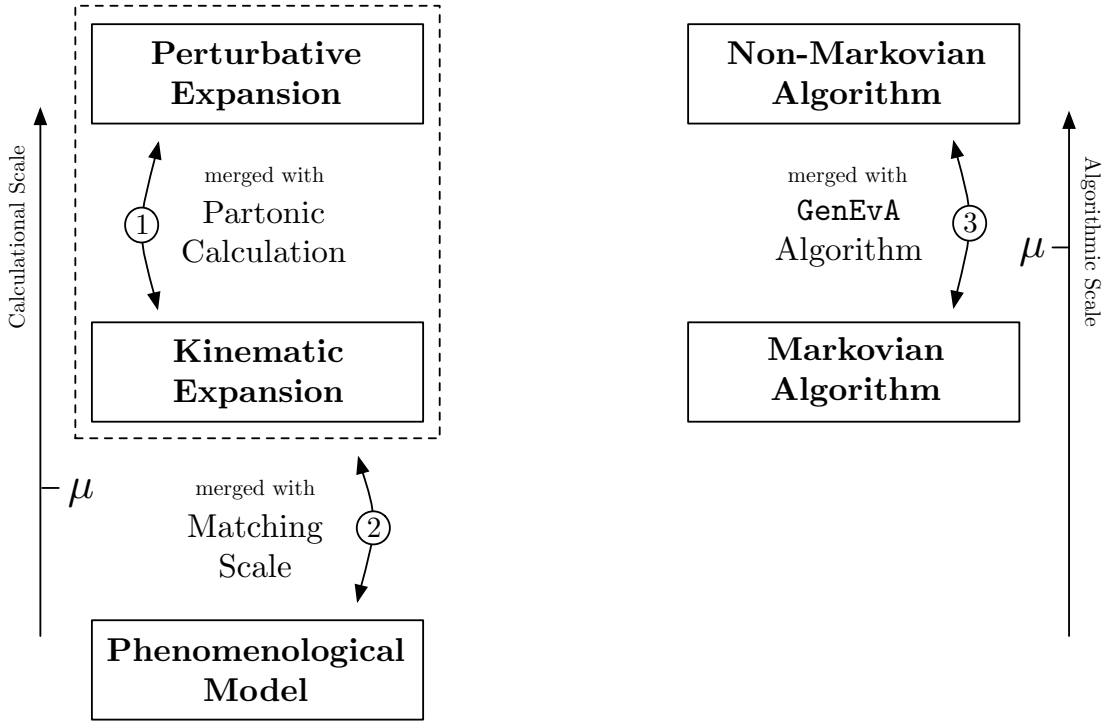


Figure 2: The three different meanings of "combine": ① Combining different expansions of QCD in a partonic calculation; ② combining a partonic calculation with a phenomenological model; and ③ combining a non-Markovian phase space generator with a Markovian generator. By isolating the issues involved in merging fixed-order calculations with parton showers, `GenEvA` offers a simple and generic method to improve Monte Carlo. Note that all these meanings of "combine" respect the division between calculations on one side and algorithms on the other, and the `GenEvA` framework is devised to cleanly separate the calculational and algorithmic issues.

ing unphysical μ dependence, and eliminating double-counting. The power and simplicity of the `GenEvA` framework comes from cleanly separating these three issues.

The first meaning of "combine" is to merge different formal expansions of QCD, in our case a perturbative expansion in α_s, with a kinematic expansion that allows resumming logarithms of some ratio r of energy scales. A successful combination of this form will result in a partonic calculation $|\mathcal{M}(\mu)|^2$ that is simultaneously correct to the calculated order in α_s while also including the desired level of logarithmic resummation. While there are formal procedures to calculate $|\mathcal{M}(\mu)|^2$ by matching QCD onto SCET [62, 61], there are numerous correct choices for $|\mathcal{M}(\mu)|^2$ that satisfy the desired properties at the leading-logarithmic level. We will also see that the issue of infrared divergences is solved entirely in the context of a partonic calculation, because one can think of the cancellation of divergences between loop diagrams and tree diagrams as just another type of formal combination.

We have included an explicit μ dependence in the partonic calculation $|\mathcal{M}(\mu)|^2$ in anticipation of the second meaning of "combine", which is to combine a formal QCD calculation with a phenomenological model based on QCD. Below some scale μ, the phenomenological

JHEP12(2008)010

parton shower will be used to fill out phase space and interface with a nonperturbative hadronization model, but we already saw that this will introduce unphysical $\alpha_s \log^2 \mu$ dependence in the final results. However, if we include leading-logarithmic resummation in the partonic calculation $|\mathcal{M}(\mu)|^2$, then this μ dependence will cancel by construction. At first glance, it appears that the partonic calculation now needs to know the details of the phenomenological shower in order to engineer the $\alpha_s \log^2 \mu$ cancellation. Remember though, that this double-logarithmic μ dependence is a property of QCD and not of a specific parton shower, so the correct μ dependence can be included in $|\mathcal{M}(\mu)|^2$ independently of the details of the shower. The only requirement is that the same formal definition of μ is used in both the partonic calculation and the phenomenological shower. In practice, the choice of μ corresponds to the choice of shower evolution variable.

Finally, the third meaning of "combine" has to do with merging different algorithmic techniques. We saw that phase space can either be covered by a fixed-multiplicity non-Markovian algorithm or with a variable-multiplicity Markovian algorithm, and for computational reasons we may wish to use different algorithms in different regions of phase space. Because of the problem of double-counting, one needs to merge these two algorithms in such a way that a given point in phase space is covered once and only once. One option is to define a phase space cut μ, as mentioned already, and use a non-Markovian algorithm above that scale μ and a Markovian algorithm below μ. We denote such phase space with a matching scale as dMC(μ), and an example of an efficient dMC(μ) generator is presented in the companion paper [1].

Let us return to the traditional approach in eq. (1.1) for creating a fully differential hadronic cross section, now with the final state multiplicity n made manifest:

$$\mathrm{d}\sigma^{\text{trad.}} = \sum_n \mathrm{MC}_{\mu_n} \left(|\mathcal{M}_n|^2 \, \mathrm{d}\Phi_n \right) . \tag{2.1}$$

Here, $|\mathcal{M}_n|^2$ is a fixed-order calculation with n final state partons, $\mathrm{d}\Phi_n$ is n-body phase space, and MC_{μ_n} represents a phenomenological parton shower that starts at the scale μ_n. As mentioned already, current approaches to merging fixed-order calculations with parton showers require either a modification to $|\mathcal{M}_n|^2$, to MC_{μ_n}, or to both, and `GenEvA` is no exception. The novelty of `GenEvA` is that these modifications can be made completely generic by applying the logic of figure 2.

First, a fixed-order calculation lacks the correct behavior for small r, so we have to substitute a partonic calculation that includes a proper logarithmic resummation,

$$\textcircled{1} \qquad |\mathcal{M}_n|^2 \to |\mathcal{M}_n(\tilde{\mu}_n)|^2 ,$$

where at this point $\tilde{\mu}_n$ is an unspecified infrared scale. Second, to avoid large $\alpha_s \log^2 \mu_n$ dependence when the phenomenological parton shower is applied, the infrared scale in the partonic calculation must be the same as the starting scale of the shower,

$$\textcircled{2} \qquad \tilde{\mu}_n \to \mu_n .$$

Finally, the traditional approach to phase space involves two different phase space algorithms, a non-Markovian $\mathrm{d}\Phi_n$ generator and a Markovian MC_{μ_n} generator. For any specific

JHEP12(2008)010

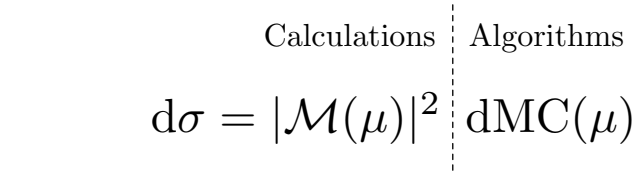


Figure 3: A schematic summary of the GenEvA master formula. The fully exclusive differential cross section comes from interfacing a partonic calculation $|\mathcal{M}(\mu)|^2$ with a phenomenological model through the use of Monte Carlo space $\mathrm{dMC}(\mu)$. The matching scale μ allows this interface to be smooth and defines a clean separation between calculational and algorithmic issues.

phase space point, the partonic calculation is simply a number that is unchanged by the action of MC_{μ_n}, so $|\mathcal{M}_n(\mu_n)|^2$ can be factored out separately for each n. The quantity $\sum_n \mathrm{MC}_{\mu_n}(\mathrm{d}\Phi_n)$ involves multiple covering of phase space, since the action of the phenomenological model starting from a phase space point in $\mathrm{d}\Phi_n$ will populate regions of phase space $\mathrm{d}\Phi_{m>n}$ by the splitting of particles. To solve the problem of double-counting we introduce a μ-aware phase space generator

$$\text{③} \qquad \mathrm{MC}_{\mu_n}(\mathrm{d}\Phi_n) \to \mathrm{dMC}_n(\mu_n)\,,$$

which cleanly separates phase space populated by the $\mathrm{d}\Phi_n$ generator and the subsequent evolution using the phenomenological model. These three steps lead to the GenEvA master formula from eq. (1.2)

$$\mathrm{d}\sigma^{\texttt{GenEvA}} = \sum_n |\mathcal{M}_n(\mu_n)|^2 \mathrm{dMC}_n(\mu_n)\,. \tag{2.2}$$

There are two interesting features of eq. (2.2). First, as summarized in figure 3, this formula can be interpreted as an explicit separation of calculational issues from algorithmic issues. $|\mathcal{M}_n(\mu_n)|^2$ certainly encodes the challenges of merging together different QCD expansion schemes into a single partonic cross section. $\mathrm{dMC}_n(\mu_n)$ is slightly more complicated because it encodes both algorithmic information as well as the physics of phenomenological models. However, to the extent to which we (unfairly) regard phenomenological models as QCD-inspired numerical algorithms, $\mathrm{dMC}_n(\mu_n)$ encodes the algorithmic challenges of distributing QCD-like events. Because calculational issues are now separated from algorithmic ones, the partonic calculations $|\mathcal{M}_n(\mu_n)|^2$ can be determined through physics consideration alone, independently of the details of $\mathrm{dMC}_n(\mu_n)$.

Second, as summarized in figure 4, the same scale μ_n appears both in $|\mathcal{M}_n(\mu_n)|^2$ and in $\mathrm{dMC}_n(\mu_n)$. A potential mismatch as in figure 1 between the calculational scale μ and the algorithmic scale μ is avoided because these two scales are now forced to be the same to cancel the $\alpha_s \log^2 \mu$ dependence. In this way, all the user needs to know is that full QCD information as encoded in $|\mathcal{M}(\mu)|^2$ will be used above μ, and phenomenological models of QCD (which the user need not specify) will be used below μ. Of course, the precise definition of μ must always be specified, and in this paper, we will use the invariant mass between particles to define μ, leaving a discussion of generalizations for the companion paper [1].

JHEP12(2008)010

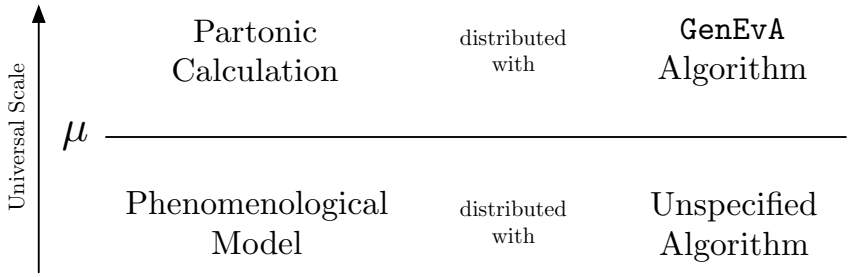


Figure 4: The user's view of the GenEvA framework. Partonic calculations as defined by $|\mathcal{M}(\mu)|^2$ are used above a universal scale μ and can be distributed with the GenEvA algorithm [1]. A phenomenological model is used below the scale μ and is implemented using an unspecified algorithm, usually a parton shower interfaced with a hadronization scheme. Because the same scale μ defines both the calculational separation and the algorithmic separation, the user is free to choose the best partonic calculations from physics considerations alone, without having to worry about the details of the algorithmic implementation.

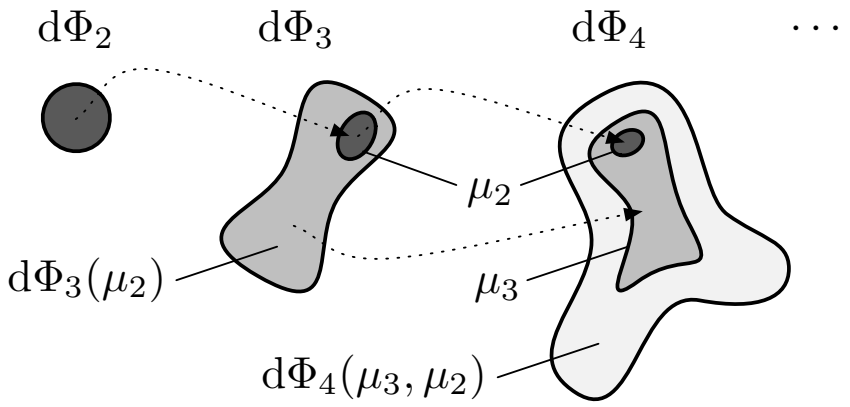


Figure 5: The scale-dependent phase space $\mathrm{d}\Phi_n(\{\mu_i\}_{i<n})$. The regions of $\mathrm{d}\Phi_{3,4}$ labeled μ_2 are those reached by the phenomenological model acting on $\mathrm{d}\Phi_2$ with starting scale μ_2. The region of $\mathrm{d}\Phi_4$ labeled μ_3 is reached by the phenomenological model acting on $\mathrm{d}\Phi_3$ when started at the scale μ_3. Excluding these regions from $\mathrm{d}\Phi_{3,4}$ defines $\mathrm{d}\Phi_3(\mu_2)$ and $\mathrm{d}\Phi_4(\mu_2, \mu_3)$. When interfaced with a phenomenological model, the set of all $\mathrm{d}\Phi_n(\{\mu_i\}_{i<n})$ regions gives a complete covering of phase space with no double-counting by construction.

In the remainder of this section, we will consider the various elements in eq. (2.2) in more detail, starting with a more precise definition of the phase space $\mathrm{dMC}(\mu)$, explaining how to include additional flexibility in defining more scales like μ, and then discussing methods to determine the partonic calculation $|\mathcal{M}(\mu)|^2$.

2.3 A new approach to phase space

As discussed in the previous section, the GenEvA framework requires a phase space generator $\mathrm{dMC}(\mu)$ that, by construction, does not introduce double-counting when interfaced

JHEP12(2008)010

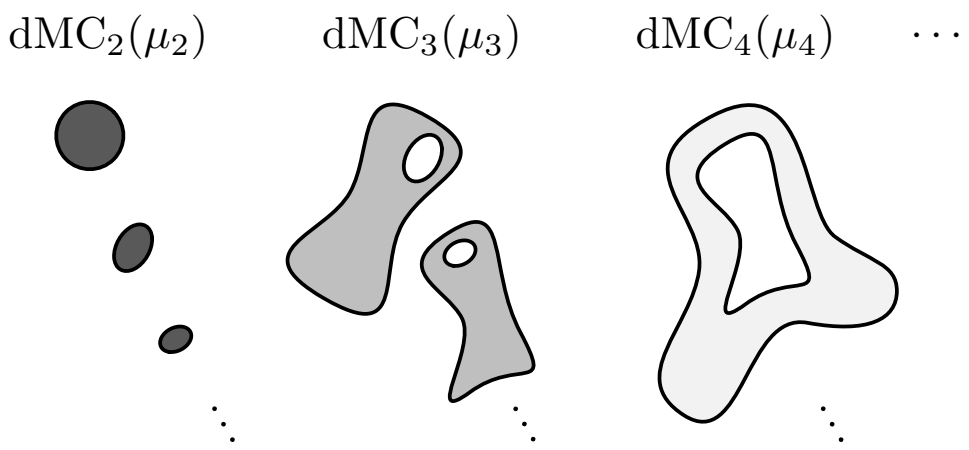


Figure 6: Definition of "Monte Carlo space" $\mathrm{dMC}_n(\mu_n)$ as all regions of scale-dependent phase space $\mathrm{d}\Phi_n(\{\mu_i\}_{i<n})$ that can be reached by the phenomenological model starting from μ_n. $\mathrm{dMC}_2(\mu_2)$ consists of $\mathrm{d}\Phi_2$ together with the part of $\mathrm{d}\Phi_{3,4,\ldots}$ labeled by μ_2 in figure 5, $\mathrm{dMC}_3(\mu_3)$ consists of $\mathrm{d}\Phi_3(\mu_2)$ together with the part of $\mathrm{d}\Phi_{4,\ldots}$ labeled by μ_3 (but excluding that labeled by μ_2), and so on. This way of organizing phase space emphasizes that while partonic calculations are defined on scale-dependent phase space $\mathrm{d}\Phi(\mu)$, they affect all regions of Monte Carlo space $\mathrm{dMC}(\mu)$ through the phenomenological model.

with a phenomenological model for filling out phase space. To be specific, we take the phenomenological model to be a parton shower that takes an n-body configuration Φ_n in $\mathrm{d}\Phi_n$ and starts showering at some scale μ_n to produce additional partons. In other words, the parton shower produces m-body configurations with $m \geq n$, where the precise regions of higher-dimensional phase space that are populated are determined by the choice of starting scale μ_n and the details of the parton shower. Calling these regions $\mathrm{MC}_n(\mu_n)$, we can think of the parton shower as defining a map

$$\Phi_n \overset{\text{pheno. model}}{\longrightarrow} \mathrm{MC}_n(\mu_n)\,. \tag{2.3}$$

To eliminate double-counting, we simply have to exclude those regions from $\mathrm{d}\Phi_n$ that can be reached by acting the shower on $\mathrm{d}\Phi_i$ starting from the scale μ_i for any $i < n$. As illustrated in figure 5, the remaining parts of $\mathrm{d}\Phi_n$ define phase space with a matching scale $\mathrm{d}\Phi_n(\{\mu_i\}_{i<n})$.[4] Note that $\mathrm{d}\Phi_n(\{\mu_i\}_{i<n})$ is a function of all the lower-dimensional matching scales $\{\mu_i\}_{i<n}$ and is independent of the matching scale μ_n.

Having solved the issue of double-counting, it is now more convenient to directly talk about "Monte Carlo space" $\mathrm{dMC}_n(\{\mu_i\}_{i\leq n})$, which is defined as $\mathrm{d}\Phi_n(\{\mu_i\}_{i<n})$ plus the collection of all relevant regions excluded from $\mathrm{d}\Phi_{m>n}$, as illustrated in figure 6. More precisely, $\mathrm{dMC}_n(\{\mu_i\}_{i\leq n})$ is the image of $\mathrm{d}\Phi_n(\{\mu_i\}_{i<n})$ under the parton shower map in eq. (2.3). To simplify our notation, we will mostly suppress the implicit dependence on the scales $\mu_{i<n}$ and only write $\mathrm{dMC}_n(\mu_n)$.

[4] In figure 5, $\mathrm{d}\Phi_4(\mu_2, \mu_3)$ strictly speaking only depends on μ_3 since the region of $\mathrm{d}\Phi_4$ mapped out by μ_2 lies completely inside μ_3. In general, this need not be the case, see for example figure 11.

JHEP12(2008)010

By construction, a covering of phase space with no double-counting is given by the map

$$\sum_{n=2}^{n_{\max}} \mathrm{d}\Phi_n(\{\mu_i\}_{i<n}) \overset{\text{pheno. model}}{\longrightarrow} \sum_{n=2}^{n_{\max}} \mathrm{dMC}_n(\mu_n)\,, \tag{2.4}$$

where $n_{\max}$ will be determined by the maximal available number of external particles in the partonic calculation. As long as the scales μ_n for $n \leq n_{\max}$ and the parton shower satisfy certain mild constraints to guarantee no dead zones, eq. (2.4) also provides a complete covering of phase space which is one-to-one and onto, i.e. every region of phase space is covered exactly once:

$$\sum_{n=2}^{n_{\max}} \mathrm{dMC}_n(\mu_n) \to \sum_{n=2}^{\infty} \mathrm{d}\Phi_n\,. \tag{2.5}$$

Making the dependence on $n_{\max}$ explicit, the `GenEvA` master formula is

$$\mathrm{d}\sigma = \sum_{n=2}^{n_{\max}} |\mathcal{M}_n(\mu_n)|^2\, \mathrm{dMC}_n(\mu_n)\,. \tag{2.6}$$

A crucial point is that an n-body partonic calculation $|\mathcal{M}_n(\mu_n)|^2$ is defined just as a function of Φ_n, but the extrapolation to all of $\mathrm{dMC}_n(\mu_n)$ is provided by the phenomenological model. In principle, all eq. (2.6) requires is a phase space generator that can generate scale-dependent phase space $\mathrm{d}\Phi(\mu)$. Such a generator could be built out of existing Monte Carlo tools by including a phase space veto. Because $\mathrm{d}\Phi(\mu)$ implicitly depends on the precise choice of phenomenological model, doing so might be nontrivial, but there is no conceptual difficulty. Once such a generator exists, we can think entirely in the language of partonic calculations, but still generate fully hadronized events with the aid of QCD phenomenological models.

2.4 Variable resolution scale

When discussing strategies for calculating the partonic result $|\mathcal{M}(\mu)|^2$, it will prove very useful to define a projection map from n-body phase space to m-body phase space

$$\Phi_n \to \Phi_m \qquad \text{with} \qquad m < n\,. \tag{2.7}$$

This phase space projection map will be the key to creating our NLO/LO/LL merged sample. Such a map can be achieved by defining a resolution scale $\bar{\mu}$, which would resolve more and more partons as the scale is lowered. Note that this scale does not have to coincide with the scale μ we use to divide up phase space.

As we raise the scale $\bar{\mu}$, the projection map will cluster together two partons once their "distance" becomes smaller than $\bar{\mu}$. The phase space map given in eq. (2.7) will give $\Phi_n \to \Phi_m$ at the resolution scale $\bar{\mu}_m$. As an example, consider the event shown in figure 7, where at the matching scale μ there are 7 resolved partons. At the scale $\bar{\mu}_5$, two pairs of particles have been clustered together such that the event has 5 resolved partons. Similarly, at the scales $\bar{\mu}_3$ and $\bar{\mu}_2 = E_{\mathrm{CM}}$, we have 3 and 2 resolved partons, respectively.

JHEP12(2008)010

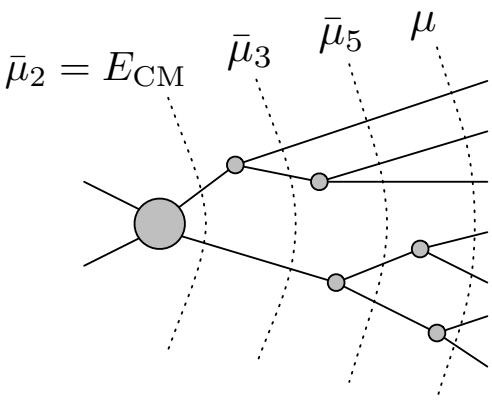


Figure 7: Phase space with a variable resolution scale. The shaded blobs represent one possible phase space projection map as in eq. (2.7) based on successive $2 \to 1$ recombinations. At the matching scale μ, the event has 7 final state partons. Going from the matching scale μ to the resolution scale $\bar{\mu}_5$ maps a 7-body configuration to a 5-body configuration. Further increasing the scale to $\bar{\mu}_3$ and $\bar{\mu}_2 = E_{\mathrm{CM}}$ yields a 3-body and eventually a 2-body configuration. The ability to project a given point of phase space to a lower-dimensional phase spaces allows for partonic calculations $|\mathcal{M}(\mu, \bar{\mu}, \ldots)|^2$ that depend on multiple scales $\{\mu\}$.

There are several such maps available in the literature including the k_T algorithm [68, 69] and the momentum mappings often found in dipole showers [70, 24]. In fact, any recursive cluster-based jet algorithm is an example of such a map, but to be useful, eq. (2.7) has to respect the symmetry and singularity structure of QCD, which means that the clustering procedure should respect flavor, and the scale $\bar{\mu}$ should roughly determine the "soft-collinear distance". Such maps give a sequence of resolution scales, and in the past, it has been mainly the ability to obtain such a sequence of scales which has been used in the literature. An example is the CKKW procedure [23], which uses a flavor-aware variant of the k_T algorithm to identify the scales needed to implement leading-logarithmic improvements to tree-level calculations. In `GenEvA`, the actual four-momenta determined by eq. (2.7) are also important, and we crucially assume that the projection eq. (2.7) returns a set of *on-shell* four-momenta, such that partonic calculations can be defined in terms of

$$|\mathcal{M}_n(\Phi_n, \mu_n; \Phi_m, \bar{\mu}_m; \ldots)|^2 \,. \tag{2.8}$$

In the `GenEvA` algorithm [1], we will show that we can use the evolution variable of the parton shower itself as the resolution variable. Thus, the projection in eq. (2.7) will effectively be the inverse of the parton shower map in eq. (2.3).[5] This reversal property of the projection means that we have a consistent way to not just increase but also *decrease* the matching scale at will. This could potentially be very important because it allows one to *raise* the value of $n_{\max}$ in eq. (2.6) and thus use additional partonic calculations after the events have been generated or even passed through a detector simulation. In fact, a

[5]Strictly speaking, eq. (2.7) will be a pseudo-left-inverse of eq. (2.3), meaning that running the shower and applying the projection will yield the identity map up to discrete ambiguities.

JHEP12(2008)010

reversable projection can be used to *define* $\mathrm{dMC}(\mu)$, as any given phase space integration region $\mathrm{d}\Phi_n$ can be mapped to the proper $\mathrm{dMC}_m(\mu_m)$ region as needed in figure 6. The presence of a reversable phase space projection is one of the reasons why the `GenEvA` algorithm is efficient and versatile, but this reversibility property is not strictly needed for the `GenEvA` framework discussed here.

2.5 Strategy for partonic calculations

In the language of eq. (1.2), the strategy to systematically improve Monte Carlo is simply to eliminate phenomenological models by pushing the scale μ as low as possible (ideally to Λ_{QCD}) with improved QCD partonic calculations $|\mathcal{M}(\mu)|^2$. Of course, one would also like to refine and tune phenomenological models based on experimental and theoretical input, but that is not the focus of the present work, where we are concerned primarily with improving the perturbative description of QCD.

How does one determine a partonic calculation $|\mathcal{M}(\mu)|^2$? As discussed, the choice can be made by physics considerations alone without worrying about algorithmic implications,[6] and the physics goal is to determine an expression for $|\mathcal{M}(\mu)|^2$ that is formally correct both in a perturbative expansion in α_s and in a kinematic expansion about r, where r denotes the ratio of two kinematic scales, such as the invariant mass between two partons compared to the center-of-mass energy. The details of how to define partonic calculations are discussed in subsequent sections, and we give just a schematic overview here.

As is well known, an expansion to fixed order in α_s does not yield a good description of perturbative QCD over all of phase space. The presence of double-logarithmic terms $(\alpha_s \log^2 r)^n$ at each order in the perturbative expansion invalidates the fixed-order expansion for small values of r. These double-logarithmic terms can be resumed to all orders in perturbation theory by doing a soft-collinear expansion of QCD. In this paper we will only work to leading-logarithmic order, leaving a treatment of subleading logarithms for future work.

The most naive partonic calculation is one that completely avoids the issue of logarithmic resummation and only uses tree-level (LO) matrix elements

$$|\mathcal{M}_n^{\mathrm{LO}}(\mu)|^2 \simeq |\mathcal{M}_n^{\mathrm{tree}}|^2 \,. \tag{2.9}$$

Note that the infrared divergences in the tree-level diagrams are regulated because phase space has a matching scale μ that imposes a phase space restriction.

Because of the double-logarithmic sensitivity to the scale μ from running the phenomenological model, we want to supplement $|\mathcal{M}^{\mathrm{LO}}(\mu)|^2$ with the correct leading-logarithmic (LL) μ dependence. As observed by CKKW [23], this can be accomplished by multiplying the tree-level diagram with a μ-dependent Sudakov factor Δ,

$$|\mathcal{M}_n^{\mathrm{LO/LL}}(\mu)|^2 \simeq |\mathcal{M}_n^{\mathrm{tree}}|^2 \, \Delta(E_{\mathrm{CM}}, \mu) \,. \tag{2.10}$$

[6] In practice, the availability of fast numerical methods to evaluate $|\mathcal{M}(\mu)|^2$ might affect the precise form of the partonic calculation, especially in the presence of loop expressions.

JHEP12(2008)010

Because the Sudakov factor can be formally expanded as $\Delta = 1 + \mathcal{O}(\alpha_s)$, this LO/LL partonic cross section is simultaneously correct to leading order in an α_s expansion and to leading-logarithmic order in the soft-collinear limit.

The story becomes more complicated at next-to-leading order (NLO) because of the need to cancel infrared divergences between n-body virtual and $(n+1)$-body real-emission diagrams. Though in practice, one typically uses Catani-Seymour subtractions [71, 72] to calculate NLO observables, for the moment, we will use a slicing method [73 – 78], which is conceptually simpler. Schematically, one writes

$$|\mathcal{M}_n^{\text{NLO}}(\mu)|^2 \simeq |\mathcal{M}_n^{\text{real}}|^2 + |\mathcal{M}_n^{\text{virtual}}|^2 + \int_\mu |\mathcal{M}_{n+1}^{\text{real}}|^2 \,, \tag{2.11}$$

where the integral $\int_\mu$ represents slicing part of the $(n+1)$-body real-emission phase space to cancel the infrared divergences in the n-body virtual diagrams. Any divergences in the n-body real emission are regulated by the matching scale μ as in the tree-level case, but the effect of those divergences can be used again in a slicing scheme to cancel infrared divergences in the $(n-1)$-body virtual diagrams.

Adding leading-logarithmic information to the NLO calculation is much more involved. A naive approach is simply incorrect,

$$|\mathcal{M}_n^{\text{NLO/LL}}(\mu)|^2 \neq |\mathcal{M}_n^{\text{NLO}}(\mu)|^2 \, \Delta(E_{\text{CM}}, \mu) \,, \tag{2.12}$$

because the $\mathcal{O}(\alpha_s)$ pieces in the Sudakov factor change the $\mathcal{O}(\alpha_s)$ behavior of the fixed-order calculation. An achievement of MC@NLO [44] was to figure out an expression for $|\mathcal{M}^{\text{NLO/LL}}(\mu)|^2$ that simultaneously has NLO and LL accuracy. In ref. [44], a specific algorithmic method to implement $|\mathcal{M}^{\text{NLO/LL}}(\mu)|^2$ was used that generated events with sometimes negative weights, but in the context of `GenEvA` the same $|\mathcal{M}^{\text{NLO/LL}}(\mu)|^2$ expression can be used to generate NLO/LL accurate events with manifestly positive weights.

At this point, `GenEvA` is simply reproducing known results from PS/ME merging (LO/LL) and PS/NLO merging (NLO/LL) in the language of partonic calculations $|\mathcal{M}(\mu)|^2$. The advantage of this language is that it makes it straightforward to merge these two results into a partonic calculation that has combined NLO/LO/LL accuracy. As already mentioned in section 2.4, and as we will explain in more detail in section 3.6, one can define a scale $\bar{\mu}_m > \mu$ at which only $m < n$ partons are resolved. This can be achieved using the phase space projection map of eq. (2.7). If one has LO/LL information for n partons, and additional NLO/LL information for m partons,[7] one can supplement the LO/LL partonic cross section with this higher-order information, using

$$|\mathcal{M}_n^{\text{NLO/LO/LL}}(\mu)|^2 = |\mathcal{M}_n^{\text{LO/LL}}(\mu)|^2 \, \frac{|\mathcal{M}_m^{\text{NLO/LL}}(\bar{\mu}_m)|^2}{|\mathcal{M}_m^{\text{LO/LL}}(\bar{\mu}_m)|^2} \,. \tag{2.13}$$

This result describes NLO observables correct to NLO, LO observables correct to LO, all with the correct LL behavior.

[7] Here, NLO/LL information for m partons means that one has performed an $(m-1)$-body one-loop calculation.

JHEP12(2008)010

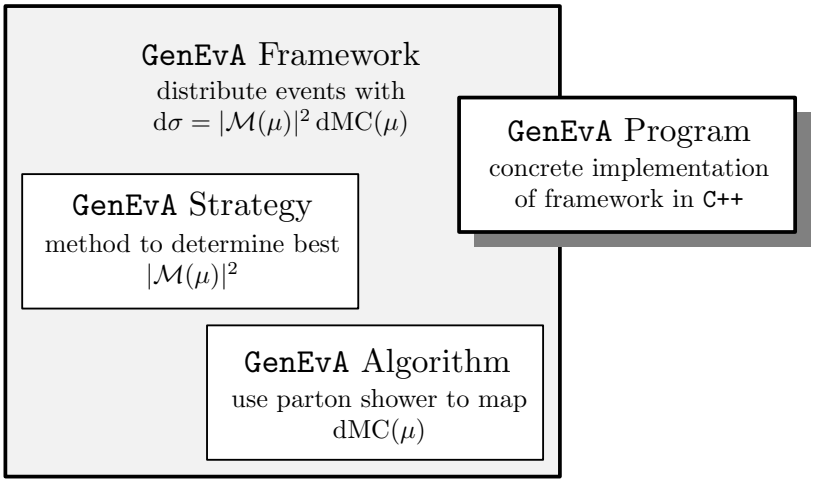


Figure 8: The four different meanings of "GenEvA". The GenEvA *framework* is an umbrella for the idea of distributing partonic calculations with a matching scale, $|\mathcal{M}(\mu)|^2$, across phase space with a matching scale, $\mathrm{dMC}(\mu)$. The GenEvA *strategy* asserts that different QCD expansions can be merged into a single partonic quantity $|\mathcal{M}(\mu)|^2$. The GenEvA *algorithm* is a method for using a parton shower as a phase space generator. The GenEvA *program* is a concrete implementation of the GenEvA framework employing the GenEvA algorithm.

Though we do not demonstrate an NNLO/NLO/LO/NLL/LL merged sample in the present work, whether or not such a sample can be built in principle only depends on whether it is possible to derive an expression for $|\mathcal{M}(\mu)|^2$ that simultaneously allows $\mathrm{N}^i\mathrm{LO}$ observables to be correct to $\mathrm{N}^i\mathrm{LO}$ and $\mathrm{N}^j\mathrm{LL}$ observables to be correct to $\mathrm{N}^j\mathrm{LL}$. Of course, since we will interface with a phenomenological model that might not have the correct NLL behavior, we now have to push μ as low as possible to make sure that our $\mathrm{N}^i\mathrm{LO}/\mathrm{N}^j\mathrm{LL}$ calculation describes as much of phase space as possible. Using the language of partonic calculations, we see that how far Monte Carlo can be improved on the perturbative side is in principle a pencil-and-paper question that can be separated from algorithmic issues.

2.6 Summary of GenEvA

Just as "fixed-order calculation" and "parton shower" referred to various concepts, GenEvA itself refers to four different concepts, summarized in figure 8. GenEvA is:

1. A Monte Carlo *framework* based on the idea of generating events according to

$$d\sigma = |\mathcal{M}(\mu)|^2 \, dMC(\mu) \tag{2.14}$$

   by distributing a generic partonic calculation with a matching scale, $|\mathcal{M}(\mu)|^2$, using a generic phase space generator with a matching scale, $\mathrm{dMC}(\mu)$. This framework could be implemented using traditional Monte Carlo tools without ever referencing the GenEvA phase space algorithm.

2. A *strategy* for improving Monte Carlo based on merging different QCD expansions to determine the best partonic calculations $|\mathcal{M}(\mu)|^2$. This strategy will be the subject of

JHEP12(2008)010

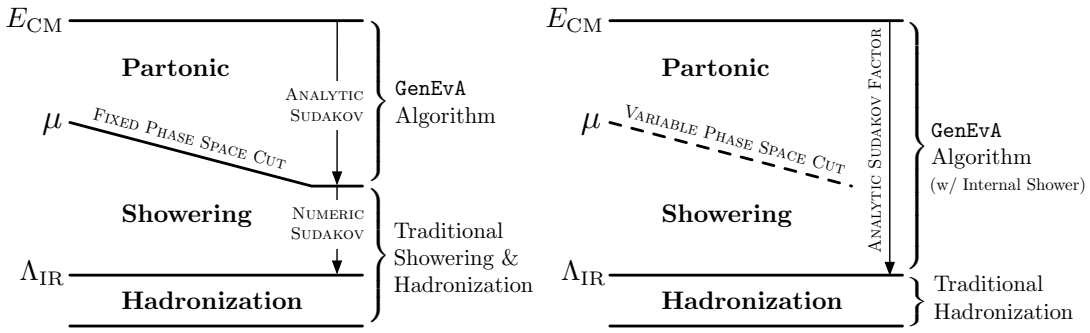


Figure 9: The three regimes of event generation. Left panel: The GenEvA algorithm [1] could be used only in the partonic regime, with the showering regime covered by a traditional, already tuned, showering program. Right panel: Alternatively, both partonic and showering regimes can be covered by the GenEvA algorithm using its internal parton shower. In this case the matching scale between the partonic and showering regimes remains freely adjustable through the reversable phase space projection. While these two approaches differ in how Sudakov factors are calculated, they both give results that are accurate to leading-logarithmic order.

the remaining sections and yields a formula for an NLO/LO/LL merged calculation and should be generalizable to NNLO or NLL.

3. An *algorithm* for generating phase space with a variable matching scale, dMC(μ), by analytically reweighting a parton shower to fixed-multiplicity phase space. This algorithm is detailed in ref. [1].

4. A computer *program* that gives a concrete implementation of the GenEvA framework by using the GenEvA algorithm. An alpha version of this software is available from the authors upon request, and is used here to show the results of the NLO/LO/LL calculation.

We emphasize that while the GenEvA algorithm for generating phase space is quite novel, the important physics behind the GenEvA framework is captured not by the phase space integration dMC(μ) but by the GenEvA strategy to define the partonic calculation $|\mathcal{M}(\mu)|^2$. Though we need a phase space generator with a variable matching scale in order to create an NLO/LO/LL merged sample, the physics behind this merging comes from finding a suitable form of $|\mathcal{M}(\mu)|^2$.

To summarize the complete GenEvA approach to Monte Carlo, it is instructive to separate the phenomenological model into a parton shower component and a hadronization component. If we now divide event generation into three regimes as in figure 9—partonic, showering, and hadronization — then the GenEvA framework improves our ability to describe the partonic regime through improved partonic calculations $|\mathcal{M}(\mu)|^2$. The showering regime can either be described by a traditional showering program such as Pythia [9, 10] or Herwig [11, 14, 15] or by the internal parton shower used in the GenEvA algorithm.[8] Regardless of which method is used, there will be no leading-logarithmic μ dependence in the event

[8] As we discuss more in the companion paper [1], the reason to use a traditional algorithm is that they have

JHEP12(2008)010

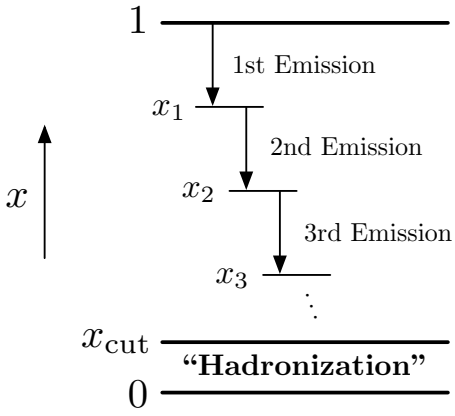


Figure 10: Structure of the toy theory. Starting at the scale $x = 1$, the system can generate "radiation" at scales x_1, x_2, ..., with the x_i values always decreasing. At the scale x_{cut}, the perturbative radiation matches onto a "hadronization" scheme.

generation as long as $|\mathcal{M}(\mu)|^2$ contains the correct Sudakov factors and the same definition of μ is used in the partonic and showering regimes. Finally, the hadronization regime uses some model of the strong interactions to hadronize the obtained partons into the hadrons observed in the collider, and `GenEvA` inherits the same smooth showering/hadronization interface as traditional frameworks.

3. GenEvA in a toy example

To see how the `GenEvA` framework works in practice, it is instructive to first consider a toy example that includes all of the issues involved in constructing partonic calculations $|\mathcal{M}(\mu)|^2$, without the technical complications introduced by full QCD. In particular, the toy theory will only have a single unambiguous scale. In actual QCD, the scale at which an emission occurs will be ambiguous, both because of the ambiguity in the choice of evolution variable and the ambiguity of how to pair together daughter particles to form mother particles. We will see how to deal with these additional complications in section 4.

To build the analogy with QCD, we need an example theory with both analytic cross-section information as well as a separate phenomenological description. We will use a slightly modified version of the toy example introduced in ref. [44]. After reviewing the toy theory, we will describe "Monte Carlo space" $\text{dMC}(\mu)$ and the kinds of partonic calculations $|\mathcal{M}(\mu)|^2$ that can be distributed across Monte Carlo space.

3.1 Review of toy theory

Consider a system that can radiate off "photons" as in figure 10. After emission of a

already been tuned to data, though there are some technical issues regarding evolution variables that would have to be overcome to use this option. If the `GenEvA` algorithm is used in the showering regime, then the user gains the freedom to adjust the scale μ used in the partonic calculations even after detector simulation.

JHEP12(2008)010

photon, the system has energy x left for further emissions with

$$0 \leq x < x_{\text{start}} \leq 1\,, \tag{3.1}$$

where x_{start} was the energy of the system before the emission (with $x_{\text{start}} = 1$ initially), so the energy of the radiated photon is $x_{\text{start}} - x$.[9] The different phase space integration ranges for n emissions are given by

$$\begin{aligned} &\mathrm{d}\Phi_0 = 1\,, \qquad \mathrm{d}\Phi_1 = \mathrm{d}x_1\,, \qquad \mathrm{d}\Phi_2 = \mathrm{d}x_1\,\mathrm{d}x_2\,\theta(x_1 > x_2)\,, \qquad \dots\,, \\ &\mathrm{d}\Phi_n = \mathrm{d}x_1 \prod_{i=2}^{n} \mathrm{d}x_i\,\theta(x_{i-1} > x_i)\,, \end{aligned} \tag{3.2}$$

and the restriction $0 \leq x_i \leq 1$ is implied.

The Born "cross section" corresponds to no emissions at zeroth order in perturbation theory. At first order in perturbation theory, there is a contribution to the cross section from real and virtual photon emissions. We define

$$\frac{\mathrm{d}\sigma_B}{\mathrm{d}x} = B\,\delta(x)\,, \qquad \frac{\mathrm{d}\sigma_V}{\mathrm{d}x} = \alpha\left(\frac{B}{2\epsilon} + V\right)\delta(x)\,, \qquad \frac{\mathrm{d}\sigma_R}{\mathrm{d}x} = \alpha B\,\frac{R(x)}{x^{1+2\epsilon}}\,, \tag{3.3}$$

where we are doing a perturbative expansion in α, and ϵ is the dimensional regularization parameter in $d = 1 - 2\epsilon$ dimensions. We need

$$\lim_{x\to 0} R(x) = 1\,, \tag{3.4}$$

such that the integral over the real contribution cancels the infrared divergence in the virtual contribution,

$$\int_0^1 \mathrm{d}x\,\frac{R(x)}{x^{1+2\epsilon}} = \int_0^1 \mathrm{d}x\,\frac{1}{x^{1+2\epsilon}} + \int_0^1 \mathrm{d}x\,\frac{R(x)-1}{x} = -\frac{1}{2\epsilon} + \text{finite}\,, \tag{3.5}$$

yielding the total cross section to $\mathcal{O}(\alpha)$

$$\sigma_{\text{NLO}} = B + \alpha V + \alpha B \int_0^1 \mathrm{d}x\,\frac{R(x)-1}{x}\,. \tag{3.6}$$

Next, we define a function $Q(x)$ which plays the role of the splitting function in QCD and reproduces the full radiation in the singular limit $x \to 0$. That is, from eq. (3.4) we require

$$\lim_{x\to 0} Q(x) = 1\,, \tag{3.7}$$

such that

$$\lim_{x\to 0} \frac{R(x) - Q(x)}{x} = \text{finite}\,. \tag{3.8}$$

[9] Unlike ref. [44], we take x to be the energy of the system available for radiation, rather than the energy of the emitted photon. Thus, x is continuously decreasing and plays the role of a parton shower evolution variable. Note that this means that this model has the somewhat unusual property of having singularities if the photons radiate with maximum energy, rather than with minimum energy.

JHEP12(2008)010

The simplest choice would be $Q(x) = 1$. In general, a nontrivial dependence on x is possible, in which case we require $Q(x)$ to be positive between $x = 0$ and $x = 1$. The function $Q(x)$ can be used to define

$$\Delta_Q(x_1, x_2) = \exp\left[-\alpha \int_{x_2}^{x_1} \mathrm{d}x \, \frac{Q(x)}{x}\right], \tag{3.9}$$

which plays the role of the Sudakov factor, and which we will often just call Δ for simplicity. Finally, the splitting function together with the Sudakov factor can be used to define a parton shower with the differential probability to branch (emit a photon) at x given by

$$\mathrm{d}\mathcal{P}(x) = \alpha \, \frac{Q(x)}{x} \, \Delta(x_{\text{start}}, x) \, \mathrm{d}x \, . \tag{3.10}$$

As before, x_{start} is the energy of the system prior to the branching. We also define a lower cutoff x_{cut} on the parton shower, which corresponds to the scale at which a QCD parton shower would be matched onto a hadronization model. In the following, the parton shower defined by eq. (3.10) will serve as the phenomenological model in our toy theory.

3.2 Phase space with a matching scale

As discussed in section 2.3, to avoid any double-counting between the parton shower started from a system with n emissions and one with $m > n$ emissions, we want to define phase space in the presence of matching scales. Because we are working with a single-scale system, it is straightforward to amend the ordinary phase space $\mathrm{d}\Phi_n$ in eq. (3.2) with a set of matching scales $\{\mu_i\}_{i<n}$. We have

$$\mathrm{d}\Phi_n(\{\mu_i\}_{i<n}) = \mathrm{d}x_1 \, \theta(x_1 > \mu_0) \prod_{i=2}^{n} \mathrm{d}x_i \, \theta(x_{i-1} > x_i) \, \theta(x_i > \mu_{i-1}) \, , \tag{3.11}$$

where the restriction $0 \leq x_i \leq 1$ is again implied. An illustration of eq. (3.11) for $n \leq 2$ is shown in figure 11. Before going on to explain eq. (3.11) in more detail, we want to highlight a few main points:

1. For each n, the scale μ_n is the scale at which we start the "parton shower" when acting on n-emission phase space $\mathrm{d}\Phi_n$. In general, $\mu_n \equiv \mu_n(\Phi_n)$ can be a function of $\Phi_n \equiv \{x_i\}_{i \leq n}$. This can lead to some counter-intuitive but still well-defined situations. The example shown in figure 11 has a seemingly pathological situation where $\mu_1(x_1) < \mu_0$ for certain values of x_1, but this is perfectly consistent and even potentially useful. The only restriction on the functional form of $\mu_n(\Phi_n)$ for eq. (3.11) to make sense is that $\mu_n(\Phi_n) \leq x_n$.

2. In accordance with our general discussion in section 2.3, the restriction $x_i > \mu_{i-1}$ in eq. (3.11) cuts out the region of $\mathrm{d}\Phi_n$ for any $n \geq i$ that gets populated by the parton shower when acting on $\mathrm{d}\Phi_{i-1}$, thus guaranteeing that there is no phase space double-counting.

JHEP12(2008)010

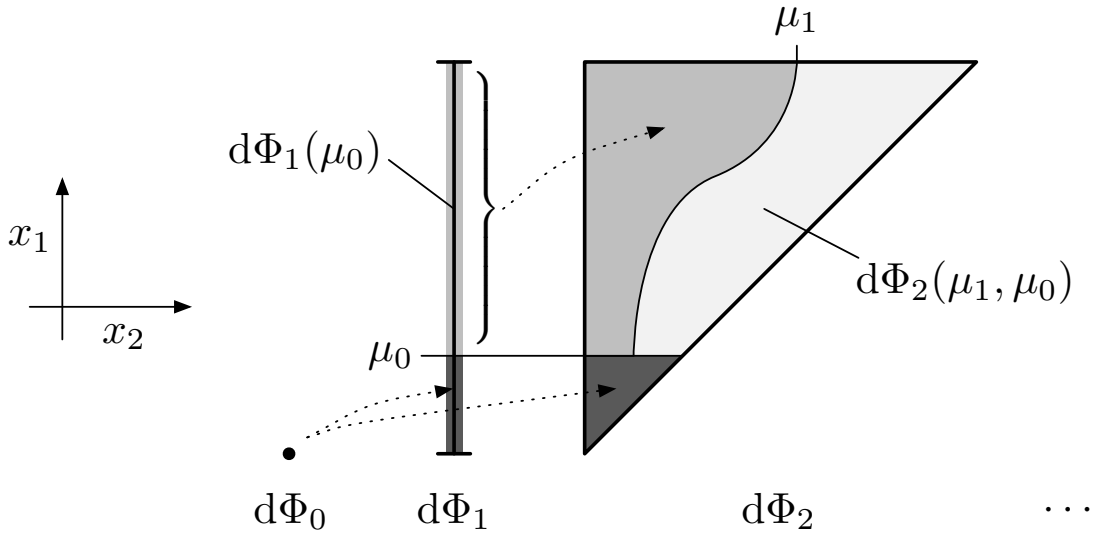


Figure 11: The scale-dependent phase space $\mathrm{d}\Phi_n(\{\mu_i\}_{i<n})$ in the toy theory. The scale μ_0 is a fixed number, whereas $\mu_1 \equiv \mu_1(x_1)$ can be a function of x_1. $\mathrm{d}\Phi_1(\mu_0)$ is given by $x_1 > \mu_0$, and $\mathrm{d}\Phi_2(\mu_0, \mu_1)$ is given by $x_1 > \mu_0$ and $\mu_1(x_1) < x_2 < x_1$.

3. If we have only calculations for up to $n_{\max}$ emissions, then to avoid dead zones in phase space, we have to start the parton shower at $\mu_{n_{\max}}(\Phi_{n_{\max}}) = x_{n_{\max}}$ when acting on the $n_{\max}$-emission contribution, such that $\mathrm{d}\Phi_n$ for $n > n_{\max}$ is completely covered by the parton shower. The easiest way to get full phase space coverage with no dead zones is to simply use a single fixed matching scale μ, with $\mu_n(\Phi_n) = \mu$ for $n < n_{\max}$ and $\mu_{n_{\max}}(\Phi_{n_{\max}}) = x_{n_{\max}}$. This also avoids some of the seeming pathologies mentioned above.

To explain eq. (3.11) in more detail, we take a closer look at how the toy parton shower works. Let $\mathrm{MC}_{x_1,x_2,\ldots}(x)$ denote an event for which emissions have occurred at x_1, x_2, …, and the system has energy x left to radiate. An event in its initial state is denoted by MC(1), which means that no photons have been emitted, and the total energy available for radiation is $x = 1$. The event then evolves according to the parton shower eq. (3.10), giving

$$\begin{aligned}
\mathrm{MC}(1) &\to \Delta(1, x_{\mathrm{cut}})\,\mathrm{MC}(x_{\mathrm{cut}}) + \int_{x_{\mathrm{cut}}}^{1} \mathrm{d}x_1 \left[\alpha\,\frac{Q(x_1)}{x_1}\,\Delta(1, x_1)\right] \mathrm{MC}_{x_1}(x_1) \\
&\to \Delta(1, x_{\mathrm{cut}})\,\mathrm{MC}(x_{\mathrm{cut}}) + \int_{x_{\mathrm{cut}}}^{1} \mathrm{d}x_1 \left[\alpha\,\frac{Q(x_1)}{x_1}\,\Delta(1, x_1)\right] \Delta(x_1, x_{\mathrm{cut}})\,\mathrm{MC}_{x_1}(x_{\mathrm{cut}}) \\
&\quad + \int_{x_{\mathrm{cut}}}^{1} \mathrm{d}x_1 \int_{x_{\mathrm{cut}}}^{x_1} \mathrm{d}x_2 \left[\alpha\,\frac{Q(x_1)}{x_1}\,\Delta(1, x_1)\right] \left[\alpha\,\frac{Q(x_2)}{x_2}\,\Delta(x_1, x_2)\right] \mathrm{MC}_{x_1,x_2}(x_2) \\
&\to \cdots .
\end{aligned} \tag{3.12}$$

The first term denotes the possibility that no emissions happen down to the "hadronization scale" x_{cut}, which occurs with the no-branching probability $\Delta(1, x_{\mathrm{cut}})$ and leaves an event

JHEP12(2008)010

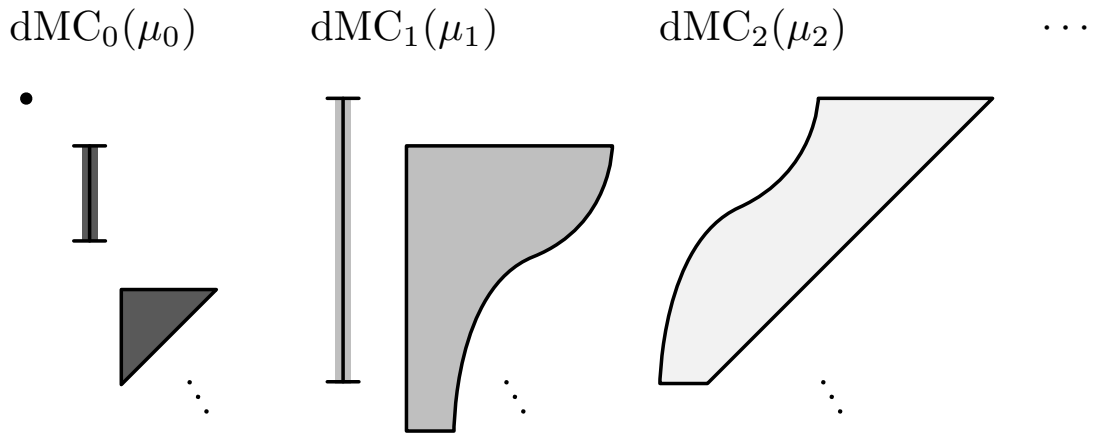


Figure 12: Definition of "Monte Carlo space" $\mathrm{dMC}_n(\mu_n)$ in the toy theory, which is obtained by collecting the various parts of $\mathrm{d}\Phi_n$ in figure 11.

with energy x_{cut} left for radiation. The second term denotes the possibility that exactly one emission happens above the scale x_{cut} and so on. The Sudakov factors are defined in such a way that the total probability is conserved,

$$\int_{x_{\mathrm{cut}}}^{1} \mathrm{d}x\, \alpha\, \frac{Q(x)}{x}\, \Delta(1,x) + \Delta(1,x_{\mathrm{cut}}) = 1\,, \tag{3.13}$$

so if we start with some number of events MC(1), the number of events after showering will still be the same.

Now imagine running the parton shower starting at the scale $\mu_n(\Phi_n)$ on an event for which n emissions have occurred at $\Phi_n = \{x_1, \ldots, x_n\}$. In this case

$$\begin{aligned}\mathrm{MC}_{x_1,\ldots,x_n}(\mu_n) \to{}& \Delta(\mu_n, x_{\mathrm{cut}})\, \mathrm{MC}_{x_1,\ldots,x_n}(x_{\mathrm{cut}}) \\ &+ \int_{x_{\mathrm{cut}}}^{\mu_n} \mathrm{d}x_{n+1} \left[\alpha \frac{Q(x_{n+1})}{x_{n+1}}\, \Delta(\mu_n, x_{n+1})\right] \mathrm{MC}_{x_1,\ldots,x_n,x_{n+1}}(x_{n+1}) \\ &\to \cdots\,.\end{aligned} \tag{3.14}$$

By assumption, the emissions are ordered with $x_{n+1} < x_n$, so for eq. (3.14) to be consistent, we need to choose $\mu_n(\Phi_n) \leq x_n$. If in addition we consider an $(n+1)$-emission event $\mathrm{MC}_{x_1,\ldots,x_n,x_{n+1}}(\mu_{n+1})$, then to avoid double-counting with the n-emission event, we have to restrict the range of x_{n+1} to $x_{n+1} > \mu_n$. Similarly, considering the effect of all of the 0- through n-emission samples, to avoid double-counting regions of phase space when running the parton shower, we always need to take $x_i > \mu_{i-1}$, as anticipated in eq. (3.11).

Putting together the phase space restriction from eq. (3.11) with the parton shower evolution of eq. (3.12), we can now define differential "Monte Carlo space" $\mathrm{dMC}_n(\mu_n)$ as

JHEP12(2008)010

the result of running the parton shower on n-emission phase space $\mathrm{d}\Phi(\{\mu_i\}_{i<n})$:

$$\begin{aligned}\mathrm{dMC}_n(\{\mu_i\}_{i\leq n}) = \mathrm{d}\Phi_n(\{\mu_i\}_{i<n}) \Bigg\{ & \Delta(\mu_n, x_{\mathrm{cut}}) \\ & + \mathrm{d}x_{n+1}\, \theta(\mu_n > x_{n+1} > x_{\mathrm{cut}}) \left[\alpha\, \frac{Q(x_{n+1})}{x_{n+1}}\, \Delta(\mu_n, x_{n+1})\right] \Delta(x_{n+1}, x_{\mathrm{cut}}) \\ & + \mathrm{d}x_{n+1}\, \mathrm{d}x_{n+2}\, \theta(\mu_n > x_{n+1} > x_{n+2} > x_{\mathrm{cut}}) \left[\alpha\, \frac{Q(x_{n+1})}{x_{n+1}}\, \Delta(\mu_n, x_{n+1})\right] \\ & \quad \times \left[\alpha\, \frac{Q(x_{n+2})}{x_{n+2}}\, \Delta(x_{n+1}, x_{n+2})\right] \Delta(x_{n+2}, x_{\mathrm{cut}}) \\ & + \cdots \Bigg\}, \end{aligned} \tag{3.15}$$

where the restriction $0 \leq x_i \leq 1$ is again implied. Note that the Monte-Carlo space $\mathrm{dMC}_n(\{\mu_i\}_{i\leq n})$ has variable multiplicity. An illustration of eq. (3.15) for $n \leq 2$ is shown in figure 12.

With the appropriate choice of μ_n, as discussed above,

$$\sum_{n=0}^{n_{\max}} \mathrm{dMC}_n(\mu_n) \to \sum_{n=0}^{\infty} \mathrm{d}\Phi_n(x_{\mathrm{cut}}) \tag{3.16}$$

gives a covering of phase space with no dead zones and no double-counting. Here, $\mathrm{d}\Phi_n(x_{\mathrm{cut}})$ represents running the hadronization model on the partons at the scale x_{cut}. The arrow indicates that $\mathrm{dMC}_n(\mu_n)$ includes additional weight information from the splitting functions and Sudakov factors.

3.3 The master formula

A compact master formula for the fully differential cross section coming from our as-of-yet undefined partonic calculation interfaced with the parton shower is given in analogy with eq. (2.6):

$$\mathrm{d}\sigma = \sum_{n=0}^{n_{\max}} |\mathcal{M}_n(\mu_n)|^2\, \mathrm{dMC}_n(\mu_n)\,. \tag{3.17}$$

As already mentioned, $|\mathcal{M}_n(\mu_n)|^2$ is a function of Φ_n. More generally, $|\mathcal{M}_n(\mu_n)|^2$ can also depend on the scales $\{\mu_i\}_{i<n}$ that are implicit in $\mathrm{dMC}_n(\mu_n)$.

It is at this point that the different concepts behind the parton shower come into play. As an instructive exercise, imagine we only know the Born cross section B and the splitting function $Q(x)$. In that case, the best approximation to the differential cross section we can write down is

$$\mathrm{d}\sigma = |\mathcal{M}_0(\mu_0)|^2\, \mathrm{dMC}_0(\mu_0) \qquad \text{with} \qquad \mu_0 = 1\,, \quad |\mathcal{M}_0(\mu_0 = 1)|^2 = B\,. \tag{3.18}$$

However, there is an entirely equivalent description of the same physics in terms of the differential cross section

$$\mathrm{d}\sigma = |\mathcal{M}_0(\mu_0)|^2\, \mathrm{dMC}_0(\mu_0) + |\mathcal{M}_1(\mu_1)|^2\, \mathrm{dMC}_1(\mu_1)\,, \tag{3.19}$$

where we now have to choose $\mu_1 = x_1$ to avoid dead zones, and the partonic calculations are defined as

JHEP12(2008)010

	$\lvert\mathcal{M}_0(\mu_0)\rvert^2$	$x\lvert\mathcal{M}_1(x)\rvert^2$
Parton Shower (LL)	$B\,\Delta_Q(1,\mu_0)$	$\alpha B\,Q(x)\,\Delta_Q(1,x)$
Tree-Level (LO)	B	$\alpha B\,R(x)$
Sudakov-Improved (LO/LL)	$B\,\Delta_Q(1,\mu_0)$	$\alpha B\,R(x)\,\Delta_Q(1,x)$
NLO Slicing (NLO)	$\bar{B}(\mu_0)$	$\alpha B\,R(x)$
NLO Subtraction (NLO/LL)	$\bar{B}(\mu_0)\,\Delta_Q(1,\mu_0)$	$\alpha B\,(R(x)-Q(x))+\alpha\bar{B}(\mu_0)\,Q(x)\,\Delta_Q(1,x)$
NLO Elegant (NLO/LL)	$\sigma_{\text{NLO}}\,\Delta_T(1,\mu_0)$	$\alpha B\,R(x)\,\Delta_T(1,x)$

Table 1: Summary of the various possible one-emission partonic calculations in the toy theory. At the one-emission level, all of the theory information can be summarized by a zero-emission calculation $|\mathcal{M}_0(\mu_0)|^2$ with a matching scale μ_0 and a one-emission calculation $|\mathcal{M}_1(x)|^2$ with a matching scale $\mu_1 = x$. The definitions of the samples and symbols are given in the text.

Parton shower (LL):

$$|\mathcal{M}_0(\mu_0)|^2 = B\,\Delta(1,\mu_0)\,,$$
$$|\mathcal{M}_1(x_1)|^2 = \alpha B\,\frac{Q(x_1)}{x_1}\,\Delta(1,x_1)\,. \tag{3.20}$$

In eq. (3.18), we would say that the boundary between the partonic calculation and the parton shower is $\mu = 1$, whereas in eq. (3.20), we would say that the boundary is some scale $\mu = \mu_0$. This is an example of the ambiguity of the parton shower having a meaning above or below the scale μ, as discussed in figure 2. This flexibility in changing the value of μ is a feature and not a bug, and it comes from the fact that the parton shower is both a phenomenological model based on QCD and a well-defined expansion of QCD. The only difference between the two descriptions are the words and perhaps algorithms we use above and below μ. But the freedom to incorporate a Sudakov factor Δ into a partonic calculation $|\mathcal{M}(\mu)|^2$ as needed is what makes it possible to define partonic calculations that include leading-logarithmic resummation.

3.4 Partonic matrix elements for the first emission

With the toy master formula eq. (3.17) in hand, we simply need to define the best partonic calculations $|\mathcal{M}(\mu)|^2$ we can with the available information from the toy theory. Apart from issues of Monte Carlo efficiency, this definition can be dictated by physics considerations alone, as the `GenEvA` framework can distribute any $|\mathcal{M}(\mu)|^2$ across dMC(μ). We will start in this subsection by only considering the first emission in our toy theory, writing for simplicity $x \equiv x_1$. The generalization to multiple emission will be given in the next subsection. Keeping in mind the freedom offered by eq. (3.19), we can define everything in terms of

$$|\mathcal{M}_0(\mu_0)|^2 \qquad \text{and} \qquad |\mathcal{M}_1(\mu_1 = x)|^2\,. \tag{3.21}$$

A summary of the partonic calculations in this subsection is given in table 1.

The most naive use of the analytic cross section information from eq. (3.3) is to choose

JHEP12(2008)010

Tree level (LO):

$$|\mathcal{M}_0(\mu_0)|^2 = B\,,$$
$$|\mathcal{M}_1(x)|^2 = \alpha B\,\frac{R(x)}{x}\,. \tag{3.22}$$

The total and one-emission cross sections resulting from eq. (3.22) are[10]

$$\sigma = B + \alpha B \int_{\mu_0}^1 \mathrm{d}x\,\frac{R(x)}{x}\,,$$
$$\frac{\mathrm{d}\sigma}{\mathrm{d}x} = \alpha B\,\frac{1}{x}\begin{cases} R(x) & \text{for} \quad x > \mu_0 \\ Q(x)\,\Delta(\mu_0, x) & \text{for} \quad x < \mu_0\,, \end{cases} \tag{3.23}$$

which is what one would obtain if events from a tree-level event generator were passed to a parton shower program.[11]

The main problem with the choice in eq. (3.22) is that it does not sum the leading-logarithmic contributions, which although technically of higher order in perturbation theory, can be numerically very important. This leads to the logarithmic dependence on the unphysical matching scale μ_0 in both σ and $\mathrm{d}\sigma/\mathrm{d}x$ in eq. (3.23). The reason is of course that since the parton shower does include this resummation, it has logarithmic dependence on the starting scale μ_0. Thus, if the tree-level results are combined with the parton shower using eq. (3.17), we are left with this logarithmic dependence on μ_0 in eq. (3.23).

To improve the situation, we can resum the leading logarithms in the partonic calculation by supplementing the naive tree-level result with Sudakov factors:

Sudakov-improved (LO/LL):

$$|\mathcal{M}_0(\mu_0)|^2 = B\,\Delta(1, \mu_0)\,,$$
$$|\mathcal{M}_1(x)|^2 = \alpha B\,\frac{R(x)}{x}\,\Delta(1, x)\,. \tag{3.24}$$

At leading order in α, this result is identical to the tree-level result eq. (3.22). The resulting total and one-emission cross sections are now given by

$$\sigma = B + \alpha B \int_{\mu_0}^1 \mathrm{d}x\,\frac{R(x) - Q(x)}{x}\,\Delta(1, x)\,,$$
$$\frac{\mathrm{d}\sigma}{\mathrm{d}x} = \alpha B\,\frac{1}{x}\begin{cases} R(x)\,\Delta(1, x) & \text{for} \quad x > \mu_0 \\ Q(x)\,\Delta(1, \mu_0)\,\Delta(\mu_0, x) & \text{for} \quad x < \mu_0\,. \end{cases} \tag{3.25}$$

The factor of $\Delta(1, \mu_0)$ in $\mathrm{d}\sigma/\mathrm{d}x$ for $x < \mu_0$ comes from $|\mathcal{M}_0(\mu_0)|^2$ and cancels the μ_0 dependence in $\Delta(\mu_0, x)$, which is generated by the parton shower. Furthermore, since we

[10] In what follows, $\mathrm{d}\sigma/\mathrm{d}x$ always refers to the inclusive differential cross section obtained by integrating over all values of x_i with $i \geq 2$.

[11] Actually, eq. (3.22) is slightly better than what one would get by blindly running a tree-level event generator through a parton shower program, because implicit in the definition of $|\mathcal{M}_0(\mu_0)|^2$ is that the parton shower will run from μ_0 to avoid double-counting. Out of the box, an ordinary parton shower program would start the parton shower at $x = 1$ for the 0-emission piece, which would lead to double-counting.

JHEP12(2008)010

know from eq. (3.8) that $(R(x) - Q(x))/x$ is finite as $x \to 0$, the logarithmic dependence on μ_0 in the total cross section cancels, too.

Since eq. (3.25) does not exhibit leading-logarithmic dependence on the matching scale μ_0, eq. (3.24) provides a well-behaved merging between matrix elements and the parton shower. This result is consistent with the observation made in ref. [23] that matrix element results need to be supplemented with Sudakov factors to allow for a merging with parton shower algorithms. What is important, however, is that we did not need a special algorithm to implement eq. (3.24), because this is just a choice of what partonic matrix elements to use in the master formula eq. (3.17). Including the correct Sudakov factors is required by physics considerations alone and, for example, has nothing to do with the fact that the `GenEvA` algorithm happens to have an efficient mechanism to implement this LO/LL merged sample.

The total and one-emission cross sections in eqs. (3.23) and (3.25) are correct to $\mathcal{O}(1)$ and $\mathcal{O}(\alpha)$, respectively, which is of course because we only used tree-level information so far. To also get the total cross section correct at $\mathcal{O}(\alpha)$ (NLO) we need to improve the partonic calculation by including the virtual corrections to the zero-emission cross section, given by $\mathrm{d}\sigma_V/\mathrm{d}x$ in eq. (3.3). There are numerous strategies one could consider. If we do not wish to reproduce the leading-logarithmic resummation, and only want to get observables accurate to NLO, then we can implement a naive slicing scheme [73–78]:

NLO slicing (NLO):

$$\begin{aligned} |\mathcal{M}_0(\mu_0)|^2 &= \tilde{B}(\mu_0)\,, \\ |\mathcal{M}_1(x)|^2 &= \alpha B\,\frac{R(x)}{x}\,, \end{aligned} \tag{3.26}$$

where

$$\begin{aligned} \tilde{B}(\mu_0) &= B + \alpha V + \alpha B \lim_{\epsilon\to 0}\biggl[\frac{1}{2\epsilon} + \int_0^{\mu_0} \mathrm{d}x\,\frac{R(x)}{x^{1+2\epsilon}}\biggr] \\ &= B + \alpha V + \alpha B \int_0^{\mu_0} \mathrm{d}x\,\frac{R(x)-1}{x} - \alpha B \int_{\mu_0}^1 \mathrm{d}x\,\frac{1}{x}\,. \end{aligned} \tag{3.27}$$

The cross sections following from eq. (3.26) are

$$\begin{aligned} \sigma &= \tilde{B}(\mu_0) + \alpha B \int_{\mu_0}^1 \mathrm{d}x\,\frac{R(x)}{x} = B + \alpha V + \alpha B \int_0^1 \mathrm{d}x\,\frac{R(x)-1}{x} = \sigma_{\mathrm{NLO}}\,, \\ \frac{\mathrm{d}\sigma}{\mathrm{d}x} &= \alpha\,\frac{1}{x}\begin{cases} B\,R(x) & \text{for} \quad x > \mu_0 \\ \tilde{B}(\mu_0)\,Q(x)\,\Delta(\mu_0, x) & \text{for} \quad x < \mu_0\,. \end{cases} \end{aligned} \tag{3.28}$$

The total cross section now equals the correct NLO cross section σ_{NLO} from eq. (3.6) and has no μ_0 dependence. Furthermore, all observables are accurate to NLO in the $\mu_0 \to 0$ limit. However, as in the tree-level result eq. (3.22), the leading logarithmic dependence is not resummed, meaning that the one-emission cross section $\mathrm{d}\sigma/\mathrm{d}x$ does not have the correct Sudakov suppression for small x and has large μ_0 dependence from both $\tilde{B}(\mu_0)$ and $\Delta(\mu_0, x)$.

JHEP12(2008)010

To include the resummation, Sudakov factors must be included in the partonic calculation in such a way that all observables are still accurate to NLO when they are expanded in α. This can be achieved using a subtraction method similar to the one proposed by Frixione and Webber in MC@NLO [44]. For the purpose of this discussion, we do not need to know the actual algorithm used in that paper to generate their results, only the corresponding partonic calculation itself:

NLO subtraction (NLO/LL):

$$\begin{aligned} |\mathcal{M}_0(\mu_0)|^2 &= \bar{B}(\mu_0)\,\Delta_Q(1,\mu_0)\,, \\ |\mathcal{M}_1(x,\mu_0)|^2 &= \alpha B\,\frac{R(x)-Q(x)}{x} + \alpha\bar{B}(\mu_0)\,\frac{Q(x)}{x}\,\Delta_Q(1,x)\,, \end{aligned} \tag{3.29}$$

where we made the dependence of Δ on $Q(x)$ explicit, and we defined

$$\begin{aligned} \bar{B}(\mu_0) &= \tilde{B}(\mu_0) + \alpha B \int_{\mu_0}^1 \mathrm{d}x\,\frac{Q(x)}{x} \\ &= B + \alpha V + \alpha B \int_0^{\mu_0} \mathrm{d}x\,\frac{R(x)-Q(x)}{x} - \alpha B \int_0^1 \mathrm{d}x\,\frac{1-Q(x)}{x}\,. \end{aligned} \tag{3.30}$$

Note that we have used the fact that the one-emission piece can in general depend on μ_0. The cross sections predicted by the substraction method in eq. (3.29) are

$$\begin{aligned} \sigma &= \bar{B}(\mu_0) + \alpha B \int_{\mu_0}^1 \mathrm{d}x\,\frac{R(x)-Q(x)}{x} = \sigma_{\mathrm{NLO}}\,, \\ \frac{\mathrm{d}\sigma}{\mathrm{d}x} &= \alpha\,\frac{1}{x}\begin{cases} B\,R(x) - Q(x)[B - \bar{B}(\mu_0)\Delta_Q(1,x)] & \text{for} \quad x > \mu_0 \\ \bar{B}(\mu_0)\,Q(x)\,\Delta_Q(1,\mu_0)\Delta_Q(\mu_0,x) & \text{for} \quad x < \mu_0\,. \end{cases} \end{aligned} \tag{3.31}$$

By expanding the above expressions to $\mathcal{O}(\alpha)$, one can easily show that the predictions of the subtraction method are equal at NLO to the predictions of the slicing method in eq. (3.28). Furthermore, in addition to producing the correct NLO total cross section, the one-emission cross section in the subtraction method has no leading-logarithmic μ_0 dependence and exhibits the correct Sudakov suppression for small x.

In ref. [44], the form of eq. (3.29) was selected mainly for its algorithmic simplicity and the lack of explicit ϵ dependence. The important difference is that in ref. [44] each of the two terms in $|\mathcal{M}_1(x,\mu_0)|^2$ is generated by a separate event sample. This can be inefficient and lead to events with negative weights, since in general, the relative size of the two terms can be very different, and the first term can also become negative. In contrast, in our approach, the form of $|\mathcal{M}_1(x,\mu_0)|^2$ in eq. (3.29) is just another choice of a partonic calculation, so it can be used in the master formula eq. (3.17) to generate a single, positive-weight event sample.

Although we could in principle use eq. (3.29) in our approach, it still has some drawbacks. First, due to the rather complicated structure of eqs. (3.29) and (3.30), they are tedious to generalize to more emissions or higher orders in α. Second, since $|\mathcal{M}_1(x,\mu_0)|^2$

JHEP12(2008)010

depends on the subtraction function $Q(x)$, it is in general not positive definite, and for pathological choices of $Q(x)$ can lead to negative cross sections.[12]

However, there is no deep conceptual reason why subtractions must be used to generate Monte Carlo. Indeed, there are many different ways to get results that are accurate at NLO after expanding in α, that have no leading-logarithmic μ_0 dependence, and that have the correct Sudakov suppression in the small x limit. One particularly elegant method proposed by Nason in ref. [46] in a slightly different context is

NLO elegant (NLO/LL):

$$\begin{aligned} |\mathcal{M}_0(\mu_0)|^2 &= \sigma_{\text{NLO}}\, \Delta_T(1,\mu_0)\,, \\ |\mathcal{M}_1(x)|^2 &= \alpha\, \sigma_{\text{NLO}}\, \frac{T(x)}{x}\, \Delta_T(1,x)\,, \end{aligned} \tag{3.32}$$

where σ_{NLO} is the total NLO cross section from eq. (3.6), $T(x)$ is an effective "splitting function"

$$T(x) = \frac{B}{\sigma_{\text{NLO}}}\, R(x)\,, \tag{3.33}$$

and Δ_T is the Sudakov factor obtained from $T(x)$ in analogy to eq. (3.9). The cross sections are now

$$\begin{aligned} \sigma &= \sigma_{\text{NLO}} \left[\Delta_T(1,\mu_0) + \int_{\mu_0}^1 \mathrm{d}x\, \alpha\, \frac{T(x)}{x}\, \Delta_T(1,x) \right] = \sigma_{\text{NLO}}\,, \\ \frac{\mathrm{d}\sigma}{\mathrm{d}x} &= \alpha\, \sigma_{\text{NLO}}\, \frac{1}{x} \begin{cases} T(x)\, \Delta_T(1,x) & \text{for} \quad x > \mu_0 \\ Q(x)\, \Delta_T(1,\mu_0)\, \Delta_Q(\mu_0,x) & \text{for} \quad x < \mu_0\,. \end{cases} \end{aligned} \tag{3.34}$$

As in the subtraction method, the total cross section is identical to the NLO result, and the one-emission cross section has the correct Sudakov suppression with no leading logarithmic μ_0 dependence. Note that the latter again relies on the fact that $R(x)$ and $Q(x)$ have the same singularities for $x \to 0$, so Δ_T and Δ_Q resum the same leading logarithms.

The form of eq. (3.32) is identical to the merged LO/LL example of eq. (3.24) by replacing $Q(x) \to T(x)$ and $B \to \sigma_{\text{NLO}}$. The nice feature of eq. (3.32) is that it is not only conceptually simple, but it also uses only functions that are completely well defined in an NLO calculation without ever needing to introduce an ad-hoc subtraction function. It is also clear from the above expressions that the cross sections are always guaranteed to be positive. In section 4, we show how to generalize this method to the realistic case of QCD. On the other hand, there is also no reason why subtractions could not be used in `GenEvA` if a subtraction method is more convenient for theoretical or numerical reasons.

3.5 Multiple emissions

The results from the previous subsection can be extended straightforwardly to multiple emissions in the partonic regime. At tree level, the partonic cross sections are given by

[12] Of course, for any reasonable choice of $Q(x)$, eq. (3.29) will lead to a positive weight NLO/LL event sample. Pathologies occur when, say, $R(x) - Q(x) \sim \mathcal{O}(\alpha^{-1})$.

JHEP12(2008)010

Tree level (LO):

$$|\mathcal{M}_n^{\rm LO}(\mu_n)|^2 = \frac{\alpha^n f_n^{\rm tree}(x_1, x_2, \ldots, x_n)}{x_1 x_2 \cdots x_n}\,, \tag{3.35}$$

where the symbol $f_n^{\rm tree}$ represents an ordinary tree-level calculation in the toy theory with the α dependence and x_i singularities made manifest. For example, $f_0^{\rm tree} = B$ and $f_1^{\rm tree}(x_1) = BR(x_1)$. We assume that in the singular regions of phase space

$$\lim_{x_n \to 0} \frac{f_n^{\rm tree}(x_1, \ldots, x_{n-1}, x_n) - f_{n-1}^{\rm tree}(x_1, \ldots, x_{n-1}) Q(x_n)}{x_n} = \text{finite}\,, \tag{3.36}$$

the analog of which is indeed true in the collinear limit of QCD as long as $Q(x)$ is generalized to include flavor information.

As discussed before, the tree-level results do not resum the leading-logarithmic behavior, which can be fixed by including Sudakov factors appropriately:

Sudakov-improved (LO/LL):

$$|\mathcal{M}_n^{\rm LO/LL}(\mu_n)|^2 = \frac{\alpha^n f_n^{\rm tree}(x_1, x_2, \ldots, x_n)}{x_1 x_2 \cdots x_n}\, \Delta_Q(1, x_1)\, \Delta_Q(x_1, x_2) \cdots \Delta_Q(x_n, \mu_n)\,. \tag{3.37}$$

This is precisely the result that was advocated in ref. [23] in order to merge matrix elements and parton showers.[13] To see that eq. (3.37) is well behaved, note that the difference across the boundary $x_n = \mu_{n-1}$ is[14]

$$\begin{aligned}
&\lim_{\varepsilon \to 0} \left(\frac{\mathrm{d}^n \sigma}{\mathrm{d}x_1 \cdots \mathrm{d}x_n}\bigg|_{x_n = \mu_{n-1} + \varepsilon} - \frac{\mathrm{d}^n \sigma}{\mathrm{d}x_1 \cdots \mathrm{d}x_n}\bigg|_{x_n = \mu_{n-1} - \varepsilon} \right) \\
&\quad = \alpha^n \frac{f_n^{\rm tree}(x_1, \ldots, x_{n-1}, x_n) - f_{n-1}^{\rm tree}(x_1, \ldots, x_{n-1}) Q(x_n)}{x_1 \cdots x_{n-1} x_n}\, \Delta_Q(1, x_n)\bigg|_{x_n = \mu_{n-1}} + \mathcal{O}(\alpha^{n+1})\,,
\end{aligned} \tag{3.38}$$

which has no large logarithmic dependence on μ_{n-1}, assuming eq. (3.36) holds.

As the example most relevant for creating our NLO/LO/LL merged sample, we finally consider the combination of the NLO results with higher-multiplicity tree-level matrix elements. This will give a partonic calculation that reproduces the exact NLO results, but also uses the full tree-level quantum interference for multiple emissions. Using the previous definitions of $\sigma_{\rm NLO}$ and $T(x)$, we find:

[13] Ref. [23] also advocated to evaluate the couplings α at successive intermediate x_i scales in the problem, which in QCD corresponds to including extra subleading-logarithmic information.

[14] Eq. (3.38) is again the inclusive differential cross section, where we have integrated over all values of x_i with $i > n$.

JHEP12(2008)010

Best combination (NLO/LO/LL):

$$
\begin{aligned}
|\mathcal{M}_0^{\text{NLO/LO/LL}}(\mu_0)|^2 &= \sigma_{\text{NLO}}\,\Delta_T(1,\mu_0)\,,\\
|\mathcal{M}_1^{\text{NLO/LO/LL}}(\mu_1)|^2 &= \frac{\alpha f_1^{\text{tree}}(x_1)}{x_1}\,\Delta_T(1,x_1)\,\Delta_Q(x_1,\mu_1)\,,\\
&\vdots\\
|\mathcal{M}_n^{\text{NLO/LO/LL}}(\mu_n)|^2 &= \frac{\alpha^n f_n^{\text{tree}}(x_1,x_2,\ldots,x_n)}{x_1x_2\cdots x_n}\,\Delta_T(1,x_1)\,\Delta_Q(x_1,x_2)\cdots\Delta_Q(x_n,\mu_n)\,.
\end{aligned}
\tag{3.39}
$$

For $n=1$ and $\mu_1 = x_1$ this reduces to eq. (3.32).

The factor of $\Delta_T(1,x_1)$ appears in $|\mathcal{M}_n(\mu_n)|^2$ to ensure that the total cross section has no large logarithmic dependence on μ_1.[15] In fact, this is the only difference between the LO/LL and NLO/LO/LL expressions for $|\mathcal{M}_n(\mu_n)|^2$, but it is crucial to ensure that all NLO observables (including the total cross section) are accurate to NLO with no large logarithms. In particular, it is straightforward to integrate eq. (3.39) to show that

$$
\sigma = \sigma_{\text{NLO}} + \alpha^2 \int_{\mu_0}^1 \mathrm{d}x_1 \int_{\mu_1}^{x_1} \mathrm{d}x_2\, \frac{f_2^{\text{tree}}(x_1,x_2) - f_1^{\text{tree}}(x_1)Q(x_2)}{x_1x_2}\,\Delta_T(1,x_1)\,\Delta_Q(x_1,x_2) + \mathcal{O}(\alpha^3)
\tag{3.40}
$$

which has no large logarithmic dependence in μ_0 or μ_1 as long as eq. (3.36) holds.

3.6 The importance of phase space projection

In the above discussion, we never needed to talk about the phase space projection feature from eq. (2.7) to write down our partonic calculations $|\mathcal{M}(\mu)|^2$. The reason is that the toy theory is a single-scale theory, so phase space projection acts trivially. Given a phase space point $\{x_1,x_2,\ldots,x_n\}$, we can simply define the intermediate resolution scales as $\bar{\mu}_{n-1} = x_{n-1}$, $\bar{\mu}_{n-2} = x_{n-2}$, and so on, such that the projected phase space points are

$$
\{x_1,x_2,\ldots,x_n\} \overset{\bar{\mu}_{n-1}}{\longrightarrow} \{x_1,x_2,\ldots,x_{n-1}\} \overset{\bar{\mu}_{n-2}}{\longrightarrow} \{x_1,x_2,\ldots,x_{n-2}\} \longrightarrow \ldots\,.
\tag{3.41}
$$

In full QCD, there are many more scale choices one could make for $\bar{\mu}$, and the difference between two different choices will yield logarithms (hopefully not large) of $\bar{\mu}_A/\bar{\mu}_B$. More importantly, in QCD there is an important requirement of final states being on shell, e.g. if we project two gluons into one mother gluon, we have to somehow shuffle momenta to put that mother gluon on shell.

So why use phase space projection in QCD if it is so hard? In the language of eq. (3.39), the reason is that we need some way of getting the analog of the factors of Δ_T to appear in $|\mathcal{M}_n(\mu_n)|^2$. In some sense, Δ_T is a property of the NLO calculation that we want the LO tree-level matrix elements to inherit, but the tree-level matrix elements do not have enough information to properly calculate Δ_T themselves. But if we take an n-emission matrix element and project it to a one-emission matrix element, then we can use our

[15] We cannot replace every Δ_Q with Δ_T because in QCD there is additional flavor structure, which would make this choice ill-defined.

JHEP12(2008)010

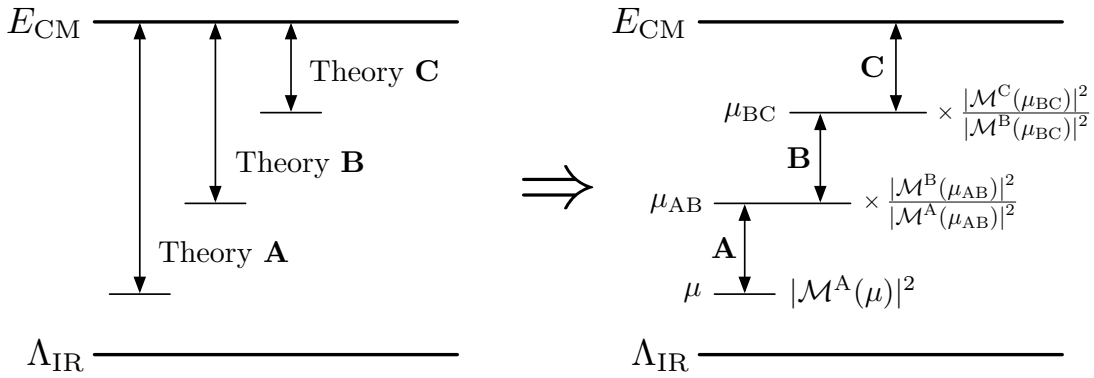


Figure 13: Using nested descriptions to define the best partonic calculations. Given a theory **A** valid down to the scale μ, the partonic calculation $|\mathcal{M}^{\rm A}(\mu)|^2$ can be improved using information from a presumably better theory **B** which is however only valid down to a scale $\mu_{\rm AB} > \mu$. The "matching coefficient" $|\mathcal{M}^{\rm B}(\mu_{\rm AB})|^2/|\mathcal{M}^{\rm A}(\mu_{\rm AB})|^2$ allows the full event sample to capture all the physics of theory **B**, while still retaining information about theory **A** below the scale $\mu_{\rm AB}$. The same logic can now be repeated with a third theory **C** which is only valid down to a scale $\mu_{\rm BC} > \mu_{\rm AB}$.

existing NLO machinery to NLO-improve the LO calculations. In formulas, we can write the n-emission cross section in eq. (3.39) as

$$|\mathcal{M}_n^{\rm NLO/LO/LL}(\mu_n)|^2 = |\mathcal{M}_n^{\rm LO/LL}(\mu_n)|^2 \times \frac{|\mathcal{M}_1^{\rm NLO/LL}(\bar{\mu}_1 = x_1)|^2}{|\mathcal{M}_1^{\rm LO/LL}(\bar{\mu}_1 = x_1)|^2}\,. \tag{3.42}$$

The ratio on the right-hand side equals precisely $\Delta_T(1, x_1)/\Delta_Q(1, x_1)$ and can be thought of as the "matching coefficient" between the NLO/LL calculation and the other Sudakov improved tree-level calculations.

In fact, eq. (3.42) is just a specific case of a more general version of matching different partonic descriptions and running between them. Imagine, we have a set of nested partonic descriptions **A**, **B**, **C**, and so on as in figure 13, then the best partonic calculation we can construct with the available information is

Best combination (A/B/C/$\cdots$):

$$|\mathcal{M}^{\rm A/B/C/\cdots}(\mu)|^2 = |\mathcal{M}^{\rm A}(\mu)|^2 \times \frac{|\mathcal{M}^{\rm B}(\mu_{\rm AB})|^2}{|\mathcal{M}^{\rm A}(\mu_{\rm AB})|^2} \times \frac{|\mathcal{M}^{\rm C}(\mu_{\rm BC})|^2}{|\mathcal{M}^{\rm B}(\mu_{\rm BC})|^2} \times \cdots\,, \tag{3.43}$$

where the various μ scales encode the multiplicities and energy scales at which the matching occurs.

Though obscured in the dMC(μ) notation, but clear from eq. (3.20), the NLO/LO/LL merged sample is equivalent to taking

$$\mathbf{A} = \text{LL}\,, \qquad \mathbf{B} = \text{LO/LL}\,, \qquad \mathbf{C} = \text{NLO/LL}\,, \tag{3.44}$$

where the LL description is just the parton shower written in the spirit of eq. (3.20). The factors of $|\mathcal{M}^{\rm LL}(\mu)|^2$ do not appear in eq. (3.42) because we use the parton shower not

JHEP12(2008)010

as partonic calculation but as phenomenological model. Though we will not pursue this direction in the present work, we expect that creating an N^{i}LO/N^{j}LL merged sample will just require finding appropriate descriptions **A**, **B**, **C**, and so on.

As a side note, the reversable phase space projection in the `GenEvA` algorithm [1] is equivalent to the statement that we have an analytic way of calculating $|\mathcal{M}^{\mathrm{LL}}(\mu)|^2$ for the parton shower at all scales μ. One of the reasons that the `GenEvA` algorithm is efficient comes from the realization that the Sudakov factors in

$$\frac{|\mathcal{M}^{\mathrm{LO/LL}}(\mu)|^2}{|\mathcal{M}^{\mathrm{LL}}(\mu)|^2} \tag{3.45}$$

cancel, reducing the number of calculations necessary to merge an LO/LL calculation with the parton shower.

The next two sections deal with the technical complications introduced by full QCD. However, the main ideas are captured by the toy example just presented, so the reader not interested in the technical details can safely skip directly to section 6 for the results of the NLO/LO/LL merged calculation in QCD.

4. The first emission in QCD

Having investigated the toy model in detail, we now move on to the realistic case of QCD. In this section, we will study how to implement the QCD cross section for $e^+e^- \to n$ jets up to $\mathcal{O}(\alpha_s)$, thus including the processes $e^+e^- \to q\bar{q}$ and $e^+e^- \to q\bar{q}g$. The extension to final states with more than three partons will be discussed in the next section, though most of the important physics considerations necessary to construct the NLO/LO/LL merged sample appear here.

The essential complication for $e^+e^- \to n$ jets compared to the toy example is the presence of multiple well-defined energy scales. This same complication appears in the process pp or $p\bar{p} \to n$ jets, so we anticipate that the same solutions present in the leptonic case should have a generalization to the hadronic case. At the end of the day, these complications are resolved by a thorough understanding of phase space, so we will begin with 3-body phase space before moving on to the relevant QCD calculations and the definition of the partonic matrix elements.

4.1 Three-body phase space

At the level of phase space, there are two complications in QCD compared with the toy model studied in the previous section. First, a single emission in QCD is specified by three independent variables, as opposed to the single variable in the toy model. Second, on top of these three variables specifying the hadronic kinematics, there are two additional variables describing the orientation of the hadronic system relative to the e^+e^- beamline.

There are many equivalent ways of choosing these five variables. For our discussion, it will be convenient to use the thrust axis to define the orientation of the hadronic system relative to the beamline, and we use Ω_2^T to describe this orientation. The remaining three variables describe the kinematics of the hadronic system relative to the thrust axis, which

JHEP12(2008)010

we choose to be the invariant mass t_a between the quark and the gluon, the invariant mass t_b between the antiquark and the gluon, as well as the azimuthal angle ϕ of the hadronic system relative to the thrust axis. In other words, we decompose Lorentz-invariant 3-body phase space as

$$\mathrm{d}\Phi_3 = \frac{1}{(4\pi)^5} E_{\mathrm{CM}}^2 \, \mathrm{d}\Omega_2^T \, \mathrm{d}t_a \, \mathrm{d}t_b \, \mathrm{d}\phi \,, \tag{4.1}$$

where $E_{\mathrm{CM}}^2 = (p_{e^+} + p_{e^-})^2$ and

$$t_a = \frac{(p_q + p_g)^2}{E_{\mathrm{CM}}^2} \,, \qquad t_b = \frac{(p_{\bar{q}} + p_g)^2}{E_{\mathrm{CM}}^2} \,. \tag{4.2}$$

For the remainder of the discussion, we will mainly ignore the Ω_2^T and ϕ dependence.

The key to constructing the analog of the master formula in eq. (3.17) is to have an unambiguous definition of the scale μ that will separate our QCD calculations from the phenomenological parton shower. For simplicity, we will assume that the parton shower uses virtuality

$$t \equiv \frac{Q^2}{E_{\mathrm{CM}}^2} \tag{4.3}$$

as the evolution variable. We discuss the issues with extending our results to more modern $p_\perp$-ordered showers in the companion paper [1]. To keep the notion of μ as a mass scale, we also define

$$\hat{\mu}^2 = \frac{\mu^2}{E_{\mathrm{CM}}^2} \,, \tag{4.4}$$

and use dimensionless variables in the remainder of this discussion. We anticipate running the `GenEvA` framework from $1 = E_{\mathrm{CM}}^2/E_{\mathrm{CM}}^2$ until the scale $\hat{\mu}^2$, attaching a parton shower starting from $\hat{\mu}^2$ going down to $t_{\mathrm{cut}} \simeq \Lambda_{\mathrm{IR}}^2/E_{\mathrm{CM}}^2$, and finally interfacing with a hadronization model at t_{cut}.

The total 3-parton phase space $0 \leq t_a \leq 1 - t_b \leq 1$ is shown in figure 14. The 3-jet region **3** is defined as the region where both t_a and t_b are larger than $\hat{\mu}^2$. The 2-jet region **2** corresponds to both t_a and t_b being less than $\hat{\mu}^2$. Finally, there is an ambiguous $\tilde{\mathbf{2}}_\mathbf{a}$ region where t_a is greater than $\hat{\mu}^2$ and t_b is less than $\hat{\mu}^2$, and a similar $\tilde{\mathbf{2}}_\mathbf{b}$ region where the roles of t_a and t_b are reversed. Formally, these phase space integration regions are defined as

$$\begin{aligned}
\int_{\mathbf{3}} \mathrm{d}t_a \, \mathrm{d}t_b &\equiv \int_{\hat{\mu}^2}^{1-\hat{\mu}^2} \mathrm{d}t_a \int_{\hat{\mu}^2}^{1-t_a} \mathrm{d}t_b \,, & \int_{\mathbf{2}} \mathrm{d}t_a \, \mathrm{d}t_b &\equiv \int_0^{\hat{\mu}^2} \mathrm{d}t_a \int_0^{\hat{\mu}^2} \mathrm{d}t_b \,, \\
\int_{\tilde{\mathbf{2}}_\mathbf{a}} \mathrm{d}t_a \, \mathrm{d}t_b &\equiv \int_0^{\hat{\mu}^2} \mathrm{d}t_b \int_{\hat{\mu}^2}^{1-t_b} \mathrm{d}t_a \,, & \int_{\tilde{\mathbf{2}}_\mathbf{b}} \mathrm{d}t_a \, \mathrm{d}t_b &\equiv \int_0^{\hat{\mu}^2} \mathrm{d}t_a \int_{\hat{\mu}^2}^{1-t_a} \mathrm{d}t_b \,.
\end{aligned} \tag{4.5}$$

By definition, the regions **2**, $\tilde{\mathbf{2}}_\mathbf{a}$, $\tilde{\mathbf{2}}_\mathbf{b}$ are covered by running the phenomenological parton shower on 2-body phase space starting at $\hat{\mu}^2$, with the $\tilde{\mathbf{2}}_\mathbf{a}$ and $\tilde{\mathbf{2}}_\mathbf{b}$ regions arising from small virtuality, small angle gluon radiation. Hence, they belong to $\mathrm{dMC}_2(\hat{\mu}^2)$. The remaining part of $\mathrm{d}\Phi_3$ is $\mathbf{3} \equiv \mathrm{d}\Phi_3(\hat{\mu}^2)$, and running the shower on **3** defines $\mathrm{dMC}_3(\hat{\mu}_3^2)$. Since $n_{\max} = 3$, we need to pick an appropriate scale $\hat{\mu}_3^2$ to avoid dead zones, similar to our discussion in the toy model in section 3.2. What we need is the QCD-version of the

JHEP12(2008)010

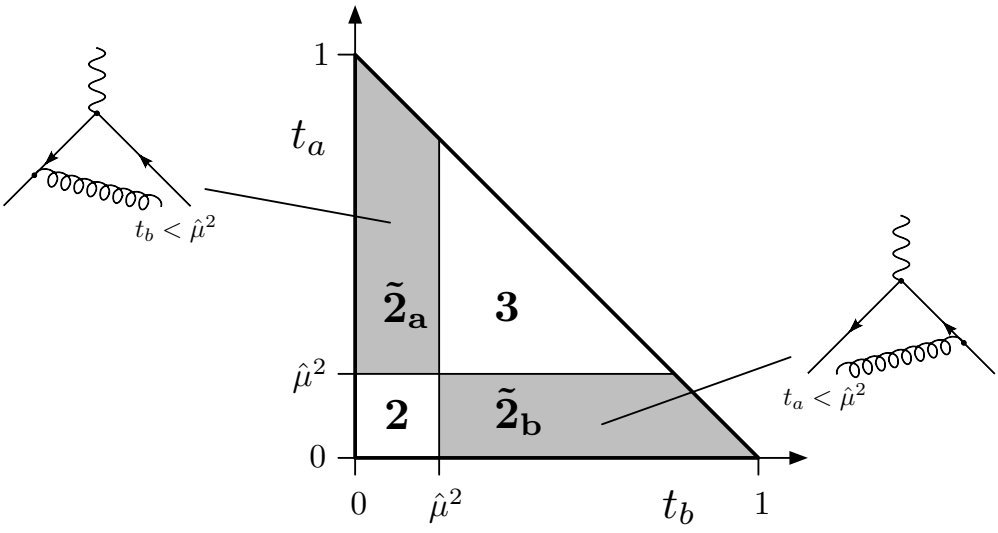


Figure 14: The available phase space $0 \leq t_a \leq 1 - t_b \leq 1$ for $e^+e^- \to q\bar{q}g$, where $t_{a,b}$ are defined in eq. (4.2). The separation between the 2-jet region and the 3-jet region is defined by a matching scale $\hat{\mu}^2$. Only region **3** has both t_a and t_b above the matching scale $\hat{\mu}^2$ and therefore corresponds to the 3-jet region of phase space. The ambiguous $\tilde{\mathbf{2}}$ regions could be generated by a parton shower cut off at $\hat{\mu}^2$ by "accidentally" populating a region closer to the QCD collinear singularity than $\hat{\mu}^2$. However, that part is also generated by emissions below $\hat{\mu}^2$ and is therefore covered by $\mathrm{dMC}_2(\hat{\mu}^2)$ and must be vetoed in $\mathrm{dMC}_3(\hat{\mu}_3^2)$.

condition $\mu_{n_{\max}} = x_{n_{\max}}$ we found in the toy model. Hence, we have to decide whether to take $\hat{\mu}_3^2$ equal to t_a or t_b, and in fact, either choice would leave no dead zones.

For the moment, imagine using the parton shower to cover all of phase space. Given the hard scattering process $e^+e^- \to q\bar{q}$, we can run the parton shower from $t_{\mathrm{start}} = 1$. The first emission can either come from a gluon being radiated from the quark at $t_{\mathrm{emit}} = t_a$ or from the antiquark at $t_{\mathrm{emit}} = t_b$. Regardless at what scale t_{emit} the emission occurs, the parton shower continues to run from that scale down, filling out the rest of phase space. Thus, knowing t_a and t_b alone does not permit a definition of the proper scale $\hat{\mu}^2 = t_{\mathrm{emit}}$, and a wrong choice could lead to potentially large logarithms $\alpha_s \log^2(t_a/t_b)$. The challenge is that a point Φ_3 gives us only access to t_a and t_b, and we have to decide how to properly account for the logarithms of the ratio of these two scales.

We will discuss several ways to resolve this ambiguity in the definition of $\hat{\mu}^2$, and for now we only provide a framework that allows the $\hat{\mu}^2 = t_a$ vs. t_b problem to be solved in principle. In particular, for now we will simply have two samples of 3-parton events, 3-parton events where we start the shower at t_a and 3-parton events where we start the shower at t_b, and let the user decide what partonic calculations to use for each sample. Thus, we will have three `GenEvA` event samples

$$\mathrm{dMC}_2(\hat{\mu}^2)\,, \qquad \mathrm{dMC}_3(\hat{\mu}_3^2 = t_a)\,, \qquad \mathrm{dMC}_3(\hat{\mu}_3^2 = t_b)\,, \tag{4.6}$$

which give a complete covering of phase space. To not introduce double-counting in the

JHEP12(2008)010

cross section, we now have to be careful to remember that there are two dMC_3 samples. We stress that this potential double-counting is conceptually very different from the double-counting between different phase space algorithms discussed earlier. Here, it arises from a scale ambiguity in QCD which has to be resolved by calculational means.

The $\tilde{\mathbf{2}}$ regions are yet another example of a scale ambiguity and how the notion of "above μ" and "below μ" can be so confusing. Since the dMC_2 sample covers the $\tilde{\mathbf{2}}_\mathbf{a}$ region, it has to be vetoed in dMC_3, despite the fact that it could have been generated by the phenomenological model with a large angle emission above $\hat{\mu}^2$. As seen in figure 14, a gluon emission could have occurred at a large angle at a large scale, while for the purposes of phase space this is part of $\mathrm{dMC}_2(\hat{\mu}^2)$, and so the actual scale of emission should be considered below $\hat{\mu}^2$.

4.2 The master formula

With the event samples of eq. (4.6), we can write the differential cross section in terms of the master formula

$$\mathrm{d}\sigma = \frac{\mathrm{d}\sigma_2(\hat{\mu}^2)}{\mathrm{d}\Phi_2}\,\mathrm{dMC}_2(\hat{\mu}^2) + \frac{\mathrm{d}\sigma_{3a}(t_a)}{\mathrm{d}\Phi_3}\,\mathrm{dMC}_3(t_a) + \frac{\mathrm{d}\sigma_{3b}(t_b)}{\mathrm{d}\Phi_3}\,\mathrm{dMC}_3(t_b)\,, \tag{4.7}$$

where we are using the notation $\mathrm{d}\sigma_i/\mathrm{d}\Phi$ instead of $|\mathcal{M}_i|^2$ because we anticipate integrating over some of the phase space variables to simplify expressions. The argument in the $\mathrm{d}\sigma_i$ functions reminds us of the scale at which we start running the parton shower for those samples. To avoid double-counting regions of phase space, the cross sections $\mathrm{d}\sigma_{3a}$ and $\mathrm{d}\sigma_{3b}$ must satisfy

$$\frac{\mathrm{d}\sigma_3}{\mathrm{d}\Phi_3} = \frac{\mathrm{d}\sigma_{3a}(t_a)}{\mathrm{d}\Phi_3} + \frac{\mathrm{d}\sigma_{3b}(t_b)}{\mathrm{d}\Phi_3}\,, \tag{4.8}$$

where $\mathrm{d}\sigma_3$ is the full 3-parton cross section. If we are not worried about large $\log^2(t_a/t_b)$ logarithms, then a simple definition of the $\mathrm{d}\sigma_{3i}$ partonic cross sections is

$$\frac{\mathrm{d}\sigma_{3a}(t_a)}{\mathrm{d}\Phi_3} = \frac{\mathrm{d}\sigma_3}{\mathrm{d}\Phi_3}\,\alpha_a(\Phi_3)\,, \qquad \frac{\mathrm{d}\sigma_{3b}(t_b)}{\mathrm{d}\Phi_3} = \frac{\mathrm{d}\sigma_3}{\mathrm{d}\Phi_3}\,\alpha_b(\Phi_3)\,, \tag{4.9}$$

where $\alpha_a(\Phi_3) + \alpha_b(\Phi_3) = 1$. In the absence of leading-logarithmic information, any choice of α_a and α_b is equally valid, however in the presence of leading-logarithmic information the α_i's have to be chosen to correctly treat logarithms of t_a/t_b.

With eq. (4.7) in hand, we can now try to find the best definitions for the partonic cross sections $\mathrm{d}\sigma_i$, but first we have to review the relevant QCD calculations.

4.3 Relevant QCD calculations

The process $e^+e^- \to q\bar{q}g$ is singular as $t_i \to 0$. To understand this singularity structure, consider the differential cross section in $d = 4 - 2\epsilon$ dimensions, and integrated over the three angular variables Ω_2^T and ϕ. One finds the well-known result [79]

$$\frac{\mathrm{d}\sigma_B}{\mathrm{d}t_a \mathrm{d}t_b} = B\,\delta(t_a)\,\delta(t_b)\,,$$

JHEP12(2008)010

$$\frac{\mathrm{d}\sigma_V}{\mathrm{d}t_a \mathrm{d}t_b} = \frac{\alpha_s C_F}{2\pi}\left(\frac{4\pi\mu_\epsilon^2}{E_{\mathrm{CM}}^2}\right)^\epsilon B\left(-\frac{2}{\epsilon^2}-\frac{3}{\epsilon}+V\right)\delta(t_a)\,\delta(t_b)\,,$$
$$\frac{\mathrm{d}\sigma_R}{\mathrm{d}t_a \mathrm{d}t_b} = \frac{\alpha_s C_F}{2\pi}\left(\frac{4\pi\mu_\epsilon^2}{E_{\mathrm{CM}}^2}\right)^\epsilon \frac{1-2\epsilon}{\Gamma(2-2\epsilon)}\,B\,\frac{R(t_a,t_b)-2\epsilon(t_a+t_b)^2}{(1-t_a-t_b)^{2\epsilon}t_a^{1+2\epsilon}t_b^{1+2\epsilon}}\,, \tag{4.10}$$

where μ_ϵ is the standard renormalization scale of dimensional regularization, and has nothing to do with the scale μ that separates QCD calculations from the parton shower. Here, B denotes the Born cross section

$$B = N_c\,Q_q^2\,\frac{4\pi\alpha_{\mathrm{em}}^2}{3E_{\mathrm{CM}}^2}\,, \tag{4.11}$$

with $N_c = 3$, Q_q the charge of the quark, and α_{em} the fine-structure constant.[16] We have also defined

$$R(t_a,t_b) = (1-t_a)^2+(1-t_b)^2\,, \qquad V = -8+\frac{7\pi^2}{6}\,. \tag{4.12}$$

The virtual diagram only contributes at $t_a = t_b = 0$, but is infrared divergent in 4 dimensions. However, the real emission also diverges as $t_i \to 0$. Integrating these expressions over the allowed phase space $0 \le t_a \le 1-t_b \le 1$ we find for the various contributions to the total cross section

$$\begin{aligned}
\sigma_B &= B\,,\\
\sigma_V &= \frac{\alpha_s C_F}{2\pi}\left(\frac{4\pi\mu_\epsilon^2}{E_{\mathrm{CM}}^2}\right)^\epsilon B\left(-\frac{2}{\epsilon^2}-\frac{3}{\epsilon}+V\right),\\
\sigma_R &= \frac{\alpha_s C_F}{2\pi}\left(\frac{4\pi\mu_\epsilon^2}{E_{\mathrm{CM}}^2}\right)^\epsilon B\left(\frac{2}{\epsilon^2}+\frac{3}{\epsilon}+\frac{19}{2}-\frac{7\pi^2}{6}\right),
\end{aligned} \tag{4.13}$$

such that the infrared $1/\epsilon$ divergences cancel and the total cross section gives the well-known result

$$\sigma_{\mathrm{NLO}} = \sigma_B+\sigma_V+\sigma_R = B\left[1+\frac{\alpha_s C_F}{2\pi}\left(V+\frac{19}{2}-\frac{7\pi^2}{6}\right)\right] = B\left(1+\frac{3}{2}\,\frac{\alpha_s C_F}{2\pi}\right) \tag{4.14}$$

in the $\epsilon \to 0$ limit.

The singularity structure of the real emission in QCD is reproduced by the well-known Altarelli-Parisi [66] splitting function

$$f(t_i,z_i) = \frac{1}{t_i}\,\frac{1+z_i^2}{1-z_i}\,, \tag{4.15}$$

where z_i are defined as the energy fractions

$$z_a = \frac{E_q}{E_q+E_g} = \frac{1-t_b}{1+t_a}\,, \qquad z_b = \frac{E_{\bar{q}}}{E_{\bar{q}}+E_g} = \frac{1-t_a}{1+t_b}\,. \tag{4.16}$$

[16] For simplicity, we are only including $e^+e^- \to \gamma^* \to$ partons. An intermediate Z-boson can easily be included by the appropriate change in the Born cross section B.

JHEP12(2008)010

Rewriting the splitting functions in terms of the variables t_a and t_b, taking into account the relevant Jacobian factors, we can define the analog of the Q functions in the toy example

$$\frac{Q_a(t_a,t_b)}{t_a t_b} = \frac{1}{t_a(t_a+t_b)}\left[1+\left(\frac{1-t_b}{1+t_a}\right)^2\right], \quad \frac{Q_b(t_a,t_b)}{t_a t_b} = \frac{1}{t_b(t_a+t_b)}\left[1+\left(\frac{1-t_a}{1+t_b}\right)^2\right], \tag{4.17}$$

where Q_a is the splitting function for $q \to qg$ and Q_b for $\bar{q} \to \bar{q}g$. One can easily verify that the sum of the two splitting functions reproduces all the singular behavior of the function $R(t_a, t_b)$

$$\frac{R(t_a,t_b) - Q_a(t_a,t_b) - Q_b(t_a,t_b)}{t_a t_b} = \text{finite as } t_i \to 0\,. \tag{4.18}$$

For later convenience, we also define a Sudakov factor

$$\Delta_Q(t_1,t_2) = \exp\left[-\frac{\alpha_s C_F}{2\pi}\int_{t_2}^{t_1} \mathrm{d}t_a \int_0^{1-t_a} \mathrm{d}t_b\, \frac{Q_a(t_a,t_b)}{t_a t_b}\right], \tag{4.19}$$

which is symmetric in $a \leftrightarrow b$. The t_b integration range goes down to zero because this corresponds to the proper integration range in the original z_a-dependent splitting function.

For technical reasons relating to the conservation of four-momentum, the parton shower in the companion paper [1] actually uses the splitting function

$$\frac{\tilde{Q}_a(t_a,t_b)}{t_a t_b} = \frac{t_b}{\sqrt{1-t_a}}\frac{1+z_*^2}{1-z_*}, \qquad z_* = \frac{1}{2}\left(1 - \frac{2t_b + t_a - 1}{\sqrt{1-t_a}}\right), \tag{4.20}$$

and $\tilde{Q}_b(t_a,t_b) = \tilde{Q}_a(t_b,t_a)$. These $\tilde{Q}$ functions are perfectly consistent with the condition of eq. (4.18) as longer as the simultaneous limit $t_a, t_b \to 0$ is avoided. This double-soft divergence leads to extra subleading logarithms in the Sudakov factor. Because we only ever work to leading-logarithmic order, this is not a problem, though the specific distributions obtained using Q vs. $\tilde{Q}$ will of course differ. Eventually, one would want to tune the $\tilde{Q}$ functions to minimize this subleading effect. In addition, one can use the difference as an estimate of subleading-logarithmic effects.

4.4 Partonic cross sections at LO

In analogy with section 3.4, we can define partonic cross sections for the master formula eq. (4.7) using the above QCD calculations. For simplicity, we will largely ignore the dependence on Ω_2^T and ϕ and focus only on t_a and t_b.

If we only had access to a parton shower, then it might distribute events according to:

Parton shower (LL):

$$\begin{aligned}\sigma_2(\hat{\mu}^2) &= B\left\{\Delta_Q^2(1,\hat{\mu}^2)\right. \\ &\qquad \left. + \frac{\alpha_s C_F}{2\pi}\left[\int_{\tilde{\mathbf{2}}_\mathbf{a}} \mathrm{d}t_a \mathrm{d}t_b \frac{Q_a(t_a,t_b)}{t_a t_b}\Delta_Q^2(1,t_a) + \int_{\tilde{\mathbf{2}}_\mathbf{b}} \mathrm{d}t_a \mathrm{d}t_b \frac{Q_b(t_a,t_b)}{t_a t_b}\Delta_Q^2(1,t_b)\right]\right\}, \\ \frac{\mathrm{d}\sigma_{3a}(t_a)}{\mathrm{d}t_a \mathrm{d}t_b} &= \frac{\alpha_s C_F}{2\pi}\, B\, \frac{Q_a(t_a,t_b)}{t_a t_b}\, \Delta_Q^2(1,t_a)\,,\end{aligned}$$

JHEP12(2008)010

$$\frac{\mathrm{d}\sigma_{3b}(t_b)}{\mathrm{d}t_a \mathrm{d}t_b} = \frac{\alpha_s C_F}{2\pi} B \frac{Q_b(t_a, t_b)}{t_a t_b} \Delta_Q^2(1, t_b) \,. \tag{4.21}$$

The details of the parton shower algorithm that yield this kind of distribution are given in the companion paper [1], though we note that this shower treats the gluon emission from the quark and antiquark symmetrically. The fact that the Sudakov factors are always squared comes from the fact that in a symmetric shower, the no-branching probability for the quark and the antiquark are tied together.

For the present purposes, the most relevant information about eq. (4.21) is that these functions include the correct leading-logarithmic resummation through the Sudakov factors while reproducing the Born cross section:

$$\sigma_2(\hat{\mu}^2) + \int_{\mathbf{3}} \mathrm{d}t_a \, \mathrm{d}t_b \, \frac{\mathrm{d}\sigma_{3a}(t_a)}{\mathrm{d}t_a \mathrm{d}t_b} + \int_{\mathbf{3}} \mathrm{d}t_a \, \mathrm{d}t_b \, \frac{\mathrm{d}\sigma_{3b}(t_b)}{\mathrm{d}t_a \mathrm{d}t_b} = B \,. \tag{4.22}$$

The integrations over the $\tilde{\mathbf{2}}$ regions in σ_2 are necessary to get the total cross section correct, and correspond to the fact that in the shower, a gluon can split off at such a large angle from the quark that it becomes singular with the antiquark and vice verse, effectively yielding 2-jet events, as shown in figure 14.

If we now have the tree-level real-emission calculation and ignore leading logarithms, then the most naive partonic cross section is

Tree level (LO):

$$\begin{aligned} \sigma_2(\hat{\mu}^2) &= B \,, \\ \frac{\mathrm{d}\sigma_3}{\mathrm{d}t_a \mathrm{d}t_b} &= \frac{\alpha_s C_F}{2\pi} B \frac{R(t_a, t_b)}{t_a t_b} \,. \end{aligned} \tag{4.23}$$

At this point, we do not have enough information to define separate $\mathrm{d}\sigma_{3a}$ and $\mathrm{d}\sigma_{3b}$ cross sections and will rely on some unspecified α_i function to determine the relative weight of the two different dMC_3 samples, as in eq. (4.8).

While eq. (4.23) gives the correct leading order behavior for 3-jet differential distributions, the total cross section has a strong dependence on the unphysical matching scale, which can be seen by expanding the resulting expression for the total cross section around small values of $\hat{\mu}^2$. This gives

$$\sigma = B \left[1 + \frac{\alpha_s C_F}{2\pi} \left(4 \log^2 \hat{\mu} + 6 \log \hat{\mu} + \frac{5}{2} - \frac{\pi^2}{3} \right) + \mathcal{O}(\hat{\mu}^2) \right] . \tag{4.24}$$

This double-logarithmic dependence on the matching scale is of course due to having used a fixed-order calculation, which by construction does not sum any of the leading-logarithmic behavior in the cross section.

The most precise way to determine the properly resummed expressions is to use renormalization group evolution in an effective field theory setup to determine the partonic cross sections. However, one can use the fact that parton showers do sum the leading-logarithmic behavior correctly to derive expressions at this order. Just as in the toy model, we can Sudakov-improve the matrix element by finding a suitable merging of eqs. (4.21) and (4.23).

JHEP12(2008)010

Since there are two event samples describing the single gluon emission, care has to be taken to minimize the effect of large $\log^2(t_a/t_b)$ logarithms. The simplest way to do so is to find an expression that reduces to eq. (4.21) if we were to take $R(t_a,t_b) = Q_a(t_a,t_b) + Q_b(t_a,t_b)$, which guarantees the Sudakov-improved result will have the same singularity structure as the leading-logarithmic result. We choose

Sudakov-improved (LO/LL):

$$\begin{aligned}
\sigma_2(\hat{\mu}^2) &= B\left\{\Delta_Q^2(1,\hat{\mu}^2) + \frac{\alpha_s C_F}{2\pi}\left[\int_{\tilde{\mathbf{2}}_\mathbf{a}} \mathrm{d}t_a\,\mathrm{d}t_b\,\frac{Q_a}{t_a t_b}\,\Delta_Q^2(1,t_a) + \int_{\tilde{\mathbf{2}}_\mathbf{b}} \mathrm{d}t_a\,\mathrm{d}t_b\,\frac{Q_b}{t_a t_b}\,\Delta_Q^2(1,t_b)\right]\right\},\\
\frac{\mathrm{d}\sigma_{3a}(t_a)}{\mathrm{d}t_a\mathrm{d}t_b} &= \frac{\alpha_s C_F}{2\pi}\,B\,\frac{R(t_a,t_b)}{t_a t_b}\left[\frac{Q_a}{Q_a+Q_b}\,\Delta_Q^2(1,t_a)\right],\\
\frac{\mathrm{d}\sigma_{3b}(t_b)}{\mathrm{d}t_a\mathrm{d}t_b} &= \frac{\alpha_s C_F}{2\pi}\,B\,\frac{R(t_a,t_b)}{t_a t_b}\left[\frac{Q_b}{Q_a+Q_b}\,\Delta_Q^2(1,t_b)\right],
\end{aligned} \tag{4.25}$$

where we have omitted the dependence of the splitting functions on the phase space variables. The Sudakov factors in $\mathrm{d}\sigma_3$ are weighted by the relative splitting probabilities of their respective shower history, which is equivalent to the approach of ref. [24]. By integrating these expressions and expanding around small values of $\hat{\mu}^2$, one can easily verify that the leading-logarithmic dependence on $\hat{\mu}^2$ has vanished,[17] and to leading order in α_s one finds

$$\sigma = B\left[1 + \frac{\alpha_s C_F}{4\pi}(5 - 12\log 2) + \mathcal{O}(\alpha_s^2)\right]. \tag{4.26}$$

Note that the Sudakov-improved and tree-level results for $\mathrm{d}\sigma_3$ agree to leading order in α_s.

Eq. (4.25) is by no means unique. For example, the dominant shower history used in the CKKW prescription [23] is morally equivalent to

Dominant history Sudakov-improved (LO/LL):

$$\begin{aligned}
\sigma_2(\hat{\mu}^2) &= B\,\Delta_Q^2(1,\hat{\mu}^2)\,,\\
\frac{\mathrm{d}\sigma_{3a}(t_a)}{\mathrm{d}t_a\mathrm{d}t_b} &= \frac{\alpha_s C_F}{2\pi}\,B\,\frac{R(t_a,t_b)}{t_a t_b}\,\theta(t_b > t_a)\,\Delta_Q^2(1,t_a)\,,\\
\frac{\mathrm{d}\sigma_{3b}(t_b)}{\mathrm{d}t_a\mathrm{d}t_b} &= \frac{\alpha_s C_F}{2\pi}\,B\,\frac{R(t_a,t_b)}{t_a t_b}\,\theta(t_a > t_b)\,\Delta_Q^2(1,t_b)\,.
\end{aligned} \tag{4.27}$$

The reason that this expressions works is that

$$\frac{R(t_a,t_b)\,\theta(t_b > t_a) - Q_a(t_a,t_b)}{t_a t_b} = \text{finite as } t_a \to 0\,. \tag{4.28}$$

The lack of the $\tilde{\mathbf{2}}$ integrations in σ_2 can contribute at most at the subleading-logarithmic level. As expected, there is no leading-logarithmic dependence in the total cross section for the dominant history method[18]

$$\sigma = B\left[1 + \frac{\alpha_s C_F}{4\pi}\left(5 - \frac{2\pi^2}{3} - 12\log 2\right) + \mathcal{O}(\alpha_s^2)\right]. \tag{4.29}$$

[17] The fact that the subleading single-logarithmic terms $\alpha_s \log\hat{\mu}$ vanished as well is an accident, and will not in general be true for the $\tilde{Q}$ splitting functions or for gluon splitting functions.

[18] Again, the lack of subleading single-logarithmic terms $\alpha_s \log\hat{\mu}$ is an accident.

JHEP12(2008)010

4.5 Partonic cross sections at NLO

We now want to include the virtual corrections to the 2-jet rate to obtain partonic expressions that give both differential distributions as well as the total cross section correct to $\mathcal{O}(\alpha_s)$. As in the toy model, we will slowly build up towards an answer that has all the desired properties. Ignoring the leading logarithmic dependence of the partonic cross sections, we can use an NLO slicing method [73 – 78]

NLO slicing (NLO):

$$\begin{aligned}\sigma_2(\hat{\mu}^2) &= \tilde{B}(\hat{\mu}^2)\,,\\ \frac{\mathrm{d}\sigma_3}{\mathrm{d}t_a \mathrm{d}t_b} &= \frac{\alpha_s C_F}{2\pi}\, B\, \frac{R(t_a,t_b)}{t_a t_b}\,,\end{aligned} \tag{4.30}$$

where

$$\tilde{B}(\hat{\mu}^2) = B\left[1 + \frac{\alpha_s C_F}{2\pi}\lim_{\epsilon\to 0}\left(-\frac{2}{\epsilon^2} - \frac{3}{\epsilon} + V + \int_{\mathbf{2}+\tilde{\mathbf{2}}_\mathbf{a}+\tilde{\mathbf{2}}_\mathbf{b}} \mathrm{d}t_a\,\mathrm{d}t_b\,\frac{R(t_a,t_b)}{t_a^{1+2\epsilon}\,t_b^{1+2\epsilon}}\right)\right]. \tag{4.31}$$

Note the cancellation between the terms that are divergent as $\epsilon \to 0$, such that $\tilde{B}$ is a finite expression, which only depends on $\hat{\mu}^2$. It is a simple exercise to show that the total cross section reproduces the NLO cross section in eq. (4.14). As for the LO calculation, we do not have enough information to separately determine $\mathrm{d}\sigma_{3a}$ and $\mathrm{d}\sigma_{3b}$ at this point.

While eq. (4.30) does reproduce both the differential and total cross sections correctly to $\mathcal{O}(\alpha_s)$, it does not sum logarithms and has large $\hat{\mu}^2$ dependence. The leading logarithmic resummation can be included employing a subtraction method similar to MC@NLO [44], and we find

NLO subtraction (NLO/LL):

$$\begin{aligned}\sigma_2(\hat{\mu}^2) &= \bar{B}(\hat{\mu}^2)\Bigg\{\Delta_Q^2(1,\hat{\mu}^2)\\ &\qquad + \frac{\alpha_s C_F}{2\pi}\left[\int_{\tilde{\mathbf{2}}_\mathbf{a}} \mathrm{d}t_a \mathrm{d}t_b \frac{Q_a(t_a,t_b)}{t_a t_b}\Delta_Q^2(1,t_a) + \int_{\tilde{\mathbf{2}}_\mathbf{b}} \mathrm{d}t_a \mathrm{d}t_b \frac{Q_b(t_a,t_b)}{t_a t_b}\Delta_Q^2(1,t_b)\right]\Bigg\}\,,\\ \frac{\mathrm{d}\sigma_3}{\mathrm{d}t_a \mathrm{d}t_b} &= \frac{\alpha_s C_F}{2\pi}\Bigg[B\,\frac{R(t_a,t_b) - Q_a(t_a,t_b) - Q_b(t_a,t_b)}{t_a,t_b}\\ &\qquad + \bar{B}(\hat{\mu}^2)\left(\frac{Q_a(t_a,t_b)}{t_a t_b}\,\Delta_Q^2(1,t_a) + \frac{Q_b(t_a,t_b)}{t_a t_b}\,\Delta_Q^2(1,t_b)\right)\Bigg]\,,\end{aligned} \tag{4.32}$$

where

$$\bar{B}(\hat{\mu}^2) = B\left[1 + \frac{\alpha_s C_F}{2\pi}\left(\frac{3}{2} - \int_{\mathbf{3}} \mathrm{d}t_a\,\mathrm{d}t_b,\,\frac{R(t_a,t_b) - Q_a(t_a,t_b) - Q_b(t_a,t_b)}{t_a t_b}\right)\right]. \tag{4.33}$$

Again, this reproduces the correct differential 3-jet distributions at $\mathcal{O}(\alpha_s)$ as well as the total NLO cross section of eq. (4.14). Eq. (4.32) only defines $\mathrm{d}\sigma_3/(\mathrm{d}t_a \mathrm{d}t_b)$, but not the individual $\mathrm{d}\sigma_{3i}/(\mathrm{d}t_a \mathrm{d}t_b)$. To avoid large $\log^2(t_a/t_b)$ logarithms, we have many options. The simplest is to use eq. (4.9) to define $\mathrm{d}\sigma_{3a}$ and $\mathrm{d}\sigma_{3b}$ via

$$\alpha_a = \frac{Q_a}{Q_a + Q_b}, \qquad \alpha_b = \frac{Q_b}{Q_a + Q_b}\,. \tag{4.34}$$

JHEP12(2008)010

To see that this works, imagine that $R = Q_a + Q_b$, and note that

$$\frac{Q_a}{Q_a + Q_b} \to 1 \text{ as } t_a \to 0\,, \tag{4.35}$$

which shows that the singularity structure will correspond to the right leading-logarithmic behavior. In fact, eq. (4.34) is a general strategy for minimizing the effect of leading-logarithms between intermediate scales, and we will use a variant of this approach in the next section.

As was true in the case of the toy model, the particular form of the NLO/LL partonic cross section in eq. (4.32) is not unique. The criteria we wanted to satisfy is the proper resummation of the leading double-logarithms while simultaneously reproducing all observables to $\mathcal{O}(\alpha_s)$ accuracy. As in the toy model, we can get a simple form for an NLO/LL partonic cross sections in analogy to ref. [46]:

NLO elegant (NLO/LL):

$$\begin{aligned}\sigma_2(\hat{\mu}^2) &= \sigma_{\text{NLO}}\,\Delta_T(\hat{\mu}^2)\,,\\ \frac{\mathrm{d}\sigma_3}{\mathrm{d}t_a \mathrm{d}t_b} &= \frac{\alpha_s C_F}{2\pi}\,\sigma_{\text{NLO}}\,\frac{T(t_a,t_b)}{t_a t_b}\,\Delta_T[\min(t_a,t_b)]\,,\end{aligned} \tag{4.36}$$

where σ_{NLO} is given in eq. (4.14). The function $T(t_a, t_b)$ is defined by

$$T(t_a,t_b) = \frac{B}{\sigma_{\text{NLO}}}\,R(t_a,t_b)\,, \tag{4.37}$$

and has an accompanying "Sudakov factor"

$$\Delta_T(t) = \exp\left[-\frac{\alpha_s C_F}{2\pi}\int_{\mathbf{3}(t)} \mathrm{d}t_a\,\mathrm{d}t_b\,\frac{T(t_a,t_b)}{t_a t_b}\right], \tag{4.38}$$

where $\mathbf{3}(t) \equiv \mathrm{d}\Phi_3(t)$ denotes the region of phase space with $t_a, t_b > t$. Note that $\Delta_T(t)$ has the same double-logarithmic dependence as $\Delta_Q^2(1,t)$, ensuring that the leading-logarithmic dependence is reproduced correctly. To order α_s, we are also clearly reproducing both the differential spectrum and the total rate. As before, $\mathrm{d}\sigma_{3a}$ and $\mathrm{d}\sigma_{3b}$ can be defined using eqs. (4.9) and (4.34) to minimize the effect of $\log^2(t_a/t_b)$ logarithms. Eq. (4.36) will be the cornerstone for the NLO/LO/LL result we will present in the next section.

In our entire discussion above, we never talked about angular dependence. To keep the discussion simple, we have implicitly assumed that the angular dependence factorizes, but we do have to be mindful that in general, operator mixing can break this factorization. In any case, we can easily generalize eq. (4.36) to have angular dependence by

NLO elegant with angles (NLO/LL):

$$\begin{aligned}\frac{\mathrm{d}\sigma_2(\hat{\mu}^2)}{d\Omega_2^T} &= \sigma_{\text{NLO}}\,f(\Omega_2^T)\,\Delta_T(\hat{\mu}^2)\,,\\ \frac{\mathrm{d}\sigma_3}{d\Omega_2^T \mathrm{d}t_a \mathrm{d}t_b d\phi} &= \frac{\alpha_s C_F}{2\pi}\,B\,\frac{R(\Omega_2^T,t_a,t_b,\phi)}{t_a t_b}\,\Delta_T[\min(t_a,t_b)]\,,\end{aligned} \tag{4.39}$$

JHEP12(2008)010

where the normalized $f(\Omega_2^T)$ function contains angular information about 2-jet distributions and we have introduced the generalized real emission function $R(t_a, t_b) \to R(\Omega_2^T, t_a, t_b, \phi)$. There are two different ways of defining the function T that appears in the Sudakov factor,

$$T(\Omega_2^T, t_a, t_b, \phi) = \frac{B}{\sigma_{\text{NLO}}} R(\Omega_2^T, t_a, t_b, \phi) \qquad \text{or} \qquad \frac{B}{\sigma_{\text{NLO}}} \frac{R(\Omega_2^T, t_a, t_b, \phi)}{f(\Omega_2^T)} . \tag{4.40}$$

In either case, the Sudakov factor is given by

$$\Delta_T(t) = \exp\left[-\frac{\alpha_s C_F}{2\pi} \int_{\mathbf{3}(t)} \mathrm{d}t_a \, \mathrm{d}t_b \int \mathrm{d}\Omega_2^T \, \mathrm{d}\phi \, \frac{T(\Omega_2^T, t_a, t_b, \phi)}{t_a t_b} \right] . \tag{4.41}$$

The first choice is easier to implement and is the strategy adopted in this paper. The second option is likely to be more correct, but the difference can at most be at the subleading-logarithmic level.

5. NLO/LO/LL merging

For the first emission in QCD, we could understand everything analytically, but the calculations quickly become unwieldy with more than one emission. Both for simplicity of discussion and simplicity of the `GenEvA` strategy, it is necessary to define more generally how to construct an LO/LL sample and how to merge together nested descriptions of QCD.

As we saw in the previous section, in order to properly treat the double-logarithms of t_a/t_b, we needed two event samples, $\mathrm{dMC}_3(t_a)$ and $\mathrm{dMC}_3(t_b)$. Since the parton shower has two possible histories to generate the 3-parton final state ($q \to qg$ or $\bar{q} \to \bar{q}g$ splitting), these two event samples can be interpreted as representing these two possible shower histories. This will generalize for more partonic final states, and we will have a master formula which contains different event samples for each possible shower history, in the process giving an unambiguous definition for the phase space projection in eq. (2.7). This will allow us to write the NLO/LO/LL sample by combining the NLO/LL, LO/LL, and LL descriptions, just as in the toy model.

5.1 Generalized master formula

To build the generalized master formula, we first introduce the concept of a would-be shower history. Though phase space with a matching scale does not require us to ever talk about parton shower histories, they are convenient for resolving logarithmic ambiguities since shower histories track QCD singularities, so one should really think of these histories as "singularity histories". All of the possible leading-logarithmic ambiguities can be captured by introducing a separate event sample for every kind of allowed shower (or singularity) history. This gives the generalized master formula:

$$\mathrm{d}\sigma = \sum_{n=2}^{n_{\max}} \sum_j \frac{\mathrm{d}\sigma_{nj}(\mu_{nj})}{\mathrm{d}\Phi_n} \, \mathrm{dMC}_n(\mu_{nj}) , \tag{5.1}$$

where j labels different shower histories for n final states, and we start running the phenomenological shower at the scale μ_{nj} for each sample. Note that, similarly to eq. (4.7),

JHEP12(2008)010

eq. (5.1) does not introduce double-counting, but merely reflects the scale ambiguity of QCD, and we anticipate defining $\mathrm{d}\sigma_{nj}$ in analogy with eq. (4.8).

As anticipated in section 2.3, $\mathrm{dMC}_n(\mu_{nj})$ is defined by excising regions of n-body phase space that are covered by running the phenomenological model on $(m < n)$-body phase space starting from the various μ_{mj} scales. With so many matching scales, it is convenient to choose μ_{nj} equal to a common scale μ, except for $\mu_{n_{\max}j}$, which has to be equal to a scale "t_j" in order to cover all of phase space without $\alpha_s \log^2(t_j/t_k)$ ambiguities. Because scale-dependent n-body phase space does not depend on μ_n, there is no need for special treatment of the $n_{\max}$-body phase space. Assuming that the phenomenological model is a virtuality-ordered shower, the excised phase space is

$$\mathrm{d}\Phi_n(\mu) = \mathrm{d}\Phi_n(p_{\mathrm{CM}}; p_1, p_2, \ldots, p_n) \prod_{i \diamond j} \theta\left[(p_i + p_j)^2 > \mu^2\right], \tag{5.2}$$

where $i \diamond j$ indicates pairs of partons that are associated with a QCD singularity.

Since we are choosing all of the $\mu_{n<n_{\max}}$ scales to be the same, the master formula in eq. (5.1) is highly redundant for $n < n_{\max}$, because many of the $\mathrm{dMC}_n(\mu_{nj})$ event samples are multiply covered. Despite this redundancy, keeping the notion of a would-be shower history is convenient for the purposes of having a definition of phase space projection as in eq. (2.7). That is, a shower history defines a map between n-body phase space and a lower-dimensional m-body phase space

$$\{\Phi_n, (nj)\} \to \{\Phi_m, (mk)\} \qquad \text{for } m < n\,, \tag{5.3}$$

where the specifics of this map are determined by the details of a showering or clustering procedure. Note that this map does not necessarily give a unique map $\Phi_n \to \Phi_m$, but rather gives a series of different consistent maps depending which shower history j is being selected. One could choose to implement a unique map $\Phi_n \to \Phi_m$ by only considering one type of shower history for a given phase space point.

Given a phase space point with a would-be shower history $\{\Phi_n, (nj)\}$, we can always define the scale t_j that we would need to start the phenomenological shower from to eliminate leading-logarithmic dependence. To do so, we simply ask at what scale the phenomenological shower would have generated the n-th branching. This is the scale we use to define $\mu_{n_{\max}j}$, and it guarantees that there are no phase space dead zones.

5.2 Sudakov improvements

We now wish to extend the LO/LL result of eq. (4.25) by adding additional tree-level emissions. As discussed above, an LO/LL result is obtained by supplementing a tree-level matrix element with appropriate Sudakov factors. That is, we wish to combine the n-body tree-level matrix element (LO)

$$\frac{\mathrm{d}\sigma_n^{\mathrm{LO}}}{\mathrm{d}\Phi_n} \tag{5.4}$$

that has no notion of a matching scale μ, with Sudakov factors taken from the equivalent partonic cross section for a set of n-body parton shower histories (LL)

$$\frac{\mathrm{d}\sigma_{nj}^{\mathrm{LL}}(\mu_{nj})}{\mathrm{d}\Phi_n}, \tag{5.5}$$

JHEP12(2008)010

where j labels the different shower histories for n final states.

Note that schematically, the expression for the LL "calculation" can be written as a product of splitting functions and Sudakov factors,

$$\frac{\mathrm{d}\sigma_{nj}^{\mathrm{LL}}(\mu_{nj})}{\mathrm{d}\Phi_n} = B \prod_{r=1}^{n-2} Q_{j,r} \prod_{s=1}^{2n-2} \Delta_{j,s} \equiv B\, Q_j\, \Delta_j(\mu_{nj})\,, \tag{5.6}$$

where B is the initial hard scattering matrix element, r labels the different $1 \to 2$ splitting function vertices, and s labels the partons in the shower history. For convenience, we have introduced the notation

$$Q_j = \prod_{r=1}^{n-2} Q_{j,r}\,, \qquad \Delta_j(\mu_{nj}) = \prod_{s=1}^{2n-2} \Delta_{j,s}\,, \tag{5.7}$$

such that for the j-th shower history, Q_j is the product of all splitting functions and $\Delta_j(\mu_{nj})$ is the product of all Sudakov factors. Since the parton shower correctly reproduces the singularity structure of QCD, the tree-level calculation should share the same singularity structure as the sum over all shower histories

$$\lim_{\Phi_n \to \mathrm{sing.}} \left[\frac{\mathrm{d}\sigma_n^{\mathrm{LO}}}{\mathrm{d}\Phi_n} - B \sum_i Q_i \right] = \text{finite}. \tag{5.8}$$

Accounting for these singularities, one possible Sudakov-improved partonic cross section is

Sudakov-improved (LO/LL):

$$\frac{\mathrm{d}\sigma_{nj}^{\mathrm{LO/LL}}(\mu_{nj})}{\mathrm{d}\Phi_n} = \frac{\mathrm{d}\sigma_n^{\mathrm{LO}}}{\mathrm{d}\Phi_n} \left(B \sum_i Q_i \right)^{-1} \frac{\mathrm{d}\sigma_{nj}^{\mathrm{LL}}(\mu_{nj})}{\mathrm{d}\Phi_n}\,. \tag{5.9}$$

Expanding this result to leading order in α_s, all of the Δ factors in $\mathrm{d}\sigma_{nj}^{\mathrm{LL}}$ are equal to unity, so summing over all of the $\mathrm{d}\sigma_{nj}^{\mathrm{LO/LL}}$ expressions yields the tree-level result as desired. By construction, this answer has the correct leading-logarithmic behavior because it has the same singularity structure as the parton shower.

Eq. (5.9) is a correct LO/LL answer, but there are various choices we can make for $\mathrm{d}\sigma_{nj}^{\mathrm{LL}}$. The first is to use the naive LL cross section from eq. (5.6), which includes the sum over all possible shower histories. One can simplify this result, by following the path taken by CKKW [23], and use the dominant shower history, defined as the one with the largest value of Q_j. In QCD, there is never an ambiguity as to which shower history is dominant, and the dominant shower history always reproduces the QCD singularities. Thus,

$$\frac{Q_{\mathrm{dom}}}{\sum_j Q_j} \to 1\,, \qquad \frac{Q_{\mathrm{other}}}{\sum_j Q_j} \to 0\,. \tag{5.10}$$

Taking these limits as equalities, the CKKW procedure essentially defines:

JHEP12(2008)010

Dominant history Sudakov-improved (LO/LL):

$$\frac{\mathrm{d}\sigma_{nj}^{\mathrm{LO/LL}}(\mu_{nj})}{\mathrm{d}\Phi_n} = \begin{cases} \dfrac{\mathrm{d}\sigma_n^{\mathrm{LO}}}{\mathrm{d}\Phi_n}\Delta_j(\mu_{nj}) & j = \mathrm{dom}, \\ 0 & j = \mathrm{other}. \end{cases} \tag{5.11}$$

Eqs. (5.9) and (5.11) share the same singularity structure, so they are both valid LO/LL results.

A different route is to try to improve on eq. (5.6). As we saw in eq. (4.21), there are additional $\tilde{\mathbf{n}}$ integrations that are necessary to get the total cross section correct in the parton shower. From the shower point of view, these integrations correspond to situations where the parton shower populates a region of phase space that has QCD singularities but uses a splitting function away from the singular region. That is, two partons are "accidentally" distributed closer than the cutoff of the shower. Schematically, we can write this as

$$\frac{\mathrm{d}\sigma_{nj}^{\mathrm{LL}}(\mu_{nj})}{\mathrm{d}\Phi_n} = B\left(Q_j\,\Delta_j + \int_{\sum\tilde{\mathbf{n}}_{\mathbf{k}}} Q_k\,\Delta_k\right). \tag{5.12}$$

These $\tilde{\mathbf{n}}$ regions do not exist as separate event samples, but their effect should be captured in the n-body event sample.

At this point, it is not at all clear how to figure out which $\tilde{\mathbf{n}}_{\mathbf{k}}$ integrations correspond to which shower histories j. At the leading-logarithmic level, these $\tilde{\mathbf{n}}$ integrations can simply be ignored because they are not associated with the dominant QCD singularities and give at most subleading-logarithmic corrections. However, for high enough multiplicity, the phase space volume of all the $\tilde{\mathbf{n}}$ regions is rather large, especially for moderate μ, so numerically it is dangerous to ignore them. On the other hand, if the only reason we are keeping the $\tilde{\mathbf{n}}$ integrations is for overall normalization, then it would be simpler to just introduce *ad hoc* k-factors to restore the normalization.

As discussed in the companion paper [1], there is a numerically efficient way to accomplish the $\tilde{\mathbf{n}}$ integrations, so the user can address the $\tilde{\mathbf{n}}$ subtleties if desired. In particular, the `GenEvA` algorithm has a way to figure out which "wrong" histories k should be associated with which "right" history j, and it does this in a way that is reasonably efficient. Because the resulting differential distributions using eqs. (5.6) and (5.12) are different, the two different choices can be used to test for systematics in the LO/LL sample. Note that for $n = 3$, using eq. (5.9) with eq. (5.12) reproduces the result of eq. (4.25). To show our results in section 6, we will always use eq. (5.12) because of the reduced cross section scale dependence.

5.3 Merging nested descriptions

In the case of LO/LL merging in eq. (5.9) or NLO/LL merging in eq. (4.36), we were trying to merge together two different descriptions of the physics that both give valid predictions at some scale μ. More generally, we are interested in cases like in figure 13 where there is one description that gives a better description of the physics at some high scale and one that gives a better description at a lower scale. To build an NLO/LO/LL merged sample,

JHEP12(2008)010

we want to use an NLO/LL result at high energies, but supplement it with additional LO/LL information for subsequent emissions.

From the toy model in eq. (3.39) we learned that the NLO/LO/LL result for the partonic results with more than three partons are given by the LO/LL result, except that the first Sudakov factor changed to Δ_T. How do we obtain this new Sudakov factor, which depends on the kinematics of the first emissions? Consider the differences between the $\mathrm{d}\sigma_3$ expressions for the LO/LL and NLO/LL merged samples in eqs. (4.25) and (4.36). We will use the convention of eq. (4.34) to define $\sigma_{3a}^{\mathrm{NLO/LL}}$ for the NLO/LL result. The ratio of the two answers is

$$\frac{\mathrm{d}\sigma_{3a}^{\mathrm{NLO/LL}}(t_a)}{\mathrm{d}t_a \mathrm{d}t_b} \Bigg/ \frac{\mathrm{d}\sigma_{3a}^{\mathrm{LO/LL}}(t_a)}{\mathrm{d}t_a \mathrm{d}t_b} = \frac{\Delta_T[\min(t_a, t_b)]}{\Delta_Q^2(1, t_a)}\,, \tag{5.13}$$

and therefore can be used to extract the Sudakov factor Δ_T from the NLO/LL calculation.[19]

While eq. (5.13) is a function of 3-body phase space, we want to apply this correction factor to n-body matrix elements. This is straightforward, using the map from eq. (5.3) to define the map

$$\{\Phi_n, (nj)\} \to \{\Phi_3, (3k)\} \qquad \text{for } n > 3\,. \tag{5.14}$$

Note that this implies that the specific value of the correction factor will depend on the particular would-be shower history and on the details of the $\Phi_n \to \Phi_3$ map. By assumption, the map in eq. (5.3) respects QCD singularities, so any differences are formally beyond the order we are working to, but different choices will affect the specifics of the NLO/LO/LL merging.

Putting these pieces together, we find an expression for the best partonic calculation that `GenEvA` can currently implement

GenEvA best (NLO/LO/LL):

$$\begin{aligned}
\frac{\mathrm{d}\sigma_2^{\mathrm{NLO/LO/LL}}(\mu_2)}{\mathrm{d}\Phi_2} &= \frac{\mathrm{d}\sigma_2^{\mathrm{NLO/LL}}(\mu_2)}{\mathrm{d}\Phi_2}\,, \\
\frac{\mathrm{d}\sigma_{3j}^{\mathrm{NLO/LO/LL}}(\mu_{3j})}{\mathrm{d}\Phi_3} &= \frac{\mathrm{d}\sigma_{3j}^{\mathrm{LO/LL}}(\mu_{3j})}{\mathrm{d}\Phi_3} \times \frac{\mathrm{d}\sigma_{3j}^{\mathrm{NLO/LL}}(t_j)}{\mathrm{d}\Phi_3} \Bigg/ \frac{\mathrm{d}\sigma_{3j}^{\mathrm{LO/LL}}(t_j)}{\mathrm{d}\Phi_3}\,, \\
&\vdots \\
\frac{\mathrm{d}\sigma_{nj}^{\mathrm{NLO/LO/LL}}(\mu_{nj})}{\mathrm{d}\Phi_n} &= \frac{\mathrm{d}\sigma_{nj}^{\mathrm{LO/LL}}(\mu_{nj})}{\mathrm{d}\Phi_n} \times \frac{\mathrm{d}\sigma_{3k}^{\mathrm{NLO/LL}}(t_k)}{\mathrm{d}\Phi_3} \Bigg/ \frac{\mathrm{d}\sigma_{3k}^{\mathrm{LO/LL}}(t_k)}{\mathrm{d}\Phi_3}\,,
\end{aligned} \tag{5.15}$$

where the value of $n_{\max}$ is determined by the availability of high-multiplicity tree-level matrix elements, and $(3k)$ is determined uniquely from (nj) by eq. (5.14). Note that eq. (5.13) is still needed to define $\mathrm{d}\sigma_{3j}$. This is analogous to eq. (3.39), where an additional Sudakov factor Δ_Q was needed once extra emissions were added. The reason is, that the NLO/LL sample had $n_{\max} = 3$ and therefore took $\mu_{3j=t_j}$, whereas in general the matching scale μ_{3j} will be lower than t_j. Taking $n_{\max} = 3$ and $\mu_{3j} = t_j$, eq. (5.15) reduces to eq. (4.36).

[19] The fact that the numerator of eq. (5.13) is a function of $\min(t_a, t_b)$ instead of t_a means that there are power suppressed logarithmic ambiguities of the form $(t_a/t_b)\log(t_a/t_b)$, which are beyond the order we are working.

JHEP12(2008)010

As in the toy model, eq. (5.15) is a special case of a more general construction to merge together nested descriptions of QCD. In analogy with eq. (3.39), if we have partonic descriptions **A**, **B**, **C**, ..., then these can be consistently merged via

Best combination (A/B/C/$\cdots$):

$$\frac{\mathrm{d}\sigma_{nj}^{\mathrm{A/B/C/\cdots}}(\mu_{nj})}{\mathrm{d}\Phi_n} = \frac{\mathrm{d}\sigma_{nj}^{\mathrm{A}}(\mu_{nj})}{\mathrm{d}\Phi_n} \times \left[\frac{\mathrm{d}\sigma_{n'j'}^{\mathrm{B}}(t_{j'})}{\mathrm{d}\Phi_{n'}} \middle/ \frac{\mathrm{d}\sigma_{n'j'}^{\mathrm{A}}(t_{j'})}{\mathrm{d}\Phi_{n'}}\right] \times \left[\frac{\mathrm{d}\sigma_{n''j''}^{\mathrm{C}}(t_{j''})}{\mathrm{d}\Phi_{n''}} \middle/ \frac{\mathrm{d}\sigma_{n''j''}^{\mathrm{B}}(t_{j''})}{\mathrm{d}\Phi_{n''}}\right] \times \cdots . \tag{5.16}$$

This merging procedure does not spare the user from the need to first merge N^{i}LO and N^{j}LL descriptions that occupy the same n-body phase space. Rather, eq. (5.16) gives the user one simple option how to supplement low-scale/high-multiplicity calculations with additional high-scale/low-multiplicity information.

6. Results

In this section, we present the results obtained using the `GenEvA` program, which implements the `GenEvA` framework, as explained in section 2.6. As such, it uses the `GenEvA` algorithm to generate Lorentz-invariant phase space, and then reweights the resulting events to the distributions discussed in this work. The current implementation of `GenEvA` is capable of describing $e^+e^- \to n$ jets, but only describes massless quarks adequately. Therefore, we do not present results for the production of b-quark jets. A validation of the `GenEvA` event generator is given in the companion paper [1].

For the fixed-order tree-level matrix elements we use the `HELAS` [80] Fortran routines generated by `Madgraph` [81], and `GenEvA` currently utilizes matrix elements with up to $n_{\max} = 6$ final state partons. Final states with more than $n_{\max}$ partons are obtained by a subsequent parton shower, and in general any virtuality-ordered shower could be used for that purpose. For simplicity, here we only use the internal analytic parton shower of the `GenEvA` algorithm, which is a virtuality-ordered shower that neither includes color coherence, α_s running, nor the hadronization of the final state. For this reason, the results presented here should be viewed as a proof of concept, and not as a prediction of the shown distributions. Unless otherwise noted, all results in this section use

$$E_{\mathrm{CM}} = 1000\,\mathrm{GeV}\,, \qquad \mu = 50\,\mathrm{GeV}\,, \qquad \Lambda_{\mathrm{IR}} = 10\,\mathrm{GeV}\,. \tag{6.1}$$

As we will see in figure 15, the matching scale $\mu = 50\,\mathrm{GeV}$ is rather low in the sense that the single-logarithmic dependence that we do not account for becomes important. However, we choose a low scale in order to accentuate the differences between different partonic calculations and leave enough phase space volume available for high-multiplicity partonic states.

The notation for the various event samples used in this section is summarized in table 2. We use the notation LO$_{n_{\max}}$ for tree-level matrix elements with $2 \le n \le n_{\max}$ partons in the final state, and LL indicates that leading-logarithmic Sudakov resummation is being

JHEP12(2008)010

Notation	Given by eq.	Description
$\mathrm{LO}_{n_{\max}}$	(5.4)	Tree-Level (LO)
$\mathrm{LO}_{n_{\max}}$/LL	(5.9) with (5.12)	Sudakov-Improved (LO/LL)
$\mathrm{NLO}_2/\mathrm{LO}_3$/LL	(4.36)	NLO Elegant (NLO/LL)
$\mathrm{NLO}_2/\mathrm{LO}_{n_{\max}}$/LL	(5.15)	`GenEvA` Best (NLO/LO/LL)

Table 2: Notation and meaning for the various event samples. $\mathrm{LO}_{n_{\max}}$ corresponds to tree-level matrix elements with 2- through $n_{\max}$-emissions, and LL indicates the inclusion of leading-logarithmic information. In this work, we only consider virtual diagrams for 2-parton final states, denoted by NLO_2. Note that the 3-parton matrix elements in the LO_3/LL and $\mathrm{NLO}_2/\mathrm{LO}_3$/LL samples are *not* the same, as eqs. (4.25) and (4.36) include different Sudakov factors. $\mathrm{LO}_{n_{\max}}$/LL and $\mathrm{NLO}_2/\mathrm{LO}_3$/LL provide implementations of PS/ME merging and PS/NLO merging, respectively. The $\mathrm{NLO}_2/\mathrm{LO}_{n_{\max}}$/LL sample is the best partonic calculation implemented by `GenEvA`, and simultaneously achieves PS/ME merging and PS/NLO merging.

used. Unlike in the companion paper [1], we always use fully merged samples that give a complete covering of phase space. For example, LO_4 includes 2-, 3-, and 4-parton tree-level matrix elements and the remaining phase space is covered by the internal parton shower. The same is true for LO_4/LL, except that the matrix elements are now Sudakov-improved according to eq. (5.9). Note that even without the Sudakov improvement all divergences in LO_n are regulated by the matching scale μ. Note also, that since `GenEvA` has no dead zones, the LO_2 sample is identical to LO_2/LL, because when only $2 \to 2$ matrix elements are used the parton shower always starts running at the center-of-mass energy. The matching scale μ can also affect the meaning of $n_{\max}$. For example, an LO_4/LL sample with $\mu = E_{\mathrm{CM}}$ would be identical to an LO_2/LL sample because in this case μ would be too high to allow for any additional emissions above μ.

For one-loop corrected matrix elements with $n \leq n_{\max}$ final state partons we write $\mathrm{NLO}_{n_{\max}}$, and we currently only use NLO_2. For example, an NLO_2/LL result corresponds to an LO_2/LL sample supplemented with an appropriate NLO k-factor, while $\mathrm{NLO}_2/\mathrm{LO}_3$/LL corresponds to fully consistent $\mathcal{O}(\alpha_s)$ results, which contain all $\mathcal{O}(\alpha_s)$ corrections from both virtual 2-parton and real-emission 3-parton diagrams, and including leading-log resummation according to eq. (4.36). Note that the 3-parton matrix elements in the LO_3/LL and $\mathrm{NLO}_2/\mathrm{LO}_3$/LL samples include different Sudakov factors, as can be seen comparing eqs. (4.25) and (4.36). An $\mathrm{NLO}_2/\mathrm{LO}_4$/LL sample additionally incorporates 4-parton tree-level matrix elements following eq. (5.15), and similarly for $\mathrm{NLO}_2/\mathrm{LO}_{n_{\max}}$/LL.

6.1 Total cross section scale dependence

We begin by presenting the dependence of the total cross section on the matching scale μ. On the left panel of figure 15 we show the results for $n_{\max} = 3$ and on the right panel for $n_{\max} = 6$. At $\mu = E_{\mathrm{CM}}$, the total cross section is given by $\sigma(E_{\mathrm{CM}}) = \sigma_{\mathrm{LO}}$ for the LO samples and $\sigma(E_{\mathrm{CM}}) = \sigma_{\mathrm{NLO}}$ for the NLO samples. As the matching scale μ is lowered, higher-multiplicity matrix elements are used in place of the parton shower to generate additional emissions. Using only tree-level matrix elements, we expect a double-logarithmic μ dependence $\alpha_s \log^2 \mu$ for the LO samples, but only a single-logarithmic dependence $\alpha_s \log \mu$

JHEP12(2008)010

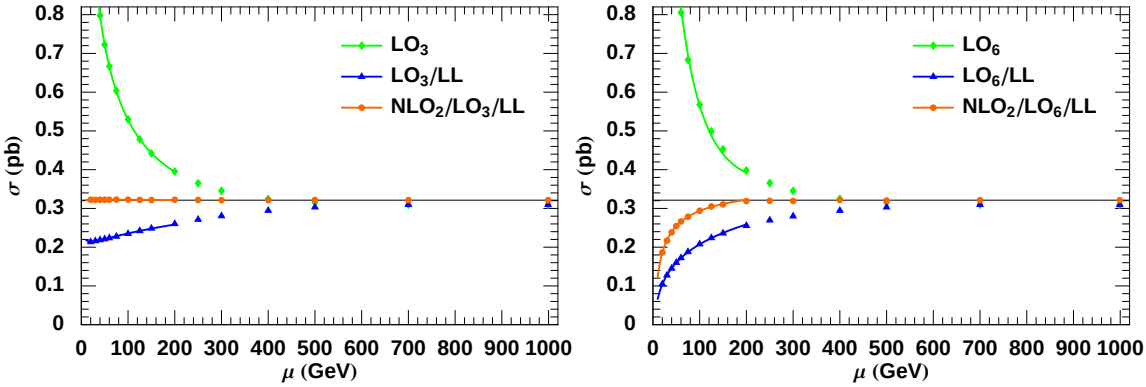


Figure 15: Dependence of the total cross section on the matching scale μ for $n_{\max} = 3$ (left panel) and $n_{\max} = 6$ (right panel) for $20\,\text{GeV} \leq \mu \leq 1000\,\text{GeV}$. The horizontal line corresponds to σ_{NLO}. The markers show the generated results and the solid lines the fit to eq. (6.2). For the LO and LO/LL samples $\sigma(E_{\text{CM}}) = \sigma_{\text{LO}}$, and for the NLO samples $\sigma(E_{\text{CM}}) = \sigma_{\text{NLO}}$. Lowering μ, higher-multiplicity matrix elements are used in place of the parton shower to generate further emissions, causing $\sigma(\mu)$ for the LO samples to scale like $\alpha_s \log^2 \mu$. With a proper LO/LL merging, the scale dependence reduces to $\alpha_s \log \mu$. Including NLO_2 information further reduces it to $(\alpha_s \log \mu)^2$ for $\text{NLO}_2/\text{LO}_6/\text{LL}$ (right panel), while for $\text{NLO}_2/\text{LO}_3/\text{LL}$ (left panel) the cross section becomes μ independent and identical to σ_{NLO} by construction.

	LO_3	LO_3/LL	$\text{NLO}_2/\text{LO}_3/\text{LL}$	LO_6	LO_6/LL	$\text{NLO}_2/\text{LO}_6/\text{LL}$
a_1	0.713	0.893	−0.037	7.40	1.203	−0.043
a_2	0.840	0.115	−0.006	2.74	0.056	−0.151

Table 3: Coefficients of the single-logarithmic (first row) and double-logarithmic (second row) μ dependence of the total cross section in figure 15. Shown is a fit to $\sigma(\hat{\mu}) = a_0 + \frac{2\alpha_s}{\pi}(a_1 \log \hat{\mu} + a_2 \log^2 \hat{\mu})$ in the range $0.02 < \hat{\mu} < 0.2$, where $\hat{\mu} = \mu/E_{\text{CM}}$ and $E_{\text{CM}} = 1000\,\text{GeV}$.

for the Sudakov-improved LO/LL samples. For the $\text{NLO}_2/\text{LO}_{n_{\max}}/\text{LL}$ samples, the full α_s dependence is included, and the μ dependence is thus expected to only start at $(\alpha_s \log \mu)^2$. For $n_{\max} = 3$, this term is absent because the total cross section is μ-independent by construction. We can clearly see in figure 15 that the scale dependence is reduced as we go from LO to LO/LL to NLO/LO/LL.

To check the expected scaling with μ explicitly, we fit the result for the total cross section obtained by `GenEvA` to the function

$$\sigma(\hat{\mu}) = a_0 + \frac{2\alpha_s}{\pi}(a_1 \log \hat{\mu} + a_2 \log^2 \hat{\mu})\,, \qquad \text{where} \qquad \hat{\mu} = \frac{\mu}{E_{\text{CM}}}\,. \tag{6.2}$$

This functional form neglects any contributions from power corrections of the form $\hat{\mu}^n$, which dominate for $\hat{\mu} \sim 1$ so we only use the range $\hat{\mu} \leq 0.2$ in the fit. The results of the fit are shown by the solid lines in figure 15, and the fitted parameters a_1 and a_2 are given in table 3. As anticipated, the LO_3 result has both single- and double-logarithmic μ dependence. Taking into account the overall factors in eq. (6.2) the values of a_1 and a_2 for the LO_3 sample are consistent with eq. (4.24). The LO_3/LL result has a much smaller a_2, indicating that the $\log^2 \hat{\mu}$ term only starts at $\mathcal{O}(\alpha_s^2)$, i.e. is a single-logarithmic effect,

JHEP12(2008)010

while a_1 is of comparable size as for the LO_3 sample. Thus, as expected, LO_3/LL still has a single-logarithmic dependence at $\mathcal{O}(\alpha_s)$. For $\mathrm{NLO}_2/\mathrm{LO}_3$/LL both a_1 and a_2 are small, indicating that there are no logarithmic terms present at $\mathcal{O}(\alpha_s)$.

For the samples with $n_{\max} = 6$, the cancellation of the μ dependence becomes more dramatic. The LO_6 sample shows a much larger μ dependence than LO_3 due to the accumulated μ dependence of all matrix elements. Nevertheless, this large μ dependence cancels in the Sudakov-improved LO_6/LL, which has similar coefficients to LO_3/LL. Finally, for $\mathrm{NLO}_2/\mathrm{LO}_6$/LL the value of a_2 is consistent with a residual μ dependence of the size $(\alpha_s \log \hat{\mu})^2$, as expected.

6.2 Interpolating kinematic extremes

Next, we study how differential distributions differ in the various implementations included in this work. We find the so-called C parameter [82] to be a particularly useful observable to highlight these effects. Given the linear sphericity tensor [83, 84]

$$S^{\alpha\beta} = \frac{\sum_i \frac{\mathbf{p}_i^\alpha \mathbf{p}_i^\beta}{|\mathbf{p}_i|}}{\sum_i |\mathbf{p}_i|} \tag{6.3}$$

constructed out of the final state three-momenta $\mathbf{p}_i$, the C-parameter is defined as

$$C = 3\left(\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1\right) , \tag{6.4}$$

where λ_i are the eigenvalues of $S^{\alpha\beta}$. Since $\sum_i \lambda_i = 1$ we have $0 \leq C \leq 1$. Low values of the C parameter are dominated by events which are mostly 2-jet-like, while larger values of C indicate more and more final state jets. For planar events $\lambda_3 = 0$ forcing $C \leq 0.75$, so that 3-jet events are confined to $C \leq 0.75$, while events with four or more jets can contribute up to $C = 1$.

We begin by studying the effect of Sudakov resummation on the tree-level matrix elements, which is the `GenEvA` analog of PS/ME merging. It is well known that the leading-logarithmic resummation for the C parameter reduces the cross section for small values of C, with the resummed expressions given in ref. [85]. It is also well known that the pure parton shower result overshoots the correct QCD result for large values of the C parameter, since the interference between the two QCD diagrams contributing to the emission of a single gluon is destructive. On the left panel of figure 16, we compare the results for the C-parameter obtained by running the pure parton shower (LO_2/LL), by using tree-level QCD matrix elements up to order α_s (LO_3), and the merged result with $n_{\max} = 3$ (LO_3/LL). We can see clearly how the LO_2/LL sample overshoots the LO_3 result for large values of C, while it suppresses the cross section for small values of C. The merged LO_3/LL result reproduces the result of the pure QCD matrix element for large values of C, where the leading logarithms are not important. The resummation of the double logarithmic terms becomes important for small values of C, and we can see the Sudakov suppression of the merged result compared to the pure matrix element for small values of C.

Note that the merged result has to approach the pure parton shower result for $C \to 0$, and we have checked that this is indeed the case. The reason this is not obvious from

JHEP12(2008)010

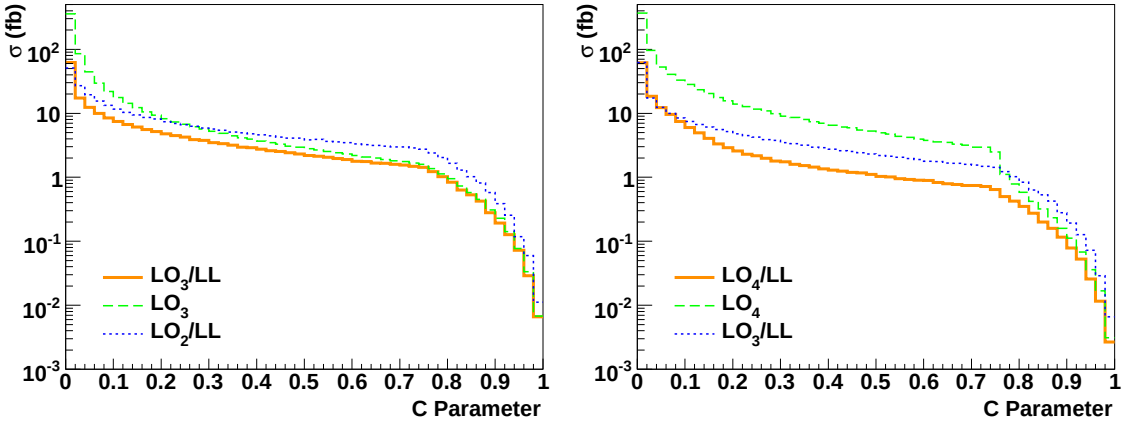


Figure 16: PS/ME merging as implemented in `GenEvA` (LO/LL) with $E_{\mathrm{CM}} = 1000\,\mathrm{GeV}$ and $\mu = 50\,\mathrm{GeV}$. The C parameter measures the "jettiness" of an event, with $C \sim 0$ giving the 2-jet region, $C < 0.75$ roughly giving the 3-jet region, and $C > 0.75$ roughly giving the 4-jet region. Left panel: The LO_3 sample has the soft-collinear divergence of the 3-parton tree-level matrix element and therefore becomes singular near $C = 0$, while the LO_3/LL sample regulates that divergence with a Sudakov factor similar to LO_2/LL. At large C, the LO_2/LL sample lacks the quantum interference of the 3-parton matrix element, while the LO_3/LL sample contains the correct interference effects and therefore tracks LO_3. Right panel: The same interpolation for LO_4/LL, which captures the correct Sudakov-suppression of the LO_3/LL sample near $C \sim 0$ and at the same time includes the extra interference effects of LO_4 near $C \sim 1$.

the figure is that the binning is too coarse for this effect to be visible, and the agreement does not happen until much smaller values of C.[20] To see more easily that the desired interpolation does occur, we include one additional tree-level emission, and compare the results for LO_4, LO_3/LL and LO_4/LL on the right panel of figure 16. One can now clearly see that the LO_4/LL sample reproduces the LO_3/LL result for small values of C. The agreement between the LO_4/LL and the LO_4 sample for large values of C is not expected to be perfect, since the Sudakov suppression has some effect all the way up to $C = 1$. However, we do see that the agreement is being obtained asymptotically.

Next, we show the effect of including NLO information, which is `GenEvA`'s analog of PS/NLO merging. As we saw in figure 15, the tree-level calculation contains double-logarithmic terms which results in a rising cross section as the scale $\hat{\mu}$ is lowered. For the C parameter distribution, this implies that the tree-level result diverges as $-\log(C)/C$ for $C \to 0$. Resumming the leading-logarithmic terms removes this dominant singularity for small values of C, however we also see in figure 15 that the cross section of the resummed result undershoots the correct (NLO) result. In figure 17, we compare the tree-level matrix elements LO_3, the resummed matrix elements LO_3/LL, as well as the resummed NLO result $\mathrm{NLO}_2/\mathrm{LO}_3$/LL. For large values of C all three results agree, as expected, because all three samples have the same interference terms from the tree-level 3-parton matrix element. How-

[20] Because of the single-logarithmic effect of the $\tilde{\mathbf{n}}$ integrations, the LO_3/LL sample first overshoots the LO_2/LL sample until finally asymptoting from above. With the $\tilde{\mathbf{n}}$ integrations turned off, the agreement happens at more moderate values of C, but still smaller than our bin size.

JHEP12(2008)010

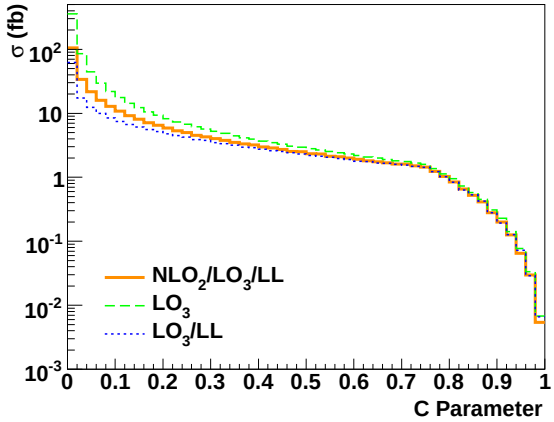


Figure 17: PS/NLO merging as implemented in `GenEvA` (NLO/LL). The $\mathrm{NLO}_2/\mathrm{LO}_3/\mathrm{LL}$, LO_3, and $\mathrm{LO}_3/\mathrm{LL}$ samples all share the same interference effects near $C = 1$, given by the 3-parton tree-level matrix element. Near $C = 0$, LO_3 is too singular, while $\mathrm{LO}_3/\mathrm{LL}$ is too Sudakov suppressed. The $\mathrm{NLO}_2/\mathrm{LO}_3/\mathrm{LL}$ sample has the correct NLO cross section information and therefore has the right normalization near $C = 0$. Since the inclusion of NLO cross-section information affects the form of the Sudakov factor for the NLO/LL sample, the $\mathrm{NLO}_2/\mathrm{LO}_3/\mathrm{LL}$ sample is *not* just a k-factor modification of the tree-level resummed $\mathrm{LO}_3/\mathrm{LL}$ result.

ever, for $C \to 0$, the $\mathrm{NLO}_2/\mathrm{LO}_3/\mathrm{LL}$ result has the correct $\mathcal{O}(\alpha_s)$ normalization, while LO_3 is too singular, and $\mathrm{LO}_3/\mathrm{LL}$ too Sudakov-suppressed. Note that the $\mathrm{NLO}_2/\mathrm{LO}_3/\mathrm{LL}$ result is *not* simply a k-factor modification of the tree-level result, because as seen in eqs. (4.25) and (4.36), the Sudakov factors in the 2- and 3-parton matrix elements must be modified to get the correct NLO cross section, while still incorporating leading-logarithmic results.

We now consider the NLO/LO/LL partonic calculation which combines NLO cross section information with higher-order tree-level matrix elements, all Sudakov-improved. We expect that in the 4-jet region ($C > 0.75$), the $\mathrm{NLO}_2/\mathrm{LO}_4/\mathrm{LL}$ sample will agree well with the $\mathrm{LO}_4/\mathrm{LL}$ result, since 3-parton states do not contribute much in this region. For small values of the C parameter, however, we expect the $\mathrm{NLO}_2/\mathrm{LO}_4/\mathrm{LL}$ result to be close to the $\mathrm{NLO}_2/\mathrm{LO}_3/\mathrm{LL}$ result, since the NLO cross section information is important there. The results are shown on the left panel of figure 18, which confirms the expected interpolation. On the right panel of figure 18, we give the result of the best partonic calculation current available in `GenEvA`: $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$. This sample includes additional interference terms from tree-level matrix elements with $n \leq 6$, while still maintaining NLO/LL accuracy. We clearly see that the $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$ sample interpolates between the $\mathrm{LO}_6/\mathrm{LL}$ (PS/ME merged) sample in the interference region ($C \sim 1$) and the $\mathrm{NLO}_2/\mathrm{LO}_3/\mathrm{LL}$ (PS/NLO merged) sample in the total cross section region ($C \sim 0$). In other words, the best implementation available in `GenEvA` simultaneously achieves PS/ME merging and PS/NLO merging.

Finally, on the left panel of figure 19, we illustrate the composition of the $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$ sample, showing the individual contributions of the n-parton matrix elements, where n corresponds to the number of partons resolvable at the scale μ. Near the 2-jet region ($C \sim 0$), the $n = 2$ matrix element dominates as expected. The $n = 3$ matrix element fills out the region $0 < C < 0.3$, and since it gives mostly planar events, it turns

JHEP12(2008)010

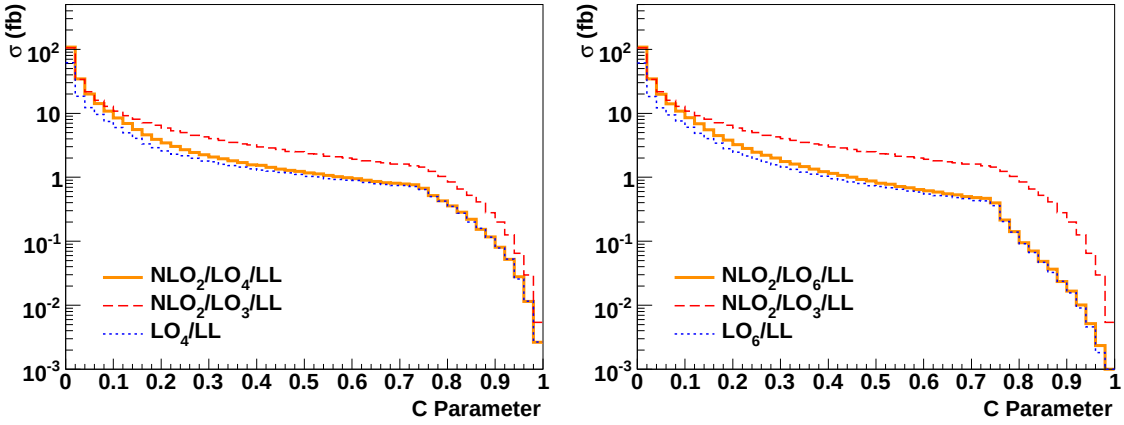


Figure 18: The combination of PS/ME merging and PS/NLO merging in `GenEvA` (NLO/LO/LL). Left panel: The LO_4/LL sample is the analog of PS/ME merging and contains the correct interference of the 4-parton matrix element near $C = 1$, while the NLO_2/LO_3/LL sample is the analog of PS/NLO merging, and has the correct cross section information near $C = 0$. The NLO_2/LO_4/LL sample smoothly interpolates between the two different regimes. Right panel: The same interpolation for $n_{\max} = 6$, where NLO_2/LO_6/LL smoothly interpolates between LO_6/LL at large C and NLO_2/LO_3/LL at small C. The kinked behavior near $C = 0.75$ is a well-known physical effect [82], unrelated to the merging procedure. It occurs because $C \leq 0.75$ for planar events, meaning that 2- and 3-parton matrix elements cannot contribute much for $C > 0.75$. The NLO_2/LO_6/LL sample is the best partonic calculation currently implemented in `GenEvA`.

off abruptly at $C = 0.75$. Despite the fact that $C < 0.75$ is supposed to correspond to the 3-jet region, the $n = 4$ matrix element dominates for $0.3 < C < 0.75$, because $\mu = 50$ GeV is so low, that the $n = 4$ matrix element is really being used to determine some of the jet substructure in the 3-jet region. Above $C = 0.75$, the $n = 4$ and $n = 5$ matrix elements give roughly equal contributions, with the region near $C = 1$ supplemented by the $n = 6$ matrix element. This is not surprising as the C parameter no longer resolves the difference between 4-jet and higher-jet event shapes. On the right panel of figure 19, the same C parameter is shown for $\mu = 100$ GeV. With this matching scale, the separation of the C parameter into n-jet-like regions roughly corresponds to the n-parton matrix elements. Note that the $n = 6$ matrix element now barely has phase space available to contribute to the total distribution, so the remaining matrix elements shift to compensate.

6.3 More results from GenEvA best

To get more intuition about the `GenEvA` Best NLO/LO/LL sample, it is helpful to use a jet algorithm to get more differential information than is available in an event shape measure like the C parameter. We use the `FastJet` package [86] for that purpose, using the inclusive k_T jet algorithm with $R = 1.0$ to identify jets, which are then ordered by their total energy.

On the left panel of figure 20, we show the invariant mass of the 2nd hardest jet. Events which are almost 2-jet-like contribute to small values of this observable, while events with well-separated partons contribute to larger values. As we saw in the previous section, adding additional tree-level matrix elements suppresses rates for well-separated partons

JHEP12(2008)010

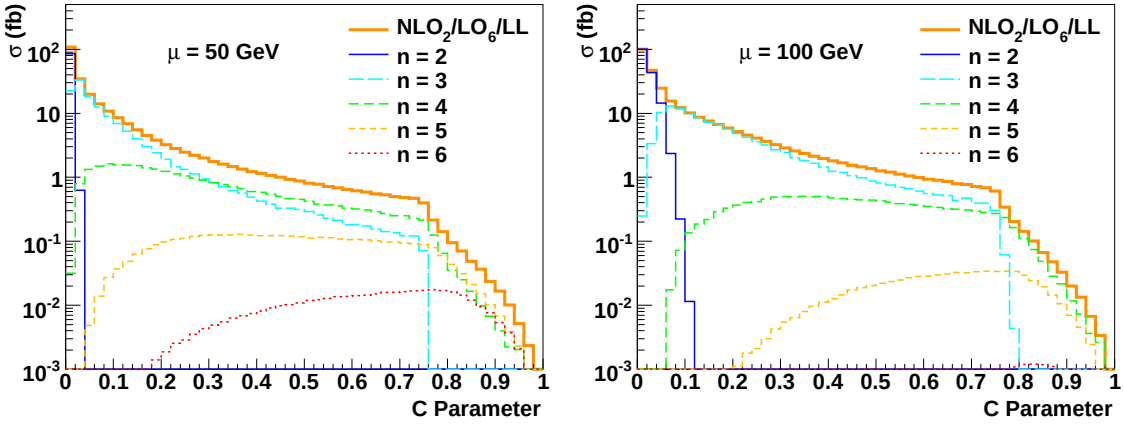


Figure 19: The five components of the `GenEvA` Best $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$ sample. The curves for different n show the individual contributions of the n-parton matrix elements, where n corresponds to the number of partons resolvable at the scale μ. Left panel: $\mu = 50\,\mathrm{GeV}$. As expected, the $n = 2$ component dominates near $C = 0$. Because μ is so low, the $n = 3$ and $n = 4$ matrix elements both contribute in the 3-jet ($C < 0.75$) region, where the crossover point $C = 0.3$ is directly related to the scale μ. Above $C = 0.75$, the $n = 3$ component can no longer contribute, and the rest of the C distribution is filled out with additional contributions from the 4-, 5-, and 6-parton matrix elements. Right panel: $\mu = 100\,\mathrm{GeV}$. For this more reasonable μ scale, the various n-jet-like regions correspond more closely to the n-parton matrix element used. While the total C parameter distribution is only single-logarithmically sensitive to changing the μ scale, the different matrix element components shift dramatically, with the $n = 6$ matrix element now having very little available phase space.

due to interference effects in the full matrix-element calculations. On the other hand, the NLO information increases the cross section in regions of phase space which contribute mostly to 2-jet-like events. Thus, we expect the $\mathrm{NLO}_2/\mathrm{LO}_3/\mathrm{LL}$ result to be higher than the $\mathrm{LO}_6/\mathrm{LL}$ sample in the whole kinematic region, which is seen in the figure. Combining these two results, we expect to reproduce the lower rate for well-separated partons, due to the interference from the tree-level matrix elements, while at the same time reproducing the enhanced cross section for almost 2-jet-like events due to the NLO information. This is clearly seen in the $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$ sample, which thus captures important physical effects that cannot be seen by considering the PS/ME merging or PS/NLO merging results alone.

The right panel of figure 20 shows the invariant mass between the 1st and 3rd hardest jets. Low values of this pairwise invariant mass correspond to the 2-jet-like region, and again the $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$ sample inherits the NLO cross section information. As the pairwise invariant mass increases, this observable gets contributions from a variety of different event types, so the $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$ sample tracks the shape of the $\mathrm{LO}_6/\mathrm{LL}$ result with an overall NLO cross section increase. At very large values of this invariant mass, all three curves give slightly different answers because this extreme kinematic region is sensitive to the exact way in which the jets are clustered, which is strongly affected by the exact ratios of the 4-, 5-, and 6-parton matrix elements.

As seen in the C parameter, the $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$ sample combines 5 different matrix el-

JHEP12(2008)010

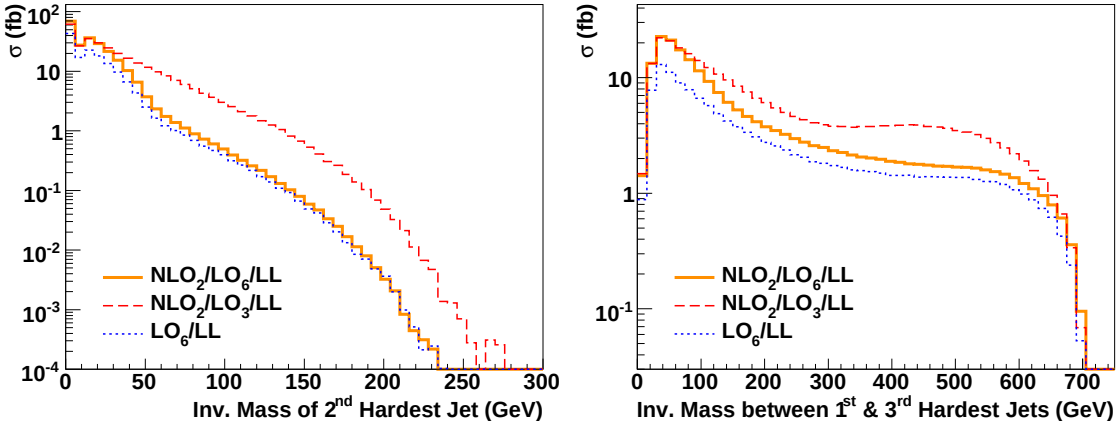


Figure 20: Comparison of the `GenEvA` best sample $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$ with the PS/ME merging analog $\mathrm{LO}_6/\mathrm{LL}$ and the PS/NLO merging analog $\mathrm{NLO}_2/\mathrm{LO}_3/\mathrm{LL}$ for two jet-based observables. Left panel: The invariant mass of the 2nd hardest jet, where we can clearly see the destructive interference in the matrix element calculation with $n > 3$ partons compared to the results where these additional partons are only generated by the parton shower. The $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$ sample interpolates between the two comparison distributions, showing that `GenEvA` can capture important physical effects that cannot be seen by considering the $\mathrm{NLO}_2/\mathrm{LO}_3/\mathrm{LL}$ or $\mathrm{LO}_6/\mathrm{LL}$ samples alone. Right panel: The invariant mass between the 1st and 3rd hardest jets. Since no particular parton multiplicity dominates at large pairwise invariant masses, the three different samples have slightly different behaviors near the endpoint.

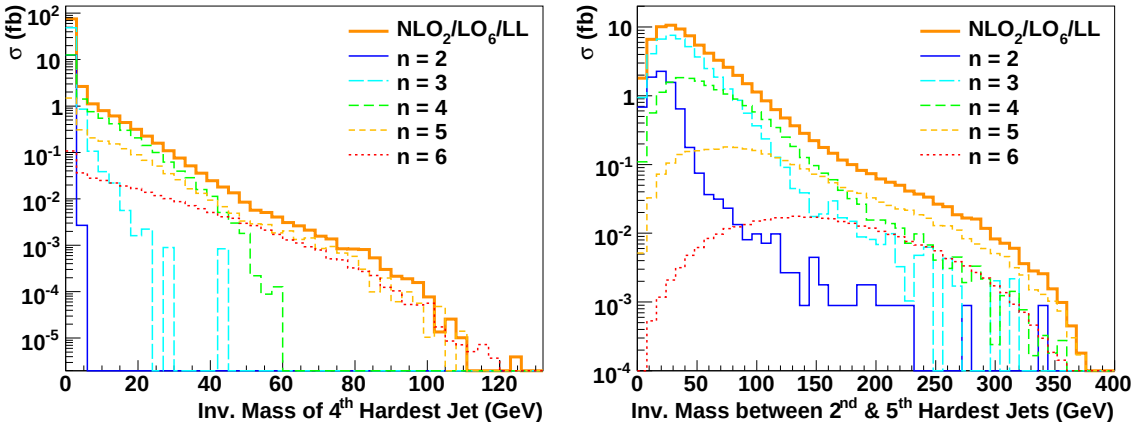


Figure 21: The five components of the $\mathrm{NLO}_2/\mathrm{LO}_6/\mathrm{LL}$ sample for two jet-based observables. The curves for different n show the contributions of the individual n-parton matrix elements. Left panel: The invariant mass of the 4th hardest jet. Since `GenEvA` currently uses virtuality to define the scale μ that separates different parton-level multiplicities, the invariant masses of jets have sharp cutoffs for the different parton-multiplicity components, with the $n = 4$ sample turning off at the matching scale around 50 GeV as expected (the small tail comes from the jet algorithm "accidentally" clustering together two different quark flavors). Right panel: The invariant mass between the 2nd and 5th hardest jets. As this pairwise invariant mass increases, the dominant n-parton matrix element smoothly changes from $n = 3$ to $n = 4$ to $n = 5$.

JHEP12(2008)010

ements, and it is interesting to see the individual matrix-element contributions for jet-based observables. On the left panel of figure 21, we show the results for the invariant mass of the 4th hardest jet. Since our evolution variable is equal to the virtuality between partons, the parton shower can only generate virtualities below the staring scale of the shower, which is chosen as $\mu = 50\,\text{GeV}$. Thus we expect that a 4th jet with invariant mass above 50 GeV can only be generated by a partonic calculation with at least 5 partons in the final state.[21] This effect is reproduced by `GenEvA`, and while there are sharp cutoffs in the individual contributions of $n = 2, 3, 4$-parton matrix elements, the combined $\text{NLO}_2/\text{LO}_6/\text{LL}$ result is relatively smooth over the entire range of the invariant mass of the 4th hardest jet. Note that there is a slight kink in the final distribution at the matching scale 50 GeV, which gives a sense of the size of the subleading-logarithmic errors one makes in this merging.

For the pairwise invariant mass between the 2nd and 5th hardest jets, shown on the right panel of figure 21, the individual partonic contributions do not cut off as sharply as for the previous case. The dominant matrix element smoothly changes from $n = 3$ for small invariant mass, to $n = 4$ for intermediate masses, to $n = 5$ towards the endpoint of the distribution. The total result is again a very smooth function over the entire range of the inter-jet invariant mass, with perhaps a slight kink at 175 GeV from the same $n = 4$ to $n = 5$ transition that gave a kink in the previous distribution.

As a final application, we study the scale dependence of a differential distribution for $n_{\text{max}} = 4$. We saw in figure 15 that the μ dependence of the total cross section was reduced when going from LO to LO/LL to NLO/LO/LL, and it is interesting to see the extent to which this reduced scale dependence is reflected in differential distributions. In figure 22, we show the invariant mass between the 1st and 3rd hardest jets for these three samples. We see that the μ dependence is lowered going from LO_4 sample (top left plot) to the LO_4/LL sample (top right plot). We also see how the direction of the scale dependence is reversed, as expected from the scale dependence of the total cross section. In the bottom plot we show the result for the $\text{NLO}_2/\text{LO}_4/\text{LL}$ sample. Including the NLO information does not change the scale dependence significantly for large values of the invariant mass, since this is the region where the real-emission diagrams are dominant. However, for small values of the invariant mass, i.e. in the 2-jet region, the scale dependence is reduced, mirroring the reduced scale dependence observed in the total cross section. The residual scale dependence is single logarithmic.

7. Conclusions

We have presented a new framework for event generation, `GenEvA`, which allows almost any partonic calculation to be interfaced with a phenomenological model such as a parton shower. Because the phenomenological model is assumed to include a description of hadronization, `GenEvA` offers a method for inclusive partonic information to be used to produce fully exclusive hadronic events.

[21] There are $n = 4$ events that can have jet masses slightly larger than the matching scale, because the cutoff only applies to singularity-producing partons. One can see the effect of, say, the $e^+e^- \to u\bar{u}d\bar{d}$ sample in the rare events that have a 4th jet with invariant mass upwards of 60 GeV.

JHEP12(2008)010

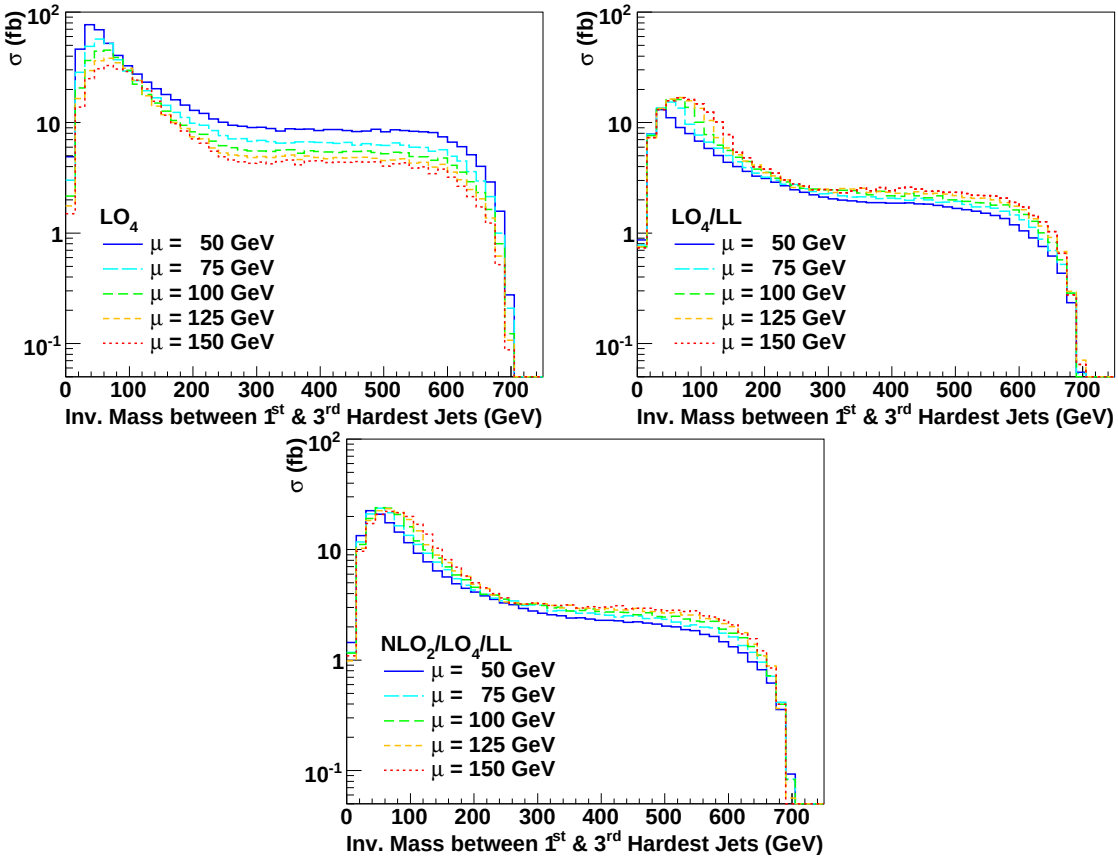


Figure 22: Comparison of the μ dependence of the LO_4, LO_4/LL, and NLO_2/LO_4/LL samples. Shown is the distribution of the invariant mass between the 1st and 3rd hardest jets for five different values of the matching scale μ. Just as in figure 15, the inclusion of leading-logarithmic and NLO cross-section information reduces the dependence on the unphysical matching scale μ. At large invariant masses, the dominant reduction comes from including leading-logarithmic information in high-multiplicity matrix elements. At small invariant masses, there is an additional reduction in the μ dependence from the inclusion of NLO cross-section information.

The main conceptual breakthrough is a definition of phase space with a matching scale. This matching scale cleanly separates partonic calculations performed in QCD from phenomenological models based on QCD. `GenEvA` avoids phase space double-counting by construction, because once `GenEvA` is interfaced with a parton shower, every point in perturbative phase space is covered once and only once. By appropriate inclusion of leading-logarithmic information in the partonic calculations, the dominant sensitivity to the unphysical scale which separates the partonic regime from the showering regime is removed.

In this way, `GenEvA` allows theorists to focus on providing the best possible partonic calculations and not on the algorithmic details of implementing those calculations. While other solutions to the problems of double-counting and shower merging exist in the literature, they are tailored to specific partonic calculations. For example, though there are various methods to merge tree-level calculations with parton showers [23 – 35], usually entirely different procedures are necessary to merge NLO calculations with parton showers [40 –

JHEP12(2008)010

53]. In contrast, GenEvA offers a generic solution to these problems, and the NLO/LO/LL merged sample provides an example of the kinds of improved partonic calculations that can be developed when algorithmic issues are separated from calculational ones.

The name GenEvA is obviously inspired by the site of the upcoming LHC experiment at CERN, and we expect to generalize the GenEvA framework to deal with additional complications present in a hadronic environment. We discuss some of these issues in our companion paper [1], and we argue there that the complications are technical, not conceptual. Much of the legwork has already been done, as modern parton showers implement initial state radiation through backwards evolution, effectively defining the notion of a variable matching scale for hadronic collisions.

From a theoretical point of view, the most interesting developments will be to implement more advanced kinds of partonic calculations. In this work, we only considered loop diagrams involving two final state particles, but there is a growing body of one-loop calculations with large numbers of final states (for a recent review see ref. [87]). Because GenEvA is agnostic as to the method of regulating IR divergences in loop diagrams, the choice about whether to use a slicing method, a subtraction method, or an elegant method to supplement NLO calculations with leading-logarithmic information can be made on the basis of physics considerations alone. Also interesting will be the implementation of calculations [62, 61, 88] based on SCET [57 – 60] because it allows for a consistent treatment of subleading logarithms.

From an experimental point of view, GenEvA offers a unique opportunity to assess Monte Carlo systematics, as a single GenEvA event can support multiple different theoretical distributions, allowing theoretical errors to be probed without additional detector simulation time [1]. The NLO/LO/LL merged sample also gives a proof-of-concept that multiple different theoretical distributions can coexist within a single Monte Carlo sample, allowing the experiments to use the best theoretical knowledge available for any given point in phase space. With the possibility for sub-GeV measurements of the top quark mass and aggressive use of Monte Carlo to extrapolate the Standard Model up to 14 TeV, we anticipate that the experimental collaborations will benefit from the flexibility and transparency of the GenEvA framework in the LHC era.

Acknowledgments

We would like to thank Johan Alwall, Lance Dixon, Walter Giele, Beate Heinemann, Zoltan Ligeti, Michelangelo Mangano, Michael Peskin, Matthew Schwartz, Torbjörn Sjöstrand, Peter Skands, Iain Stewart, and Dieter Zeppenfeld for many useful discussions. We would also like to thank Johan Alwall, Fabio Maltoni, and Tim Stelzer for patient assistance with MadGraph, and Jeffrey Anderson for help with general computing questions. This work was supported in part by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under the Contract DE-AC02-05CH11231. CWB acknowledges support from an DOE OJI award and an LDRD grant from LBNL. JT is supported by a fellowship from the Miller Institute for Basic Research in Science.

JHEP12(2008)010

References

[1] C.W. Bauer, F.J. Tackmann and J. Thaler, *GenEvA (II): a phase space generator from a reweighted parton shower*, *JHEP* **12** (2008) 011 [arXiv:0801.4028].

[2] Tevatron Electroweak Working Group collaboration, E. Brubaker et al., *Combination of CDF and D0 results on the mass of the top quark*, hep-ex/0603039.

[3] CDF collaboration, T. Aaltonen et al., *First measurement of the W boson mass in run II of the Tevatron*, *Phys. Rev. Lett.* **99** (2007) 151801 [arXiv:0707.0085].

[4] CDF collaboration, T. Aaltonen et al., *First run II measurement of the W boson mass*, *Phys. Rev.* **D 77** (2008) 112001 [arXiv:0708.3642].

[5] ATLAS collaboration, *ATLAS: detector and physics performance technical design report. Volume* 1, CERN-LHCC-99-14 [ATLAS-TDR-14].

[6] ATLAS collaboration, *ATLAS: detector and physics performance technical design report. Volume* 2, CERN-LHCC-99-15 [ATLAS-TDR-15].

[7] CMS collaboration, G.L. Bayatian et al., *CMS technical design report, volume II: physics performance*, *J. Phys.* **G 34** (2007) 995.

[8] L. Lönnblad, *ARIADNE version* 4*: a program for simulation of QCD cascades implementing the color dipole model*, *Comput. Phys. Commun.* **71** (1992) 15.

[9] T. Sjöstrand et al., *High-energy-physics event generation with PYTHIA* 6.1, *Comput. Phys. Commun.* **135** (2001) 238 [hep-ph/0010017].

[10] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA* 6.4 *physics and manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175].

[11] G. Corcella et al., *HERWIG* 6*: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010 [hep-ph/0011363].

[12] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A.D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003) 001 [hep-ph/0206293].

[13] F. Maltoni and T. Stelzer, *MadEvent: automatic event generation with MadGraph*, *JHEP* **02** (2003) 027 [hep-ph/0208156].

[14] S. Gieseke, A. Ribon, M.H. Seymour, P. Stephens and B. Webber, *HERWIG++* 1.0*: an event generator for* e^+e^- *annihilation*, *JHEP* **02** (2004) 005 [hep-ph/0311208].

[15] S. Gieseke et al., *HERWIG++* 2.0 *release note*, hep-ph/0609306.

[16] T. Gleisberg et al., *SHERPA* 1.α*, a proof-of-concept version*, *JHEP* **02** (2004) 056 [hep-ph/0311263].

[17] F.E. Paige, S.D. Protopopescu, H. Baer and X. Tata, *ISAJET* 7.69*: a Monte Carlo event generator for pp,* $\bar{p}p$ *and* e^+e^- *reactions*, hep-ph/0312045.

[18] CompHEP collaboration, E. Boos et al., *CompHEP* 4.4*: automatic computations from Lagrangians to events*, *Nucl. Instrum. Meth.* **A534** (2004) 250 [hep-ph/0403113].

[19] W. Kilian, T. Ohl and J. Reuter, *WHIZARD: simulating multi-particle processes at LHC and ILC*, arXiv:0708.4233.

JHEP12(2008)010

[20] A. Cafarella, C.G. Papadopoulos and M. Worek, *Helac-Phegas: a generator for all parton level processes*, arXiv:0710.2427.

[21] CDF collaboration, T. Aaltonen et al., *Model-independent and quasi-model-independent search for new physics at CDF*, *Phys. Rev.* **D 78** (2008) 012002 [arXiv:0712.1311].

[22] CDF collaboration, T. Aaltonen et al., *Model-independent global search for new high-pT physics at CDF*, arXiv:0712.2534.

[23] S. Catani, F. Krauss, R. Kuhn and B.R. Webber, *QCD matrix elements + parton showers*, *JHEP* **11** (2001) 063 [hep-ph/0109231].

[24] L. Lönnblad, *Correcting the colour-dipole cascade model with fixed order matrix elements*, *JHEP* **05** (2002) 046 [hep-ph/0112284].

[25] F. Krauss, *Matrix elements and parton showers in hadronic interactions*, *JHEP* **08** (2002) 015 [hep-ph/0205283].

[26] M.L. Mangano, *The so-called MLM prescription for ME/PS matching*, talk presented at the *Fermilab ME/MC Tuning Workshop*, October 4 2004.

[27] S. Mrenna and P. Richardson, *Matching matrix elements and parton showers with HERWIG and PYTHIA*, *JHEP* **05** (2004) 040 [hep-ph/0312274].

[28] A. Schalicke and F. Krauss, *Implementing the ME+PS merging algorithm*, *JHEP* **07** (2005) 018 [hep-ph/0503281].

[29] N. Lavesson and L. Lönnblad, *W + jets matrix elements and the dipole cascade*, *JHEP* **07** (2005) 054 [hep-ph/0503293].

[30] S. Höche et al., *Matching parton showers and matrix elements*, hep-ph/0602031.

[31] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, *Eur. Phys. J.* **C 53** (2008) 473 [arXiv:0706.2569].

[32] 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].

[33] N. Lavesson and L. Lönnblad, *Merging parton showers and matrix elements — back to basics*, *JHEP* **04** (2008) 085 [arXiv:0712.2966].

[34] Z. Nagy and D.E. Soper, *Parton showers with quantum interference*, *JHEP* **09** (2007) 114 [arXiv:0706.0017].

[35] Z. Nagy and D.E. Soper, *Parton showers with quantum interference: leading color, spin averaged*, *JHEP* **03** (2008) 030 [arXiv:0801.1917].

[36] R.K. Ellis and J. Campbell, *MCFM - Monte Carlo for FeMtobarn processes*, http://mcfm.fnal.gov.

[37] Z. Nagy, *NLOJet++*, http://nagyz.web.cern.ch/nagyz/Site/NLOJet++.html.

[38] P. Aurenche et al., *The PHOX family*, http://wwwlapp.in2p3.fr/lapth/PHOX_FAMILY/main.html.

[39] M. Bähr et al., *VBFNLO - NLO parton level Monte Carlo for vector boson fusion*, http://www-itp.particle.uni-karlsruhe.de/~vbfnloweb/.

JHEP12(2008)010

[40] J.C. Collins and F. Hautmann, *Soft gluons and gauge-invariant subtractions in NLO parton-shower Monte Carlo event generators*, *JHEP* **03** (2001) 016 [hep-ph/0009286].

[41] J.C. Collins, *Subtraction method for NLO corrections in Monte Carlo event generators for leptoproduction*, *JHEP* **05** (2000) 004 [hep-ph/0001040].

[42] B. Potter and T. Schorner, *Combining parton showers with next-to-leading order QCD matrix elements in deep-inelastic e p scattering*, *Phys. Lett.* **B 517** (2001) 86 [hep-ph/0104261].

[43] 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].

[44] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029 [hep-ph/0204244].

[45] S. Frixione, P. Nason and B.R. Webber, *Matching NLO QCD and parton showers in heavy flavour production*, *JHEP* **08** (2003) 007 [hep-ph/0305252].

[46] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040 [hep-ph/0409146].

[47] P. Nason and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for Z pair hadroproduction*, *JHEP* **08** (2006) 077 [hep-ph/0606275].

[48] O. Latunde-Dada, S. Gieseke and B. Webber, *A positive-weight next-to-leading-order Monte Carlo for e^+e^- annihilation to hadrons*, *JHEP* **02** (2007) 051 [hep-ph/0612281].

[49] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070 [arXiv:0709.2092].

[50] M. Krämer and D.E. Soper, *Next-to-leading order QCD calculations with parton showers. I: collinear singularities*, *Phys. Rev.* **D 69** (2004) 054019 [hep-ph/0306222].

[51] D.E. Soper, *Next-to-leading order QCD calculations with parton showers. II: soft singularities*, *Phys. Rev.* **D 69** (2004) 054020 [hep-ph/0306268].

[52] Z. Nagy and D.E. Soper, *Matching parton showers to NLO computations*, *JHEP* **10** (2005) 024 [hep-ph/0503053].

[53] M. Krämer, S. Mrenna and D.E. Soper, *Next-to-leading order QCD jet production with parton showers and hadronization*, *Phys. Rev.* **D 73** (2006) 014022 [hep-ph/0509127].

[54] S. Frixione and B.R. Webber, *The MC@NLO* 3.3 *event generator*, hep-ph/0612272.

[55] S. Frixione, P. Nason and G. Ridolfi, *The POWHEG-hvq manual version* 1.0, arXiv:0707.3081.

[56] S. Bethke, Z. Kunszt, D.E. Soper and W.J. Stirling, *New jet cluster algorithms: next-to-leading order QCD and hadronization corrections*, *Nucl. Phys.* **B 370** (1992) 310 [*Erratum ibid.* **B 523** (1998) 681].

[57] C.W. Bauer, S. Fleming and M.E. Luke, *Summing Sudakov logarithms in $B \to X_s\gamma$ in effective field theory*, *Phys. Rev.* **D 63** (2000) 014006 [hep-ph/0005275].

[58] C.W. Bauer, S. Fleming, D. Pirjol and I.W. Stewart, *An effective field theory for collinear and soft gluons: heavy to light decays*, *Phys. Rev.* **D 63** (2001) 114020 [hep-ph/0011336].

[59] C.W. Bauer and I.W. Stewart, *Invariant operators in collinear effective theory*, *Phys. Lett.* **B 516** (2001) 134 [hep-ph/0107001].

JHEP12(2008)010

[60] C.W. Bauer, D. Pirjol and I.W. Stewart, *Soft-collinear factorization in effective field theory*, *Phys. Rev.* **D 65** (2002) 054022 [hep-ph/0109045].

[61] C.W. Bauer and M.D. Schwartz, *Improving jet distributions with effective field theory*, *Phys. Rev. Lett.* **97** (2006) 142001 [hep-ph/0604065].

[62] C.W. Bauer and M.D. Schwartz, *Event generation from effective field theory*, *Phys. Rev.* **D 76** (2007) 074004 [hep-ph/0607296].

[63] C.W. Bauer and F.J. Tackmann, *Gaining analytic control of parton showers*, *Phys. Rev.* **D 76** (2007) 114017 [arXiv:0705.1719].

[64] V.V. Sudakov, *Vertex parts at very high-energies in quantum electrodynamics*, *Sov. Phys. JETP* **3** (1956) 65 [*Zh. Eksp. Teor. Fiz.* **30** (1956) 87].

[65] V.N. Gribov and L.N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438 [*Yad. Fiz.* **15** (1972) 781].

[66] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, *Nucl. Phys.* **B 126** (1977) 298.

[67] Y.L. Dokshitzer, *Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics* (in Russian), *Sov. Phys. JETP* **46** (1977) 641 [*Zh. Eksp. Teor. Fiz.* **73** (1977) 1216].

[68] S. Catani, Y.L. Dokshitzer, M. Olsson, G. Turnock and B.R. Webber, *New clustering algorithm for multi-jet cross-sections in e^+e^- annihilation*, *Phys. Lett.* **B 269** (1991) 432.

[69] S.D. Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* **D 48** (1993) 3160 [hep-ph/9305266].

[70] G. Gustafson and U. Pettersson, *Dipole formulation of QCD cascades*, *Nucl. Phys.* **B 306** (1988) 746.

[71] 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].

[72] 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].

[73] K. Fabricius, I. Schmitt, G. Kramer and G. Schierholz, *Higher order perturbative QCD calculation of jet cross-sections in e^+e^- annihilation*, *Z. Physik* **C 11** (1981) 315.

[74] G. Kramer and B. Lampe, *Jet cross-sections in e^+e^- annihilation*, *Fortschr. Phys.* **37** (1989) 161.

[75] L.J. Bergmann, *Next-to-leading log QCD calculation of symmetric dihadron production*, UMI-89-15738.

[76] W.T. Giele and E.W.N. Glover, *Higher order corrections to jet cross-sections in e^+e^- annihilation*, *Phys. Rev.* **D 46** (1992) 1980.

[77] W.T. Giele, E.W.N. Glover and D.A. Kosower, *Higher order corrections to jet cross-sections in hadron colliders*, *Nucl. Phys.* **B 403** (1993) 633 [hep-ph/9302225].

[78] B.W. Harris and J.F. Owens, *The two cutoff phase space slicing method*, *Phys. Rev.* **D 65** (2002) 094032 [hep-ph/0102128].

JHEP12(2008)010

[79] R.K. Ellis, W.J. Stirling and B.R. Webber, *QCD and collider physics*, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **8** (1996) 1.

[80] H. Murayama, I. Watanabe and K. Hagiwara, *HELAS: HELicity amplitude subroutines for Feynman diagram evaluations*, KEK-91-11.

[81] T. Stelzer and W.F. Long, *Automatic generation of tree level helicity amplitudes*, *Comput. Phys. Commun.* **81** (1994) 357 [hep-ph/9401258].

[82] R.K. Ellis, D.A. Ross and A.E. Terrano, *The perturbative calculation of jet structure in e^+e^- annihilation*, *Nucl. Phys.* **B 178** (1981) 421.

[83] G. Parisi, *Super inclusive cross-sections*, *Phys. Lett.* **B 74** (1978) 65.

[84] J.F. Donoghue, F.E. Low and S.-Y. Pi, *Tensor analysis of hadronic jets in quantum chromodynamics*, *Phys. Rev.* **D 20** (1979) 2759.

[85] S. Catani and B.R. Webber, *Resummed C-parameter distribution in e^+e^- annihilation*, *Phys. Lett.* **B 427** (1998) 377 [hep-ph/9801350].

[86] M. Cacciari, *FastJet: dispelling the N^3 myth for the k_t jet-finder*, hep-ph/0607071.

[87] Z. Bern, L.J. Dixon and D.A. Kosower, *On-shell methods in perturbative QCD*, *Ann. Phys. (NY)* **322** (2007) 1587 [arXiv:0704.2798].

[88] M.D. Schwartz, *Resummation and NLO matching of event shapes with effective field theory*, *Phys. Rev.* **D 77** (2008) 014026 [arXiv:0709.2709].

JHEP12(2008)010