

GenEvA (II): a phase space generator from a reweighted parton shower

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 introduce a new efficient algorithm for phase space generation. A parton shower is used to distribute events across all of multiplicity, flavor, and phase space, and these events can then be reweighted to any desired analytic distribution. To verify this method, we reproduce the $e^+e^- \rightarrow n$ jets tree-level result of traditional matrix element tools. We also show how to improve tree-level matrix elements automatically with leading-logarithmic resummation. This algorithm is particularly useful in the context of a new framework for event generation called GenEvA. In a companion paper [1], we show how the GenEvA framework can address contemporary issues in event generation.

KEYWORDS: Jets, Phenomenological Models, QCD.

Contents

1. Introduction	2
1.1 Motivation	2
1.2 Phase space from a parton shower	4
1.3 Benefits of GenEvA	5
2. Overview of the GenEvA algorithm	6
2.1 The master formula	6
2.2 Calculating the event weight	10
2.2.1 The parton shower probability $\mathcal{P}(\Sigma)$	10
2.2.2 The Jacobian $J(\Sigma)$	11
2.2.3 The overcounting factor $\hat{\alpha}(\Sigma)$	11
2.2.4 The matrix element $\sigma[\Phi(\Sigma)]$	12
2.3 Reweighting efficiency	12
3. The analytic parton shower	13
3.1 Overview	14
3.2 Choice of kinematic variables	15
3.3 The single-branch probability	17
3.4 The double-branch probability	18
3.5 Truncation	22
4. The Jacobian factor	27
5. The overcounting factor	29
5.1 Overview	29
5.2 The ALPHA algorithm	30
5.3 A shower “Lagrangian”	32
5.4 Choice of overcounting function	34
6. The matrix element	35
6.1 Tree-level matrix elements	35
6.2 Logarithmically-improved matrix elements	37
6.3 Further logarithmic possibilities	39
7. Results	41
7.1 Comparison with MadEvent	41
7.2 Logarithmically-improved results	43
7.3 The efficiency of GenEvA	44

8. Conclusions	48
8.1 Towards hadronic collisions	48
8.2 Heavy resonances	49
8.3 Choice of evolution variable	50
8.4 Outlook	51

1. Introduction

1.1 Motivation

Precision theoretical calculations are vital for any comparison between experimental measurements and underlying theoretical models. While in many cases we are interested in the effects of physics Beyond the Standard Model, a precise understanding of Standard Model backgrounds is always required for such a comparison to be useful. To be able to study the effect of experimental cuts and detector effects, the most useful form for theoretical predictions is as fully exclusive events with a high multiplicity of hadrons in the final state.

Ignoring the issue of hadronization and other nonperturbative effects, the challenge for constructing exclusive event generators at the perturbative level is that full quantum field theoretic results for high-multiplicity final states become prohibitively difficult both to calculate and to implement. At tree level, the number of required Feynman diagrams grows factorially with the number of partons in the final state, so that a direct calculation is not feasible. Even if an expression for the partonic calculation can be obtained, the large volume of high-multiplicity phase space makes it difficult to efficiently distribute events according to those distributions. Both of these constraints imply that in practice, it is impossible to use full quantum information to calculate typical partonic final states arising in collider environments.

It has been known for some time that for emissions at small transverse momentum, much of the quantum nature of QCD becomes subdominant, such that emissions in QCD can be approximated by classical splitting functions [2–4]. These splitting functions form the basis for so-called parton showers, which build up arbitrarily complicated final states by starting with low-multiplicity final states and recursively splitting one final state parton into two. The recursive nature of parton showers allows them to be cast into a Markov chain algorithm. Parton showers form the basis for multi-purpose event generators like *Pythia* [5, 6], *Herwig* [7–9], *Ariadne* [10], *Isajet* [11], and *Sherpa* [12], which build up QCD-like events from simple hard-scattering matrix elements.

While parton shower algorithms easily generate high-multiplicity events, the resulting distributions are not based on full QCD, and for example do not accurately describe emissions at large transverse momentum. For this reason event generators based on full matrix element information have been constructed. Examples for generators based on tree-level calculations are *AlpGen* [13], *MadEvent* [14], *CompHEP* [15], *AMEGIC* [16], *Whizard* [17], and *Helac-Phegas* [18–20], while programs using next-to-leading order (NLO) calculations

also exist, examples being MCFM [21], NLOJet [22], PHOX [23], and VBFNLO [24]. In the last few years, significant progress has been made towards combining full QCD calculations, distributed with traditional phase space generators, with parton shower algorithms [25–37]. Encouraged by these successes, it is worth considering whether additional progress could be made by considering a fundamentally new approach to event generation.

In this paper, we introduce a novel use of the parton shower as a phase space generator for generic partonic calculations. The idea is to generate an event with a given final state multiplicity using a simple parton shower algorithm, and then to reweight this event to any desired partonic distribution. This algorithm is a key component of the **GenEvA** framework for event generation that we discuss in a companion paper [1], so we will refer to the phase space generator presented here as the **GenEvA** algorithm — for **Generate Events Analytically**.

The **GenEvA** framework asserts that the choice of partonic calculations should not be determined by algorithmic needs but rather by the underlying full theory information we want to describe. For this to be possible, the **GenEvA** framework requires a “projectable” phase space generator with a built-in notion of a matching scale, and the **GenEvA** algorithm is an example of such a generator. In the companion paper [1], we implement the following partonic results for $e^+e^- \rightarrow n$ jets:

1. **LO/LL**: Tree-level (LO) matrix elements with up to 6 final state partons, improved by leading logarithms (LL),
2. **NLO/LL**: Next-to-leading order (NLO) matrix elements including the virtual 2-parton diagram, improved by leading logarithms,
3. **NLO/LO/LL**: A combination of the above two results, where the NLO calculations are supplemented by higher-multiplicity tree-level calculation, all with leading-logarithmic resummation.

While both LO/LL [25–37] and NLO/LL [38–51] results have been implemented elsewhere, the third result has never been implemented before. Furthermore, these distributions are obtained without the need for negative-weight events.

From a technical point of view, reweighting a parton shower has become possible with the introduction of the analytic parton shower in ref. [52]. There, the probabilities to obtain individual branches are independent, so that the total probability of generating a given shower history is just the product of multiple splitting probabilities. This allows one to analytically calculate the exact probability for generating a given event. Because the parton shower covers Lorentz-invariant phase space multiple times, we also introduce a method to eliminate this double-counting problem while making maximal use of the full event sample given by the parton shower.

Some of the techniques in this paper are readily applicable to existing Monte Carlo programs, and could be implemented via a simple change in bookkeeping instead of a complete code rewrite. Indeed, the idea of reweighting parton showers to matrix elements appeared in the literature over 20 years ago [53–59, 26] and is implemented in Monte Carlo programs like **Pythia** [5, 6]. The **GenEvA** algorithm offers a generalization of this well-known

method. As we will see, some algorithmic and conceptual ideas are taken directly from **ALPGEN** [13] and **MadEvent** [14], so the code optimizations present in more mature Monte Carlo programs could aid in further increasing the efficiency of the **GenEvA** algorithm.

As in the companion paper [1], we will focus on the process $e^+e^- \rightarrow n$ jets, though in the conclusions in section 8 we discuss obvious generalizations to hadronic collisions with heavy resonances. In section 2, we introduce the main ingredients of the **GenEvA** algorithm: the parton shower probability, the Jacobian transformation to Lorentz-invariant phase space, the overcounting factor, and the desired matrix element. Sections 3, 4, 5, and 6 discuss these ingredients in turn, and section 6 also contains a discussion how the structure of the **GenEvA** algorithm allows for an automatic leading-logarithmic improvement of tree-level matrix elements. In section 7 we verify that the **GenEvA** algorithm works as a proper phase space generator, showing as an example that it can reproduce the results of traditional tree-level event generators, and discuss the efficiency of the algorithm. The reader not interested in the technical details may skip sections 3 through 6 and directly go from section 2 to section 7.

Before going into the details of the **GenEvA** algorithm, we first want to discuss why the parton shower is an appropriate and beneficial starting point for event generation.

1.2 Phase space from a parton shower

What does it mean to generate phase space from a parton shower? Ignoring the fact that the parton shower is derived from the soft-collinear limit of QCD, there are three main distinguishing features. First, it is well known that n -body phase space can be built recursively from $(n - 1)$ -body phase space, and a parton shower is an explicit implementation of this recursion. Second, a single parton shower run covers all of multiplicity, flavor, and phase space (hereafter referred to just as “phase space”) in one shot, which is an example of “single-channel” integration. Third, the parton shower preserves probability, meaning that during an event run, the fraction (or “grid”) of events of a given multiplicity, flavor, and kinematic structure will remain constant. Thus, as a phase space generator, the parton shower is a recursive, single-channel, fixed-grid algorithm.

The **GenEvA** algorithm is then diametrically opposed to modern phase space integrators like **MadEvent** [14], which use non-recursive, multi-channel, adaptive-grid algorithms, based on **Vegas** [60] and the **Bases/Spring** package [61]. While event generators like **CalcHEP/CompHEP** [62, 15] do use recursive methods (though not a parton shower) to generate phase space, algorithms such as **RAMBO** [63] and **SARGE** [64] were developed in order to *avoid* using a recursive definition of phase space because of problems of efficiency. Almost every phase space generator uses some notion of adaptation, in order to avoid the problem of large over- or under-sampling of phase space regions. And advanced phase space algorithms use multi-channel integration to create a single event sample culled from separate runs.

To see why **GenEvA** is a step forward rather than a step backward, it is important to note one crucial commonality between **MadEvent** and **GenEvA**: Both know about the singularity structure of QCD from $1/Q^2$ poles. In practice, we will see that for $e^+e^- \rightarrow n$ jets, **GenEvA** performs at or above **MadEvent** efficiencies. This suggests that the use of advanced non-recursive, multi-channel, adaptive-grid algorithms by other phase space

generators is compensating for something that **GenEvA** has already built-in, namely, an excellent approximation to QCD-like events through the parton shower. In other words, despite the fact that the recursive, single-channel, fixed-grid machinery of **GenEvA** is so primitive, the parton shower is sufficiently close to the right answer that more advanced methods are simply not necessary. Whether this fact will continue to be true in hadronic collisions is an open question.

While the machinery of **GenEvA** is simple-minded, the technical details needed to turn a parton shower into a phase space generator are not. The fact that the parton shower preserves probability means that one needs a thorough understanding of exactly what that probability distribution is, and the bulk of this paper is devoted to explaining in detail how the analytic parton shower in ref. [52] can become an efficient generator of Lorentz-invariant phase space.

For example, a crucial technical detail is that since the parton shower generates the entire phase space out of successive $1 \rightarrow 2$ splittings, it will usually distribute events with more partons in the final state than we have full QCD calculations available for. The **GenEvA** algorithm allows such events to be “truncated” to lower-multiplicity final states in a way that still allows for an analytic calculation of the resulting truncation probability. This is a concrete realization of the “phase space projection” requirement of the **GenEvA** framework [1], though the form implemented here is a “reversible projection” that has additional analytic properties beyond what is strictly speaking necessary in that context.

1.3 Benefits of **GenEvA**

There are numerous benefits to using the parton shower as a phase space generator, and many of these benefits actually derive from the fact that **GenEvA** is a recursive, single-channel, fixed-grid algorithm.

First, the parton shower is based on the symmetry structures of QCD. All possible flavor and color structures are generated with approximately the right probabilities, meaning that processes that are naturally linked, such as $e^+e^- \rightarrow q\bar{q} + n$ gluons and $e^+e^- \rightarrow q\bar{q}q'\bar{q}' + m$ gluons, need not have separate phase space integrations. The single-channel nature of the **GenEvA** algorithm makes the program trivial to parallelize, because a single **GenEvA** run contains a complete sampling of all desired physics processes.

Second, the parton shower is based on the singularity structures of QCD. **GenEvA** reproduces both the singular pole and leading-logarithmic behavior of full QCD distributions, yielding events with moderate weights. More importantly, the program distributes events with appropriate Sudakov factors [65] already, allowing for an automatic leading-log improvement of any tree-level matrix element. Of course, these generated Sudakov factors can also be calculated analytically and removed or exchanged if desired.

Third, because the phase space generation is recursive, a single event can have a nested set of consistent n -body phase space variables. This allows considerable flexibility in the types of parton-level calculations that can be handled, and allows a simple implementation of the NLO/LO/LL merged sample from the companion paper [1].

Fourth, a single event can have numerous different weights corresponding to different theoretical assumptions and tunable parameters. Using the feature of truncation, **GenEvA**

allows the same event to be interpreted as coming from an n -body or m -body matrix element with n not necessarily the same as m .¹ This makes it possible to have a single set of events that is capable of describing multiple theoretical distributions, greatly simplifying the process of parton-level systematic error estimation.

Moreover, by running this single event set through the required detector simulation, one obtains the equivalent of multiple detector-improved distributions, each with roughly the statistical significance of the full number of events. Given the large computational cost of full detector simulation, it is highly desirable to reuse Monte Carlo in this way.² Because the **GenEvA** algorithm includes the proper singularities and Sudakov factors out-of-the-box, this reweighting strategy will potentially be feasible even on large Monte Carlo sets.

Finally, **GenEvA** is well-suited to situations where it is difficult or tedious to distribute events according to a certain kind of theoretical calculation. For example, the application of soft-collinear effective theory (SCET) [66–69] to event generation [70–72] is most useful once the final state kinematics and matching scales μ are already known. If we are interested in calculating subleading logarithms using SCET, it is hard to analytically determine the proper SCET amplitudes over all of phase space. However, as a template for QCD-like events, **GenEvA** determines the kinematics and scales without reference to the amplitudes, allowing in principle numeric methods to be used to calculate SCET operator mixing.

Because of the above benefits, it is tempting to identify the **GenEvA** framework introduced in the companion paper [1] with the **GenEvA** algorithm that will be presented here. This is particularly true because we will see that in the **GenEvA** algorithm, it is actually easier and more efficient to include leading-logarithmic resummation in matrix elements than not. However, we emphasize that the **GenEvA** algorithm can be used as phase-space generator independently of any particular strategy for event generation, and similarly the **GenEvA** framework could work with an alternative phase space algorithm.

2. Overview of the **GenEvA** algorithm

The idea of the **GenEvA** algorithm is straightforward: Generate events using a parton shower and reweight these events to any desired distribution in a way that avoids double-counting. In this section, we develop the master formula for the event weight and then summarize the various ingredients used in this master formula. The details of the algorithm are then given in sections 3, 4, 5, and 6. Readers not interested in the technical details should get a basic understanding of the algorithmic issues involved from this section, and can then safely skip to section 7 for a comparison to other event generators and the efficiency of the algorithm.

2.1 The master formula

Let Φ denote a point in Lorentz-invariant phase space generalized to include all additional particle information, like flavor, helicity, and color, that fully characterize an event. We

¹In a traditional phase space generator, it is possible to reweight events distributed according to one n -body matrix element to a different n -body matrix element, but the value of n must be fixed.

²Because detector simulation is a linear operation, it is even possible to try different theoretical distributions *after* detector simulation.

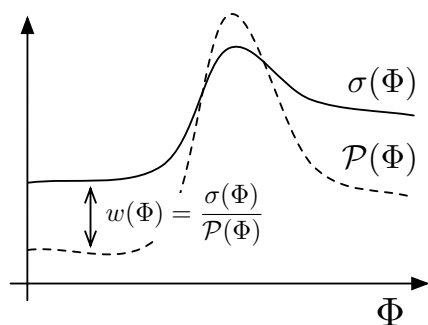


Figure 1: The simplest version of reweighting. Events are distributed according to a normalized probability distribution $\mathcal{P}(\Phi)$ and given weights $w(\Phi) = \sigma(\Phi)/\mathcal{P}(\Phi)$ such that the resulting weighted events are effectively distributed according to $\sigma(\Phi)$. In `GenEvA`, Φ corresponds to Lorentz-invariant phase space variables and $\sigma(\Phi)$ is a differential cross section.

want to distribute events according to a general distribution $\sigma(\Phi)$. That is, we want the contribution of an event with value Φ to be $d\sigma = \sigma(\Phi) d\Phi$. Even though the value of $\sigma(\Phi)$ is calculable for fixed values of Φ , it is usually very complicated to directly distribute events according to $\sigma(\Phi)$.

If we have a known normalized probability distribution $\mathcal{P}(\Phi)$ for which it is easy to distribute according to, we can choose values of Φ according to $\mathcal{P}(\Phi)$ and assign the weight

$$w(\Phi) = \frac{\sigma(\Phi)}{\mathcal{P}(\Phi)} \tag{2.1}$$

to each event with value Φ , as in figure 1. Since the probability of choosing a value Φ is $d\mathcal{P} = \mathcal{P}(\Phi) d\Phi$, the weighted set of events with values Φ will be distributed according to

$$d\sigma = w(\Phi) d\mathcal{P} = \frac{\sigma(\Phi)}{\mathcal{P}(\Phi)} \mathcal{P}(\Phi) d\Phi = \sigma(\Phi) d\Phi, \tag{2.2}$$

as desired.

Now imagine we have an easy way to distribute events in terms of a different variable Σ according to a known normalized distribution $\mathcal{P}(\Sigma)$, and there is a mapping $\Phi \equiv \Phi(\Sigma)$ which is one-to-one and onto, as in figure 2. Using $\mathcal{P}(\Sigma)$ as sampling function, the probability for choosing a value of Φ is now

$$d\mathcal{P} = \mathcal{P}(\Sigma) d\Sigma = \mathcal{P}(\Sigma) J(\Sigma) d\Phi \equiv \mathcal{P}(\Phi) d\Phi, \tag{2.3}$$

where $\Sigma \equiv \Sigma(\Phi)$ is the inverse of $\Phi(\Sigma)$ and

$$\frac{1}{J(\Sigma)} = \left| \frac{\partial(\Phi)}{\partial(\Sigma)} \right|(\Sigma) \tag{2.4}$$

is the Jacobian in the transformation from Σ to Φ . Thus, we obtain the event weight

$$w(\Phi) = \frac{\sigma(\Phi)}{\mathcal{P}[\Sigma(\Phi)] J[\Sigma(\Phi)]}. \tag{2.5}$$

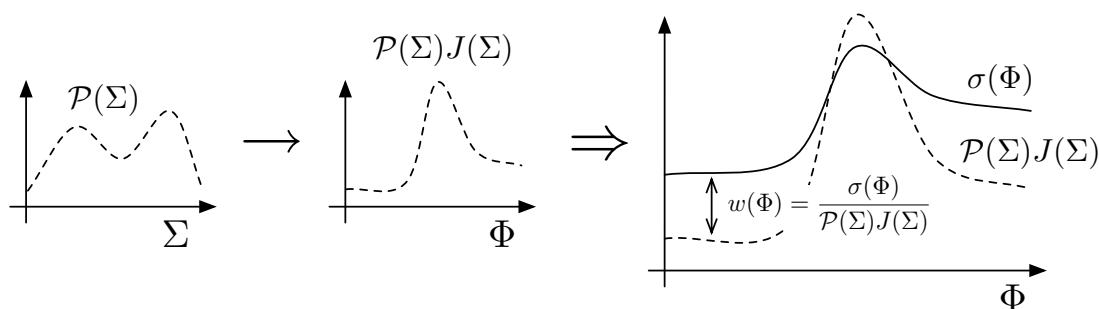


Figure 2: Reweighting using a different distribution variable Σ , where there is a one-to-one and onto map $\Sigma \rightarrow \Phi(\Sigma)$. Events are distributed according to a normalized probability distribution $\mathcal{P}(\Sigma)$, which together with the Jacobian $J(\Sigma)$ from the variable transformation $\Sigma \rightarrow \Phi(\Sigma)$ defines an effective probability distribution $\mathcal{P}(\Phi) \equiv \mathcal{P}(\Sigma)J(\Sigma)$. The weights $w(\Phi)$ can now be defined analogously to figure 1.

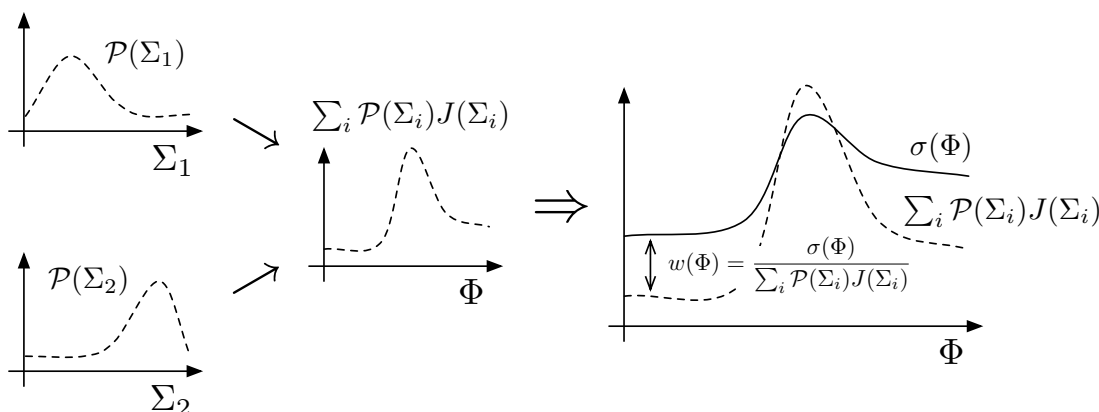


Figure 3: Reweighting when there are multiple values Σ_i that map to the same point $\Phi(\Sigma_i)$. If all Σ_i that map to a given Φ are known, then there is again an effective probability distribution $\mathcal{P}(\Phi) \equiv \sum_i \mathcal{P}(\Sigma_i)J(\Sigma_i)$ that can be used to define an appropriate weight $w(\Phi)$. In **GenEvA**, the parton shower provides a $\mathcal{P}(\Sigma)$, and it is well-known that multiple parton shower histories Σ_i can correspond to the same phase space point $\Phi(\Sigma_i)$. For this reweighting strategy to be computationally feasible, there must be an efficient way to calculate the sum over all parton shower histories $\sum_i \mathcal{P}(\Sigma_i)J(\Sigma_i)$.

It is important that the mapping $\Phi(\Sigma)$ is onto, otherwise there would be values of Φ which are never covered by $\mathcal{P}(\Sigma)$. On the other hand, $\Phi(\Sigma)$ need not be one-to-one. Instead, for a given value of Φ there could be a *discrete* number $n(\Phi)$ of Σ values which all map to that same Φ as in figure 3. In this case, the probability of choosing a value Φ is

$$d\mathcal{P} = \sum_{i=1}^{n(\Phi)} \mathcal{P}(\Sigma_i)J(\Sigma_i) d\Phi, \quad (2.6)$$

where here and in the following $\Sigma_i \equiv \Sigma_i(\Phi)$ denotes the i -th point Σ for which $\Phi(\Sigma_i) = \Phi$

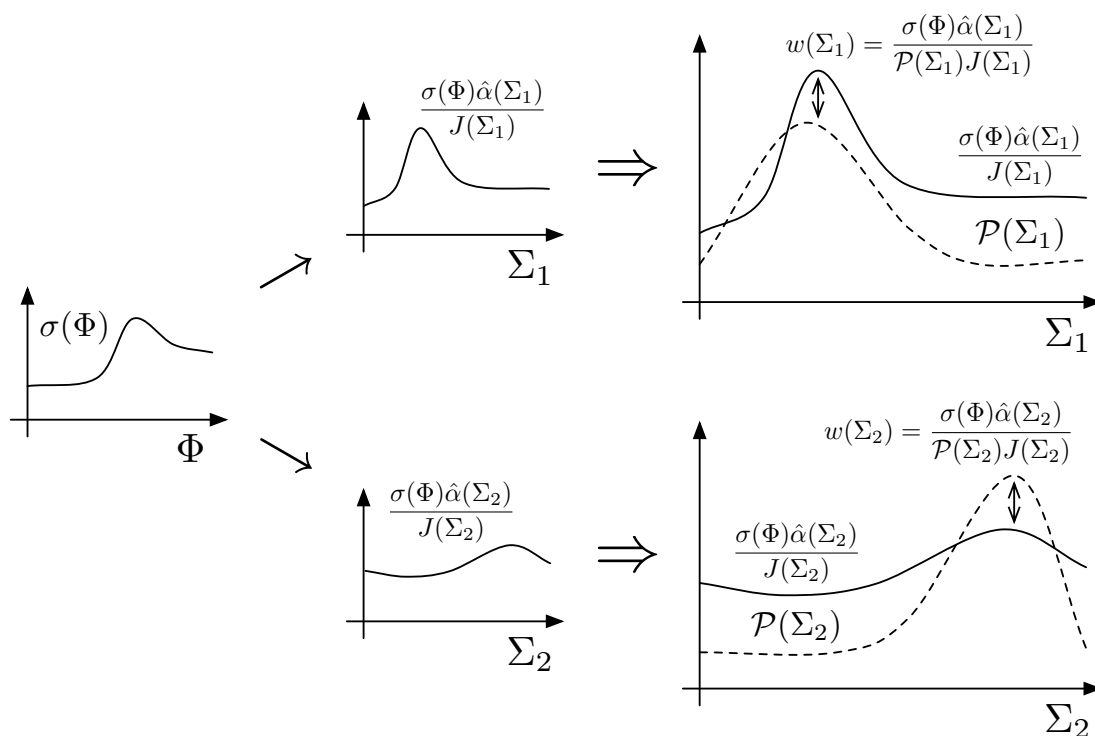


Figure 4: The reweighting approach used by **GenEvA** when the map $\Phi(\Sigma)$ is not one-to-one. Instead of taking the event weight to depend on Φ only, the event weight $w(\Sigma)$ is effectively a function of Σ . Using the overcounting factor $\hat{\alpha}(\Sigma_i)$, the desired distribution $\sigma(\Phi)$ is split up among the various Σ_i that map to the same Φ , including the Jacobian factor $J(\Sigma_i)$. As long as $\sum_i \hat{\alpha}(\Sigma_i) = 1$, the resulting events will be distributed according to $\sigma(\Phi)$ without the need to calculate $\mathcal{P}(\Sigma_i)J(\Sigma_i)$ for every Σ_i . In practice, there is a small computational cost to make sure that $\hat{\alpha}(\Sigma_i)$ is properly normalized.

and i runs from 1 to $n(\Phi)$. The weight as a function of Φ is thus

$$w(\Phi) = \frac{\sigma(\Phi)}{\sum_i \mathcal{P}[\Sigma_i(\Phi)]J[\Sigma_i(\Phi)]}. \quad (2.7)$$

In **GenEvA**, the parton shower provides the sampling function $\mathcal{P}(\Sigma)$, where Σ describes a complete parton shower history, and $\Phi(\Sigma)$ is given by the shower's final state. It is well-known that multiple parton shower histories Σ_i can produce the same phase space point $\Phi(\Sigma_i)$. In practice, it may be computationally expensive to calculate the sum in the denominator of eq. (2.7), as it requires reconstructing all possible parton shower histories Σ_i for a given point Φ . However, eq. (2.7) is dictated by our choice of characterizing events by Φ , which forces us to treat the weight of an event as a function of Φ as well. In particular, $w(\Phi)$ in eq. (2.7) cannot depend on the actual Σ_i which was used to generate the event.

To relax this condition we can simply characterize events by Σ and let $w \equiv w(\Sigma)$ be a function of Σ as in figure 4. In this case, the total contribution to $d\sigma$ for a given Φ is

$$d\sigma = \sum_i w(\Sigma_i) d\mathcal{P}_i = \sum_i \frac{\sigma(\Sigma_i)}{\mathcal{P}(\Sigma_i)} \mathcal{P}(\Sigma_i) d\Sigma = \sum_i \sigma(\Sigma_i) d\Sigma. \quad (2.8)$$

At this point, we need to define what we mean by $\sigma(\Sigma)$. From eq. (2.8) we have to require

$$d\sigma = \sum_i \sigma(\Sigma_i) J(\Sigma_i) d\Phi = \sigma(\Phi) d\Phi. \quad (2.9)$$

Note that simply taking $\sigma(\Sigma) = \sigma[\Phi(\Sigma)]/J(\Sigma)$ does not work, because it would lead to an overcounting by a factor of $n(\Phi)$ in eqs. (2.8) and (2.9). Instead, we need

$$\sigma(\Sigma) = \frac{\sigma[\Phi(\Sigma)] \hat{\alpha}(\Sigma)}{J(\Sigma)} \quad \text{with} \quad \sum_i \hat{\alpha}(\Sigma_i) = 1, \quad (2.10)$$

such that

$$d\sigma = \sum_i \sigma(\Sigma_i) J(\Sigma_i) d\Phi = \sigma(\Phi) \sum_i \hat{\alpha}(\Sigma_i) d\Phi = \sigma(\Phi) d\Phi, \quad (2.11)$$

as required.

Using eq. (2.10), the final weight as a function of Σ is

$$w(\Sigma) = \frac{\sigma(\Sigma)}{\mathcal{P}(\Sigma)} = \frac{\sigma[\Phi(\Sigma)] \hat{\alpha}(\Sigma)}{\mathcal{P}(\Sigma) J(\Sigma)}. \quad (2.12)$$

Eq. (2.12) is the master formula underlying the **GenEvA** algorithm. Alternatively, if we wish to specify $\sigma(\Sigma)$ directly, we use the notation $\sigma_i(\Phi)$, where i labels the different shower histories mapping to Φ , and the event weight is

$$w(\Sigma) = \frac{\sigma_i[\Phi(\Sigma)]}{\mathcal{P}(\Sigma) J(\Sigma)}. \quad (2.13)$$

It is clear that all the information in $\sigma_i(\Phi)$ could be absorbed into an ordinary $\sigma(\Phi)$ times an overcounting factor $\hat{\alpha}(\Sigma_i)$. However, note that the Jacobian $J(\Sigma)$ still appears in eq. (2.13), meaning that $\sigma_i(\Phi)$ is still a differential function of Φ and not Σ .

2.2 Calculating the event weight

The master formula in eq. (2.12) contains several ingredients. Here, we will give a brief overview of each of the components, while the details are discussed in the remainder of the paper.

2.2.1 The parton shower probability $\mathcal{P}(\Sigma)$

In the **GenEvA** algorithm, the parton shower plays the role of the sampling function $\mathcal{P}(\Sigma)$ for reweighting to any desired matrix element $\sigma(\Phi)$. The crucial point is that the parton shower automatically contains the right singularity structure of QCD matrix elements, which should in principle allow a reweighting with reasonable efficiency. However, for a parton shower to be used for the purposes of reweighting, it has to satisfy several additional requirements.

First, we have to know the parton shower's probability distribution $\mathcal{P}(\Sigma)$, which means we must be able to compute the exact probability with which the parton shower generates a given shower history Σ . The analytic parton shower algorithm constructed in ref. [52] has this property and is the algorithm used here and detailed in section 3. The key feature

of this shower is that energy-momentum is conserved by construction without any *ad hoc* momentum shuffling.

Second, the parton shower has to cover all of phase space. Of course, almost any parton shower already covers all possible final state configurations, in that it will generate 2-body final states, 3-body final states, and so on, with the flavors of the final state particles representing all possible allowed configurations. What is important is that the shower also has no kinematic dead zones, to ensure that the map $\Phi(\Sigma)$ is indeed onto and all of phase space is covered as necessary.

Finally, as already mentioned, the fact that the parton shower generates many final state particles seems at first glance like a liability, because matrix elements for n -body configurations with arbitrary n are not readily available. In order to reweight the generated events, we need to be able to “backtrack” (not necessarily uniquely) a high-multiplicity final state to a lower-multiplicity intermediate state, while still being able to calculate analytically the probability with which this intermediate state was generated. We call this feature “truncation”, and it is possible because of the analytic properties of the algorithm of ref. [52] in conjunction with the parton shower’s built-in notion of probability preserving ordering. This truncation is a concrete implementation of the phase space projection required by a phase space generator suitable for the **GenEvA** framework as explained in more detail in ref. [1].

2.2.2 The Jacobian $J(\Sigma)$

An event produced by **GenEvA**’s parton shower is characterized by a parton shower history Σ that consists of the set of $1 \rightarrow 2$ splittings that build up the event. In contrast, Lorentz-invariant phase space Φ characterizes an event by the set of three-momenta (symmetrized appropriately with respect to identical particles) of the on-shell final state particles. As needed for reweighting, there is a map from parton shower histories Σ to Lorentz-invariant phase space Φ , such that for every parton shower history Σ there is a unique point $\Phi(\Sigma)$. A nontrivial Jacobian factor $J(\Sigma)$ is required in changing variables from the natural variables used to describe the $1 \rightarrow 2$ splittings in the shower history to those that describe Lorentz-invariant phase space. This Jacobian will be given in section 4.

2.2.3 The overcounting factor $\hat{\alpha}(\Sigma)$

It is well known that the parton shower does not map out phase space in a one-to-one way, because in general, multiple parton shower histories can lead to the same final state four-momenta. The overcounting factor $\hat{\alpha}(\Sigma)$ takes care of this overcounting, and essentially determines how $\sigma(\Phi)$ is split up among the different $\Sigma_i(\Phi)$ that map to the same point Φ . Since we are eventually only interested in $\sigma(\Phi)$, we can choose $\hat{\alpha}(\Sigma)$ freely only subject to the constraint in eq. (2.10). The simplest possible choice would be $\hat{\alpha}(\Sigma) = 1/n(\Phi)$, where $n(\Phi)$ counts the number of Σ_i mapping to the same Φ . This choice will in general lead to a poor statistical efficiency, since the weights $w(\Sigma)$ will vary widely for the different Σ_i contributing to the same point Φ . To ensure well-behaved statistics, we need to ensure that for a given phase space point Φ , all event weights are roughly the same, thus requiring that

$\hat{\alpha}(\Sigma)$ scales like $\mathcal{P}(\Sigma)J(\Sigma)$. We will discuss the choice of $\hat{\alpha}(\Sigma)$ in the context of **GenEvA** in detail in section 5.

2.2.4 The matrix element $\sigma[\Phi(\Sigma)]$

GenEvA can distribute events according to nearly any distribution $\sigma(\Phi)$, and this quantity will depend on the underlying physics one is trying to describe. Crucially, physics considerations alone can determine the best choice for $\sigma(\Phi)$, and the algorithmic details of **GenEvA** need not influence this choice.

While $\sigma(\Phi)$ in eq. (2.12) only depends on a set of on-shell four-momenta, in many cases it can be advantageous to use eq. (2.13) and directly work in terms of $\sigma_i(\Phi)$. That is, we may want to assign different weights to events having different parton shower histories Σ_i even though they contribute to the same point $\Phi(\Sigma_i)$. Two reasons to use $\sigma_i(\Phi)$ are to improve the efficiency of **GenEvA** and to resolve logarithmic ambiguities [1]. In section 6, we will discuss tree-level fixed-order QCD matrix elements using $\sigma(\Phi)$, as well as calculations which properly sum the leading-logarithmic dependence, using $\sigma_i(\Phi)$.

2.3 Reweighting efficiency

To assess the practical usefulness of the **GenEvA** algorithm, it is important to define what we mean by efficiency. In this paper, we will use statistical efficiency as the figure of merit in contrast to unweighting efficiency. Since unweighting efficiency depends on the value of the maximum weight generated, it is sensitive to how many events were generated in a given run. Here we argue that statistical efficiency is a much better measure of the strength of an algorithm.

If N is the total number of events, then in the limit of large N the weights satisfy³

$$\sum_n w(\Phi) \rightarrow N \int d\Phi \sigma(\Phi). \quad (2.14)$$

To see this, consider the normalized distribution $\hat{\sigma}(\Phi) = \sigma(\Phi) / \int d\Phi \sigma(\Phi)$. Since both $\hat{\sigma}(\Phi)$ and $\mathcal{P}(\Phi)$ are normalized to unity, the normalized weights $\hat{w}(\Phi) = \hat{\sigma}(\Phi) / \mathcal{P}(\Phi)$ satisfy $\sum_n \hat{w}(\Phi) \rightarrow N$, from which eq. (2.14) follows.

Dividing eq. (2.14) by N we have

$$\frac{1}{N} \sum_n w(\Phi) \equiv \langle w(\Phi) \rangle \rightarrow \int d\Phi \sigma(\Phi), \quad (2.15)$$

where $\langle \cdot \rangle$ denotes the sample average. From the standard error on averages, the statistical uncertainty on the Monte Carlo estimate of $\int d\Phi \sigma(\Phi)$ is thus given by

$$\frac{\sqrt{\langle w(\Phi)^2 \rangle - \langle w(\Phi) \rangle^2}}{\sqrt{N}}. \quad (2.16)$$

³This definition of weights differs slightly from the one commonly used in Monte Carlo programs. Here, the weight of an event is independent of the number of events generated, whereas most Monte Carlo programs divide individual event weights by N , the total number of events generated in a run. We will use the run-independent definition of weights throughout this paper.

The numerator describes the width of the distribution of weights, which vanishes in the limit of all weights being equal. This is equivalent to the fact that for unweighted events, the uncertainty on the total number of events is always zero. Thus, the ratio

$$\eta_{\text{eff}} = \frac{\langle w(\Phi) \rangle^2}{\langle w(\Phi)^2 \rangle} = \frac{1}{N} \frac{[\sum_n w(\Phi)]^2}{\sum_n w(\Phi)^2} \equiv \frac{N_{\text{eff}}}{N} \leq 1, \quad (2.17)$$

provides a statistical measure of the efficiency of the reweighting. We can interpret N_{eff} as the effective number of events after reweighting in the sense that the sample of reweighted events is statistically equivalent to an unweighted sample of N_{eff} events. In the limit of all weights being equal $\eta_{\text{eff}} \rightarrow 1$ and $N_{\text{eff}} \rightarrow N$. Hence, this way of distributing events by reweighting $\mathcal{P}(\Phi)$ to $\sigma(\Phi)$ is efficient as long as $\mathcal{P}(\Phi)$ is reasonably close to $\hat{\sigma}(\Phi)$. For example, if we wish to distribute events according to a function $\sigma(\Phi)$ which has large peaks, we have to choose a sampling function $\mathcal{P}(\Phi)$ that has a similar peak structure.

The definition of N_{eff} differs from that of N_{unw} , which is the number of events that would remain after unweighting the sample,

$$N_{\text{eff}} = \frac{[\sum_n w(\Phi)]^2}{\sum_n w(\Phi)^2} \quad \text{vs.} \quad N_{\text{unw}} = \frac{\sum_n w(\Phi)}{\max[w(\Phi)]}. \quad (2.18)$$

The unweighting efficiency $\eta_{\text{unw}} = N_{\text{unw}}/N$ is highly sensitive to the tail of the weight distribution, such that even an exponentially small number of high-weight events can cause a dramatic decrease in η_{unw} . The only way to deal with this is to throw away events with too large weights, which of course introduces a systematic error which is very hard to quantify. In contrast, the statistical efficiency N_{eff}/N correctly reflects the full statistical power of the event sample, and being based on sample averages, is unaffected by individual weights.

The main disadvantage of considering the statistical efficiency is that since the statistical uncertainties are determined by N_{eff} and not by N , it is necessary to store and process a larger total number of events than in a statistically equivalent unweighted sample. However, if this becomes a serious issue one has always the option to partially unweight the sample by unweighting just those events that have small weights. That is, we can define a threshold weight w_0 , and only unweight those events with weights smaller than w_0 . The expected statistical efficiency of the partially unweighted sample is given by

$$\eta_{\text{eff}}(w_0) = \frac{1}{N} \frac{[\sum_n w(\Phi)]^2}{\sum_{w>w_0} w(\Phi)^2 + w_0 \sum_{w<w_0} w(\Phi)}. \quad (2.19)$$

Note that $\eta_{\text{eff}}(0) = \eta_{\text{eff}}$ and $\eta_{\text{eff}}(w_{\text{max}}) = \eta_{\text{unw}}$. In section 7.3, we will report the efficiency of **GenEvA** not only in terms of η_{eff} , but also in terms of the time needed to achieve a partially unweighted sample with $\eta_{\text{eff}}(w_0) = 0.9$.

3. The analytic parton shower

As discussed in section 2.2.1, a parton shower to be used as phase space generator has to fulfill three requirements: 1) It has to have an analytically calculable probability $\mathcal{P}(\Sigma)$;

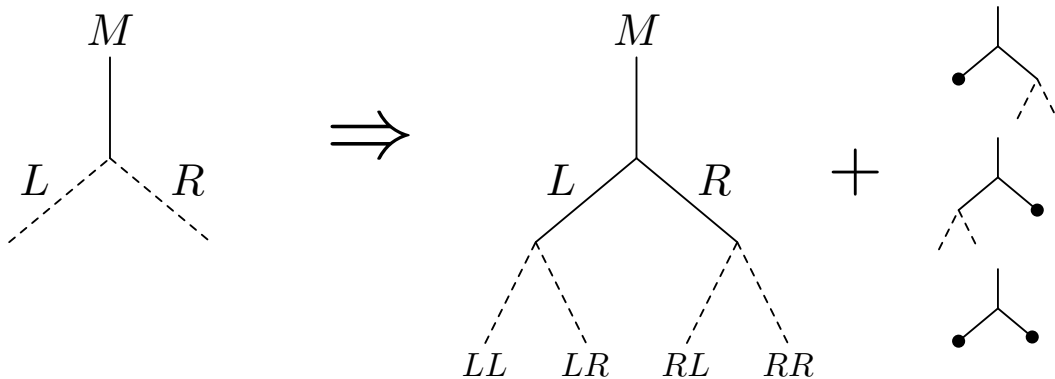


Figure 5: The structure of the parton shower. A solid line indicates an already processed particle, while a dashed line indicates a particle that has not yet been processed. A dot at the end of a solid line indicates a particle that did not branch above the shower cutoff, i.e. a final state particle. Starting from an existing branch $M \rightarrow LR$, where a mother particle M has branched into daughter particles L and R , **GenEvA**'s parton shower generates the double branch $M \rightarrow LR \rightarrow (LL/LR)(RL/RR)$, where it is understood that there is also a finite probability for L or R to remain unbranched. Understanding how the L and R branches are coupled is essential for being able to exactly compute the double-branch probability and to truncate the parton shower.

2) it has to cover all of phase space; and 3) we have to be able to analytically truncate the shower, which means we must be able to truncate any high-multiplicity final state produced by the shower to an n -body final state for any given n , while retaining analytic control over the truncation probability. In the following we will describe in detail how these requirements are met by **GenEvA**'s parton shower.

3.1 Overview

The parton shower used by the **GenEvA** algorithm to generate phase space is based on the analytic parton shower algorithm constructed in ref. [52]. As we will only be discussing $e^+e^- \rightarrow n$ jets, we will only discuss the final state parton shower. We comment on the extension to hadronic collisions and initial state showers in section 8.1.

Since the main purpose of **GenEvA**'s parton shower is to generate n -body phase space efficiently, the current implementation is only accurate to leading-logarithmic accuracy, and neglects several important physics effects a realistic parton shower should take into account, such as the running of α_s and color coherence [55, 56, 73]. In addition, the shower we consider here treats all final state particles as massless. While it is straightforward to construct an algorithm that does incorporate all of these physics issues and still maintains analytic control (see the discussion in ref. [52]), for the purposes of this paper the simple analytic parton shower discussed here will suffice.

An event produced by **GenEvA**'s parton shower is characterized by the kinematics of the initial hard-scattering process (typically a $2 \rightarrow 2$ interaction), followed by a set of $1 \rightarrow 2$ particle splittings. We mostly follow the notation of ref. [52]. We call a given $1 \rightarrow 2$ splitting a branch, with the original particle called the mother particle, and the two

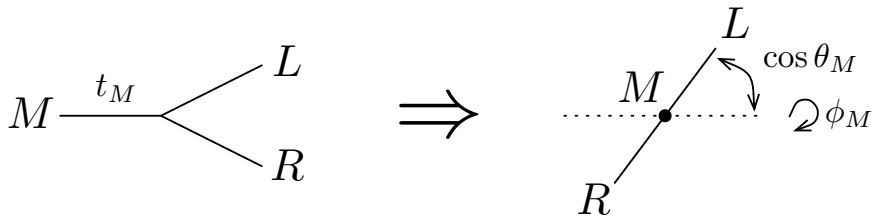


Figure 6: The kinematic variables describing a single splitting in the parton shower. The splitting functions can be chosen to be functions of the mother’s virtuality t_M and the angles $\cos\theta_M$ and ϕ_M , where $\cos\theta_M$ is the angle between the L daughter and the boost axis to the mother’s rest frame, and ϕ_M is the azimuth around the boost axis. Note that the L and R four-momenta cannot be determined until t_L and t_R are known, that is, only after the daughters themselves have been processed.

final particles called the daughter particles. We label the mother particle by M , and the daughter particles by L and R . The daughters of these daughters (the granddaughters) are labeled by LL , LR , RL , and RR .

Starting from an existing branch $M \rightarrow LR$, the shower generates a double branch $M \rightarrow LR \rightarrow (LL/LR)(RL/RR)$. This is illustrated in figure 5, where here and in the following figures a solid line indicates an already processed particle, while a dashed line indicates a particle that has not yet been processed. A dot at the end of a solid line indicates a particle that did not branch above the shower cutoff, i.e. a final state particle. The double-branch probability, i.e. the probability to obtain a specific $1 \rightarrow 2 \rightarrow 4$ double branch from a given $1 \rightarrow 2$ single branch, is the fundamental object in **GenEvA**’s parton shower, and allows it to have a well-defined notion of global evolution.

3.2 Choice of kinematic variables

In general, the splitting functions describing the $1 \rightarrow 2$ parton splittings depend on three kinematic variables, whose precise definitions depend on the details of the parton shower. One of these variables is the evolution variable of the parton shower, which is always decreasing as the shower progresses, while the other two describe the remaining kinematics of the splitting. There are several choices possible for the evolution variables, such as virtuality [56, 59], the energy-weighted emission angle [74, 7], transverse momentum [75, 76, 10], or more general variables [34].

In **GenEvA** we use the analytic parton shower algorithm defined in ref. [52], in which the evolution variable is the virtuality t_M of the mother, and the remaining two variables are chosen to be the angle $\cos\theta_M$ in the mother’s rest frame between the L daughter and the boost axis, and the azimuthal angle ϕ_M around this boost axis, as shown in figure 6. Thus, we can write the differential kinematics of the parton shower in terms of

$$d\Sigma_2(M \rightarrow LR) = dt_M d\cos\theta_M d\phi_M, \quad (3.1)$$

with the phase space limits

$$-1 \leq \cos\theta_M \leq 1, \quad 0 \leq \phi_M \leq 2\pi, \quad 0 \leq \sqrt{t_L} + \sqrt{t_R} \leq \sqrt{t_M}. \quad (3.2)$$

The primary reason to use the angles as the basic variables is that their phase space limits are trivial and completely independent of any other kinematic quantity, which also makes it trivial to satisfy energy-momentum conservation for each branch.

A complete shower history with n final state partons is described by

$$d\Sigma_n = d\Phi_2^{(0)} \prod_{i=1}^{n-2} d\Sigma_2(M_i \rightarrow L_i R_i), \quad (3.3)$$

where $d\Phi_2^{(0)}$ is the two-body phase space of the initial hard scattering and $d\Sigma_2(M_i \rightarrow L_i R_i)$ are the subsequent branches from the shower. We will discuss the precise phase space measure and its relation to ordinary n -body phase space in more detail in section 4.

Instead of the angle $\cos\theta_M$, many parton showers use the energy splitting $z_M = E_L/E_M$, describing how the energy of the mother is shared between the two daughters. Indeed, the splitting functions themselves are usually written in terms of z . The relation between z and $\cos\theta$ is given by

$$\begin{aligned} \cos\theta_M &= \frac{1}{\beta_M} \frac{2z_M - (1 + t_L/t_M - t_R/t_M)}{\lambda(t_M; t_L, t_R)}, \\ z_M &= \frac{1}{2} \left[1 + \frac{t_L}{t_M} - \frac{t_R}{t_M} + \beta_M \cos\theta_M \lambda(t_M; t_L, t_R) \right], \end{aligned} \quad (3.4)$$

where t_M is the virtuality of the mother, $t_{L,R}$ are the virtualities of the daughters, and we have defined

$$\beta_M = \sqrt{1 - \frac{t_M}{E_M^2}}, \quad \lambda(t_M; t_L, t_R) = \frac{1}{t_M} \sqrt{(t_M - t_L - t_R)^2 - 4t_L t_R}, \quad (3.5)$$

with E_M being the energy of the mother.

In **GenEvA**'s shower, $\cos\theta$ is the fundamental splitting variable, while z is a derived quantity. However, since the usual splitting functions are functions of z , we need to translate $\cos\theta$ into z . At the time the mother is branched, both daughters are still massless, so eq. (3.4) gives the relation

$$z(\cos\theta_M, E_M) = \frac{1}{2} + \frac{1}{2}\beta_M(E_M) \cos\theta_M, \quad (3.6)$$

where we made the dependence on E_M explicit. The problem is that the final value for the mother's energy E_M cannot be determined until after the mother itself and its sister have been processed and their final invariant masses are known. This is equivalent to the fact that in ordinary parton showers based on $1 \rightarrow 2$ splittings, the value of z_M has to be "re-shuffled" at a later stage in the algorithm in order to satisfy momentum conservation. For a more detailed discussion see ref. [52]. In contrast, once a value for $\cos\theta_M$ is determined, it never has to be changed afterwards, which crucially relies on the fact that its phase space limits in eq. (3.2) are independent from the rest of the shower. Nevertheless, we still need to pick a value of E_M in eq. (3.6), which we denote as E_M^{start} . From eq. (3.6) we see that the choice of E_M^{start} effectively determines the range of z allowed in the splitting

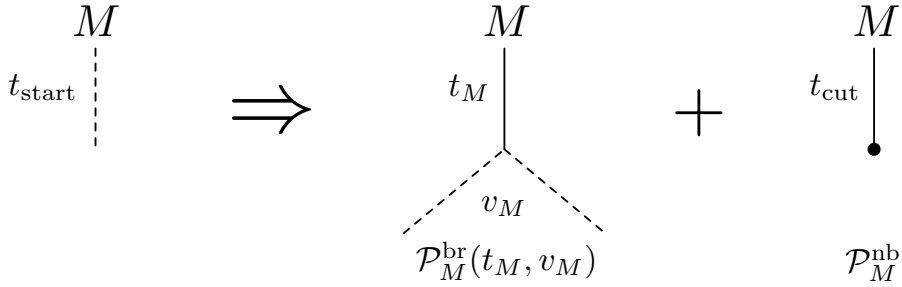


Figure 7: The single-branch probability. Starting at the scale t_{start} , the mother M can either branch at the scale $t_M > t_{\text{cut}}$ or remain unbranched down to the scale t_{cut} , which is denoted by a dot at the end of a line. The branching probability $\mathcal{P}_M^{\text{br}}(t_M, v_M)$ and the no-branching probability $\mathcal{P}_M^{\text{nb}}$ also depend on t_{start} , t_{cut} , and E_M^{start} . The variable v_M is a shorthand for the splitting angles and the daughters' flavors.

functions. Therefore, to avoid artificially large subleading-logarithmic effects, one should pick an E_M^{start} of the order of E_M . In our case, the precise choice of E_M^{start} is dictated by the requirement to have a simple analytic truncation, and will be discussed in section 3.5.

We are now ready to obtain the required splitting functions. Taking into account the appropriate Jacobian factors, we have

$$f_{M \rightarrow LR}(t_M, \cos \theta_M, \phi_M; E_M^{\text{start}}) = \frac{\alpha_s}{2\pi} \frac{\beta_M}{2} \frac{1}{2\pi} \frac{1}{t_M} P_{M \rightarrow LR}[z(\cos \theta_M, E_M^{\text{start}})], \quad (3.7)$$

where $P_{M \rightarrow LR}(z)$ are the well-known Altarelli-Parisi [3] splitting functions

$$\begin{aligned} P_{q \rightarrow qg}(z) &= C_F \frac{1+z^2}{1-z}, \\ P_{g \rightarrow gg}(z) &= C_A \left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z) \right], \\ P_{g \rightarrow q\bar{q}}(z) &= T_R [z^2 + (1-z)^2]. \end{aligned} \quad (3.8)$$

In the following, to simplify the notation, we will mostly suppress the dependence of the splitting functions on E_M^{start} , but one should keep in mind that wherever the splitting functions appear there is also an implicit dependence on E_M^{start} .

3.3 The single-branch probability

The basic ingredient in the parton shower is the single-branch probability, which determines if and how a given mother particle M splits into two daughters L and R , as illustrated in figure 7. The probability for a single branch to occur with given values of the splitting variables is a combination of the splitting function and a Sudakov factor (for a review of parton showers see for example ref. [6]):

$$\begin{aligned} \mathcal{P}_{M \rightarrow LR}^{\text{br}}(t_M, \cos \theta_M, \phi_M; t_{\text{start}}, t_{\text{cut}}) &= f_{M \rightarrow LR}(t_M, \cos \theta_M, \phi_M) \Delta_M(t_{\text{start}}, t_M) \\ &\quad \times \theta(t_{\text{start}} > t_M) \theta(t_M > t_{\text{cut}}). \end{aligned} \quad (3.9)$$

The Sudakov factor is defined as

$$\Delta_M(t_2, t_1) = \exp \left[- \int_{t_1}^{t_2} dt \int_{-1}^1 d\cos\theta \int_0^{2\pi} d\phi \sum_X f_{M \rightarrow X}(t, \cos\theta, \phi) \right], \quad (3.10)$$

where X denotes all possible flavor structures allowed for the splitting of the mother M . The Sudakov factor can be interpreted as a no-branching probability, since the probability of obtaining a value $t_M \leq t_{\text{start}}$ is given by the probability to not branch anywhere between t_M and t_{start} times the probability to branch at t_M , which is given by the splitting function evaluated at t_M . Thus the probability of having no branching anywhere above t_{cut} , where t_{cut} is the cutoff scale of the parton shower, is determined by the no-branching probability directly

$$\mathcal{P}_M^{\text{nb}}(t_{\text{start}}, t_{\text{cut}}) = \Delta_M(t_{\text{start}}, t_{\text{cut}}). \quad (3.11)$$

For simplicity, we will frequently use the shorthand notation

$$\mathcal{P}_M^{\text{br}}(t_M, v_M; t_{\text{start}}, t_{\text{cut}}) \equiv \mathcal{P}_{M \rightarrow LR}^{\text{br}}(t_M, \cos\theta_M, \phi_M; t_{\text{start}}, t_{\text{cut}}), \quad (3.12)$$

where v_M denotes all variables other than the virtuality required to describe a single splitting. Thus, v_M contains the angular variables $\cos\theta_M$ and ϕ_M , as well as the flavor information $M \rightarrow LR$. We will also use the notation

$$\int dv_M \equiv \int_{-1}^1 d\cos\theta \int_0^{2\pi} d\phi \sum_X. \quad (3.13)$$

Of course, the probability for a particle to do anything (branch at any $t_M > t_{\text{cut}}$ or not branch above t_{cut}) is equal to unity

$$\int_{t_{\text{cut}}}^{t_{\text{start}}} dt_M \int dv_M \mathcal{P}_M^{\text{br}}(t_M, v_M; t_{\text{start}}, t_{\text{cut}}) + \mathcal{P}_M^{\text{nb}}(t_{\text{start}}, t_{\text{cut}}) = 1. \quad (3.14)$$

To see this explicitly, note that the Sudakov factor satisfies

$$\begin{aligned} \frac{d}{dt_1} \Delta_M(t_2, t_1) &= \Delta_M(t_2, t_1) \int_{-1}^1 d\cos\theta \int_0^{2\pi} d\phi \sum_X f_{M \rightarrow X}(t_1, \cos\theta, \phi) \\ &= \int dv_M \mathcal{P}_M^{\text{br}}(t_1, v_M; t_2, t_{\text{cut}}), \end{aligned} \quad (3.15)$$

where we are assuming that $t_1 > t_{\text{cut}}$. This implies

$$\int_{t_1}^{t_2} dt_M \int dv_M \mathcal{P}_M^{\text{br}}(t_M, v_M; t_{\text{start}}, t_{\text{cut}}) = \Delta_M(t_{\text{start}}, t_2) - \Delta_M(t_{\text{start}}, t_1), \quad (3.16)$$

from which eq. (3.14) follows.

3.4 The double-branch probability

GenEvA's parton shower algorithm always acts on an existing $M \rightarrow LR$ branch. It takes both the L and R daughters of the branch and considers their branching into two sets of granddaughters LL/LR and RL/RR . The shower algorithm is illustrated in figure 8, and works as follows:

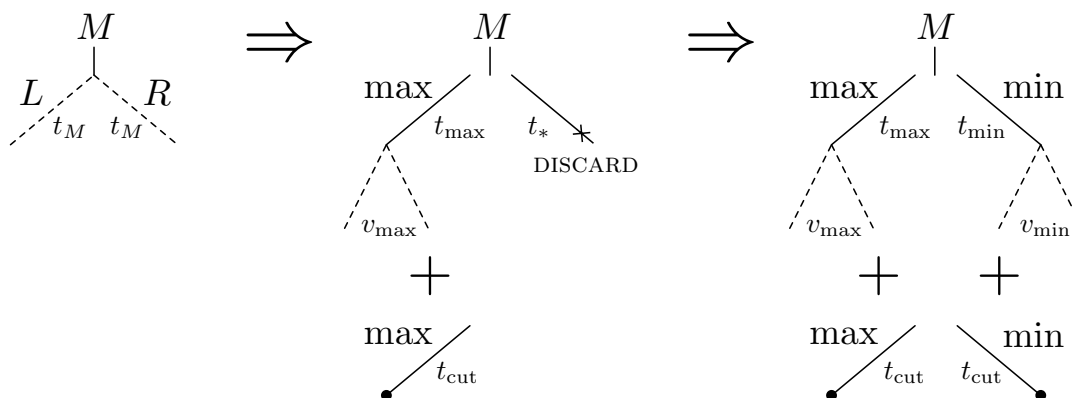


Figure 8: The GenEvA shower algorithm. Starting at the scale t_M , trial L and R branches are generated, but only the one with the larger of t_L and t_R is kept. Then, starting from the new scale $t_* = \min[t_{\max}, (\sqrt{t_M} - \sqrt{t_{\max}})^2]$, the other branch is generated. This coupling between the L and R branches is essential to obtain an analytic truncation, and means that the GenEvA shower has a notion of global evolution.

1. Pick a random branch with two unprocessed daughters L and R and a mother particle M with invariant mass t_M .
2. Choose $E_{L,R}^{\text{start}}$, set $t_{\text{start}} = t_M$, and determine values of $\{t_L, v_L\}$ and $\{t_R, v_R\}$ independently according to the single-branch probability $\mathcal{P}_i^{\text{br}}(t_i, v_i; t_M, t_{\text{cut}})$, where $i = L, R$. If $t_i < t_{\text{cut}}$, leave that particle unbranched.
3. Keep the branch of the daughter with the larger t , call this the max-daughter, and define $t_{\max} = \max(t_L, t_R)$. Discard the branch of the min-daughter.
4. Determine new values $\{t_{\min}, v_{\min}\}$ for the min-daughter according to $\mathcal{P}_{\min}^{\text{br}}(t_{\min}, v_{\min}; t_*, t_{\text{cut}})$, with t_{start} set to

$$t_* = \min[t_{\max}, (\sqrt{t_M} - \sqrt{t_{\max}})^2]. \quad (3.17)$$

If $t_{\min} < t_{\text{cut}}$, leave that particle unbranched.

5. Continue until all branches are processed.

Since eq. (3.9) contains no restrictions on the angles $\cos\theta$ and ϕ , their phase space is trivially covered. The range of $t_{L,R}$ allowed by the algorithm is

$$t_{\text{cut}} \leq t_{L,R} \leq t_M, \quad \sqrt{t_L} + \sqrt{t_R} \leq \sqrt{t_M}, \quad (3.18)$$

where the second condition is enforced by t_* in step 4 of the algorithm, as required by eq. (3.2). Hence, apart from the infrared shower cutoff t_{cut} , the algorithm covers all of phase space without leaving any dead zones. In particular, a parton shower history with a cutoff value t_{cut} and n particles in the final state covers the entire n -body phase space with the additional restriction that the minimum virtuality of any two particles is greater than

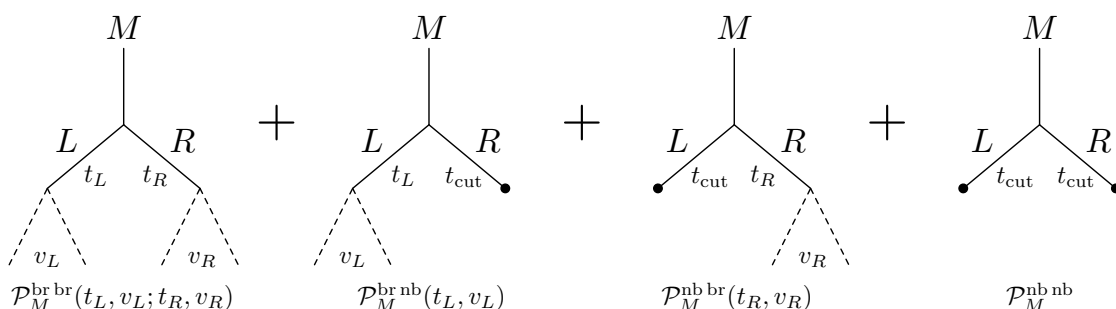


Figure 9: The double-branch probability. The **GenEvA** algorithm in figure 8 yields four different branching/no-branching combinations. They are built out of products of single-branch probabilities, with nontrivial interdependence between t_L and t_R .

t_{cut} .⁴ As there are QCD singularities associated with two partons getting close in virtuality, the t_{cut} restriction is actually helpful to regulate infrared divergences in tree-level diagrams.

The above algorithm results in an analytically calculable parton shower. It is clear that the algorithm couples the splitting of the L and the R daughters together, resulting in a double-branch probability

$$\begin{aligned} \mathcal{P}_M(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}) \\ \equiv \mathcal{P}_{M \rightarrow LR \rightarrow (LL/LR)/(RL/RR)}(t_L, \cos \theta_L, \phi_L; t_R, \cos \theta_R, \phi_R; t_M, t_{\text{cut}}), \end{aligned} \quad (3.19)$$

where $t_{L,R}$ are the final virtualities selected for the daughters, and $v_{L,R}$ denote all other variables describing the daughters' splittings, including the flavor and color structure for the LL , LR , RL , and RR granddaughters.

There are four possible cases for the double-branch probability as shown in figure 9: 1) Both the left and right daughters branch; 2) the left daughter branches and the right one does not; 3) the right daughter branches and the left one does not; and 4) neither the left nor the right daughter branches. Following the shower algorithm, the double-branch probability for case 1) is given by

$$\begin{aligned} \mathcal{P}_M^{\text{br br}}(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}) \\ = \theta(t_L > t_R) \left[\mathcal{P}_L^{\text{br}}(t_L, v_L; t_M, t_{\text{cut}}) \times \Delta_R(t_M, t_L) \mathcal{P}_R^{\text{br}}(t_R, v_R; t_*, t_{\text{cut}}) \right] \\ + \theta(t_R > t_L) \left[\mathcal{P}_R^{\text{br}}(t_R, v_R; t_M, t_{\text{cut}}) \times \Delta_L(t_M, t_R) \mathcal{P}_L^{\text{br}}(t_L, v_L; t_*, t_{\text{cut}}) \right], \end{aligned} \quad (3.20)$$

where the extra factors of $\Delta_{L,R}$ arise from discarding the first attempt to determine the smaller t value. For the other three cases we have

$$\mathcal{P}_M^{\text{br nb}}(t_L, v_L; t_M, t_{\text{cut}}) = \mathcal{P}_L^{\text{br}}(t_L, v_L; t_M, t_{\text{cut}}) \times \Delta_R(t_M, t_L) \Delta_R(t_*, t_{\text{cut}}),$$

⁴Two partons from different branches can also accidentally get closer than t_{cut} , so there are additional regions of n -body phase space that are covered in practice. We will discuss this issue more in section 6.1.

$$\begin{aligned}
 \mathcal{P}_M^{\text{nb br}}(t_R, v_R; t_M, t_{\text{cut}}) &= \mathcal{P}_R^{\text{br}}(t_R, v_R; t_M, t_{\text{cut}}) \times \Delta_L(t_M, t_R) \Delta_L(t_*, t_{\text{cut}}), \\
 \mathcal{P}_M^{\text{nb nb}}(t_M, t_{\text{cut}}) &= \Delta_L(t_M, t_{\text{cut}}) \times \Delta_R(t_M, t_{\text{cut}}),
 \end{aligned}
 \tag{3.21}$$

where it is understood that if $t_* < t_{\text{cut}}$, then $\mathcal{P}_i^{\text{br}}(t_i, v_i; t_*, t_{\text{cut}}) = 0$ and $\Delta_i(t_*, t_{\text{cut}}) = 1$.

We can formally combine all four cases by writing

$$\begin{aligned}
 \mathcal{P}_M(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}) &= \mathcal{P}_M^{\text{br br}}(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}) \theta(t_L > t_{\text{cut}}) \theta(t_R > t_{\text{cut}}) \\
 &+ \mathcal{P}_M^{\text{br nb}}(t_L, v_L; t_M, t_{\text{cut}}) \theta(t_L > t_{\text{cut}}) \delta(t_R) \delta(v_R) \\
 &+ \mathcal{P}_M^{\text{nb br}}(t_R, v_R; t_M, t_{\text{cut}}) \theta(t_R > t_{\text{cut}}) \delta(t_L) \delta(v_L) \\
 &+ \mathcal{P}_M^{\text{nb nb}}(t_M, t_{\text{cut}}) \delta(t_L) \delta(v_L) \delta(t_R) \delta(v_R).
 \end{aligned}
 \tag{3.22}$$

The δ functions are just formal crutches to be able to define values for $t_{L,R}$ and $v_{L,R}$ when no branching actually occurred. When a daughter does not branch, it is put on shell through the $\delta(t_{L,R})$ functions (for massless particles as we are considering here). The θ functions are not strictly necessary, as they are already contained in the single-branch probabilities, but to be explicit we write them out again.

As a check of the algorithm, it is straightforward to verify that the probability for anything to happen is equal to unity:

$$\int_0^{t_M} dt_L \int dv_L \int_0^{t_M} dt_R \int dv_R \mathcal{P}_M(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}) = 1.
 \tag{3.23}$$

To see this, integrate over $\min(t_L, t_R)$ using eq. (3.16), and then examine the quantity

$$\frac{d}{dt} [\Delta_L(t_M, t) \Delta_R(t_M, t)],
 \tag{3.24}$$

to obtain a simple expression for the integral over $\max(t_L, t_R)$.

It is now straightforward to calculate the probability for generating an event with a given hard $2 \rightarrow 2$ scattering and a history of branches with n particles in the final state. It is simply given by the probability to obtain the hard scattering multiplied by the double-branch probabilities for each of the following branches. As an example, consider the parton shower history shown in figure 10. The differential probability to obtain this event is given by

$$\begin{aligned}
 d\mathcal{P}(\Sigma_{\text{figure 10}}) &= \frac{d\sigma_{M \rightarrow LR}}{d\sigma_{\text{total}}} \times \mathcal{P}_M^{\text{br br}}(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}) dt_L dv_L dt_R dv_R \\
 &\times \mathcal{P}_L^{\text{br nb}}(t_{LL}, v_{LL}; t_L, t_{\text{cut}}) dt_{LL} dv_{LL} \times \mathcal{P}_R^{\text{nb nb}}(t_R, t_{\text{cut}}) \\
 &\times \mathcal{P}_{LL}^{\text{nb nb}}(t_{LL}, t_{\text{cut}}),
 \end{aligned}
 \tag{3.25}$$

where $d\sigma_{M \rightarrow LR}/\sigma_{\text{total}}$ is the probability to obtain the initial $M \rightarrow LR$ branch from a 2-body matrix element with L and R in the final state. Hence, **GenEva**'s parton shower algorithm has indeed an analytic formula for the probability $\mathcal{P}(\Sigma)$ to obtain a given parton shower history.

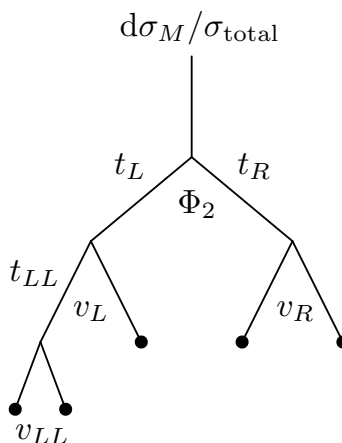


Figure 10: An example parton shower history with 5 particles in the final state. The quantity $d\sigma_{M \rightarrow LR}/\sigma_{\text{total}}$ is the probability to obtain the initial $M \rightarrow LR$ branch from a $2 \rightarrow 2$ hard scattering as a function of Φ_2 . The total probability for obtaining the full 5-particle final state is given in eq. (3.25).

3.5 Truncation

The parton shower described above will produce final states with in principle arbitrarily many final state particles. The goal of **GenEvA** is to reweight such an obtained final state to a known distribution. In practice, the exact distributions will only be available for a limited number of final state particles. Furthermore, for the purposes of the **GenEvA** framework [1], one might want to choose the scale that separates partonic calculations from phenomenological models to be different from the value t_{cut} used by the parton shower.

Consider the case where we only have partonic calculations available with up to n_{max} partons in the final state, and we want the minimum virtuality described by the partonic calculation to be $t_{\text{match}} \geq t_{\text{cut}}$. The parton shower algorithm will in general produce events with more than n_{max} partons in the final state and produce splittings with $t < t_{\text{match}}$. By assumption, these events cannot be reweighted to the partonic calculation. One way to deal with such events is to simply reject them, and only reweight events that satisfy the criteria required by the partonic calculation. However, this would reject a large number of events and would obviously result in large inefficiencies. Furthermore, because the parton shower preserves probability, unless the user calculated the rejection probability, vetoing events would make it impossible to obtain total cross section information from the phase space generator.

Instead of rejecting events with too many final state partons or too low a splitting scale, we can truncate these events to a final state with at most n_{max} particles, in such a way that each branch has virtuality above t_{match} . This is achieved by a simple truncation algorithm that truncates events to a given $\{t_{\text{match}}, n_{\text{max}}\}$:

1. Order the branches by the value of the evolution variable t .

2. Remove any branches with $t < t_{\text{match}}$. That is, remove the daughter particles from any mother whose virtuality is $t < t_{\text{match}}$, leaving the mother unbranched.
3. If the final state contains more than n_{max} particles remove additional branches, starting with the smallest value of t , until there are n_{max} final state particles.
4. Recalculate the kinematics of all remaining final state particles.
5. For events with n_{max} final state particles, redefine t_{match} to be the smallest value of the virtualities t of the remaining branches.

This algorithm clearly results in an event sample for which all branches have $t > t_{\text{match}}$ and there are no more than n_{max} final states. The last step guarantees that if the shower restarts from the event-defined t_{match} scale, then the resulting distribution of events will be identical to those if truncation were not applied. Note that this algorithm preserves probability because no events are ever rejected.

To be able to reweight the resulting events after truncation, we of course need the probability with which the truncated event was generated. As expected from the notion of global evolution, the probability for finding a truncated event should be exactly equal to the probability of generating that event with the shower cutoff t_{cut} set equal to the event-defined t_{match} , i.e.

$$\mathcal{P}_M^{\text{trunc}}(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}; t_{\text{match}}) = \mathcal{P}_M(t_L, v_L; t_R, v_R; t_M, t_{\text{match}}). \quad (3.26)$$

To see this, we write out $\mathcal{P}_M^{\text{trunc}}$ in analogy with eq. (3.22) as

$$\begin{aligned} \mathcal{P}_M^{\text{trunc}}(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}; t_{\text{match}}) &= \mathcal{P}_M^{\text{trunc br br}}(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}; t_{\text{match}}) \theta(t_L > t_{\text{match}}) \theta(t_R > t_{\text{match}}) \\ &+ \mathcal{P}_M^{\text{trunc br nb}}(t_L, v_L; t_M, t_{\text{cut}}; t_{\text{match}}) \theta(t_L > t_{\text{match}}) \delta(t_R) \delta(v_R) \\ &+ \mathcal{P}_M^{\text{trunc nb br}}(t_R, v_R; t_M, t_{\text{cut}}; t_{\text{match}}) \theta(t_R > t_{\text{match}}) \delta(t_L) \delta(v_L) \\ &+ \mathcal{P}_M^{\text{trunc nb nb}}(t_M, t_{\text{cut}}; t_{\text{match}}) \delta(t_L) \delta(v_L) \delta(t_R) \delta(v_R), \end{aligned} \quad (3.27)$$

and show that

$$\mathcal{P}_M^{\text{trunc xy}}(\dots; t_M, t_{\text{cut}}; t_{\text{match}}) = \mathcal{P}_M^{\text{xy}}(\dots; t_M, t_{\text{match}}). \quad (3.28)$$

There are three cases to consider. First, take a double branch for which truncation keeps both the left and the right branches. This occurs if t_L and t_R are initially greater than t_{match} , as in figure 11. Thus, the probability of getting given t_L and t_R values after truncation is

$$\begin{aligned} \mathcal{P}_M^{\text{trunc br br}}(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}; t_{\text{match}}) &= \mathcal{P}_M^{\text{br br}}(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}) \\ &\equiv \mathcal{P}_M^{\text{br br}}(t_L, v_L; t_R, v_R; t_M, t_{\text{match}}), \end{aligned} \quad (3.29)$$

where we used that the single-branch probability in eq. (3.9) only depends on the shower cutoff through an overall θ function. Thus, the probability to get this double branch after

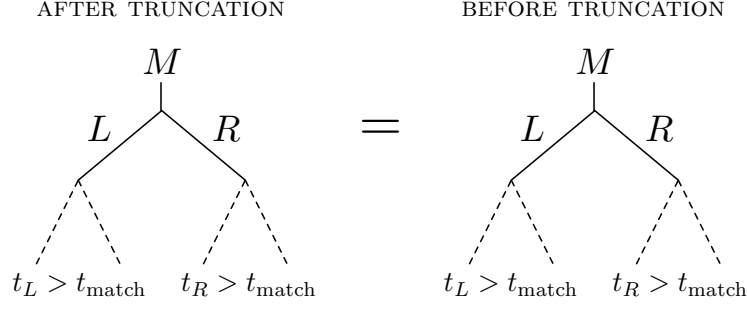


Figure 11: To get a truncated event where both daughters are branched, the original event has to have $t_{L,R} > t_{\text{match}}$.

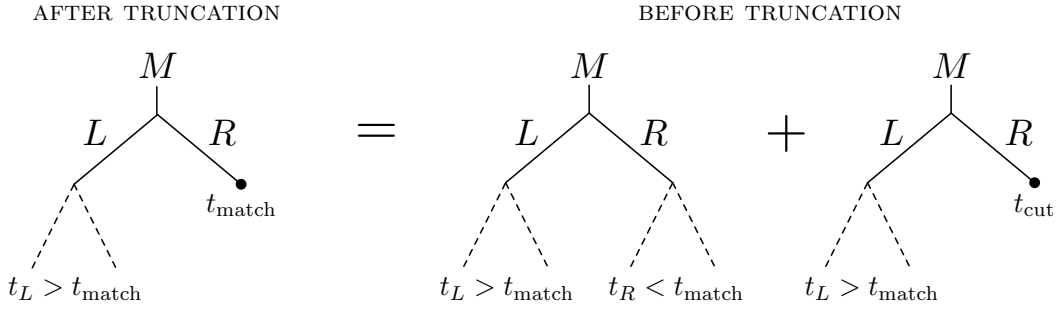


Figure 12: To get a truncated event where only the left daughter is branched, the original event has to have $t_L > t_{\text{match}}$ and $t_R < t_{\text{match}}$. Since, formally, an unbranched particle has $t < t_{\text{cut}}$, there are two different ways this can happen. For the truncated event, the shower cutoff is effectively raised from t_{cut} to t_{match} . The same holds if the roles of L and R are interchanged.

truncation is equal to the probability to get this double branch from running the shower with $t_{\text{cut}} = t_{\text{match}}$.

Next, take the case where after truncation one daughter is branched and one is unbranched. This occurs for example when $t_R < t_{\text{match}} < t_L$, which happens if the left daughter branched above t_{match} , while the right daughter did not branch at all or branched below t_{match} , as in figure 12. In this case, the truncation probability is

$$\begin{aligned}
& \mathcal{P}_M^{\text{trunc br nb}}(t_L, v_L; t_M, t_{\text{cut}}; t_{\text{match}}) \\
&= \int_{t_{\text{cut}}}^{t_{\text{match}}} dt_R \int dv_R \mathcal{P}_M^{\text{br br}}(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}) + \mathcal{P}_M^{\text{br nb}}(t_L, v_L; t_M, t_{\text{cut}}) \\
&= \mathcal{P}_L^{\text{br}}(t_L, v_L; t_M, t_{\text{cut}}) \times \Delta_R(t_M, t_L) \left[\int_{t_{\text{cut}}}^{t_{\text{match}}} dt_R \int dv_R \mathcal{P}_R^{\text{br}}(t_R, v_R; t_*, t_{\text{cut}}) + \Delta_R(t_*, t_{\text{cut}}) \right] \\
&= \mathcal{P}_L^{\text{br}}(t_L, v_L; t_M, t_{\text{match}}) \times \Delta_R(t_M, t_L) \Delta_R(t_*, t_{\text{match}}) \\
&\equiv \mathcal{P}_M^{\text{br nb}}(t_L, v_L; t_M, t_{\text{match}}), \tag{3.30}
\end{aligned}$$

where we used the definitions in eqs. (3.20) and (3.21), and in the third step we used eq. (3.16) and again that $\mathcal{P}_L^{\text{br}}$ only depends on the shower cutoff through a θ function. The

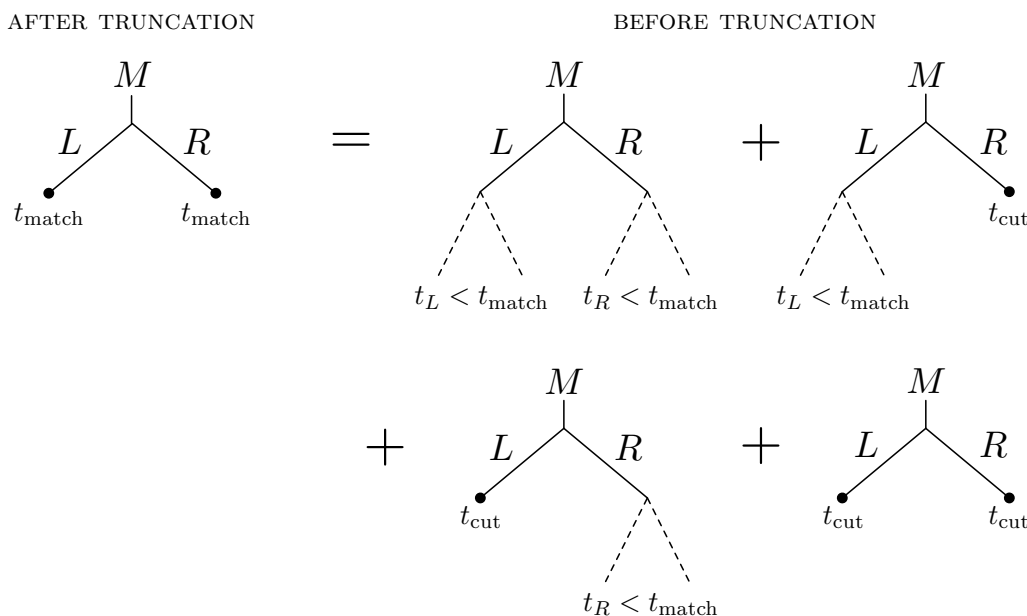


Figure 13: To get a truncated event where neither daughter branches, the original event must have $t_{L,R} < t_{\text{match}}$. Since, formally, an unbranched particle has $t < t_{\text{cut}}$, there are now four different ways for this to happen. For the truncated event, the shower cutoff is effectively raised from t_{cut} to t_{match} .

crucial feature of **GenEva**'s shower algorithm that makes this calculation possible is that the double-branch probability can be integrated analytically over the smaller virtuality t_R without knowing the value of the larger virtuality t_L . An analogous calculation holds for $t_L < t_{\text{match}} < t_R$.

Finally, take the case where after truncation, both daughters are unbranched. This occurs if both daughters either did not branch or branched below t_{match} , as shown in figure 13. In this case, the truncation probability is

$$\begin{aligned}
 \mathcal{P}_M^{\text{trunc nb nb}}(t_M, t_{\text{cut}}; t_{\text{match}}) &= \int_{t_{\text{cut}}}^{t_{\text{match}}} dt_L \int dv_L \int_{t_{\text{cut}}}^{t_{\text{match}}} dt_R \int dv_R \mathcal{P}_M^{\text{br br}}(t_L, v_L; t_R, v_R; t_M, t_{\text{cut}}) \\
 &+ \int_{t_{\text{cut}}}^{t_{\text{match}}} dt_L \int dv_L \mathcal{P}_M^{\text{br nb}}(t_L, v_L; t_M, t_{\text{cut}}) \\
 &+ \int_{t_{\text{cut}}}^{t_{\text{match}}} dt_R \int dv_R \mathcal{P}_M^{\text{nb br}}(t_R, v_R; t_M, t_{\text{cut}}) + \mathcal{P}_M^{\text{nb nb}}(t_M, t_{\text{cut}}) \\
 &= \mathcal{P}_M^{\text{nb nb}}(t_M, t_{\text{match}}), \tag{3.31}
 \end{aligned}$$

where we used eq. (3.23) and the fact that the truncation algorithm preserves probability. One can also do the t_L and t_R integrals explicitly as in eq. (3.30) to verify this result. The above calculations show that eq. (3.26) indeed holds: The probability to get a truncated event is exactly the same as the probability for it to be produced by the shower with $t_{\text{cut}} = t_{\text{match}}$, consistent with the notion of a global evolution.

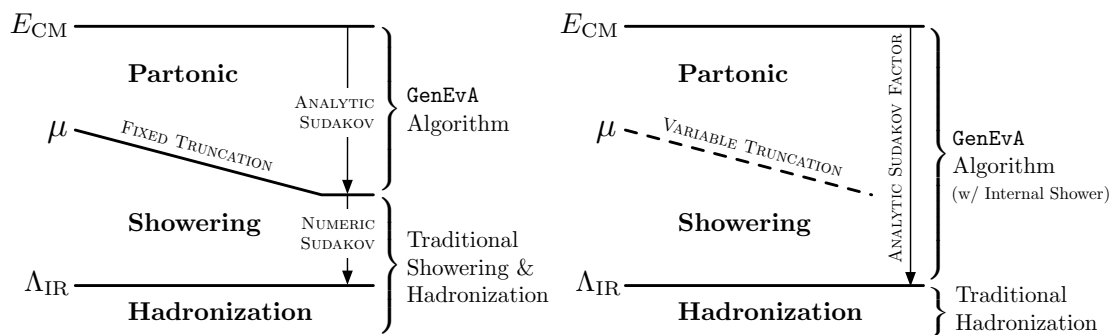


Figure 14: Two different ways of using the **GenEvA** algorithm in a complete event generation framework. In the left panel, the **GenEvA** algorithm is only used above a scale $\mu = \sqrt{t_{\text{cut}}} \gg \Lambda_{\text{IR}}$, and traditional showering and hadronization is applied starting at μ . In the right panel, **GenEvA** is used to distribute events all the way down to $\sqrt{t_{\text{cut}}} = \Lambda_{\text{IR}}$ to interface directly with a hadronization scheme. The first case allows the use of already existing parton showers that have been tuned to data, while the second case allows the scales associated with the partonic calculations to vary even after hadronization and detector simulation. In the companion paper [1], we discuss how both methods can give the same leading-logarithmic Sudakov improvements.

There is one subtlety we have completely ignored so far. In step 2 of the shower algorithm outlined at the beginning of section 3.4, we need to choose values for $E_{L,R}^{\text{start}}$ to be used in the daughters' splitting functions. To not violate eq. (3.26), we need a method to determine values for $E_{L,R}^{\text{start}}$ which gives the same values before and after truncation. In other words, we need an expression for $E_{L,R}^{\text{start}}$ that is invariant under truncation.

Certainly the easiest choice would be to take $E_{L,R}^{\text{start}} \equiv E_{CM}$. However, as noted in section 3.2, we would like $E_{L,R}^{\text{start}}$ to be of the order of the daughters' actual energies. In principle, there are many possible choices. In analogy with $t_{\text{start}} = t_M$, we might attempt to take $E_{L,R}^{\text{start}} = E_M$, where E_M is the mother's energy. However, this does not work, because E_M is not invariant under truncation.⁵ Instead, we take $E_{L,R}^{\text{start}} = E_M^{\text{max}}$, where E_M^{max} is the maximum possible energy of the mother and is determined recursively as follows: Given the E_M^{max} used to branch the daughters, we define the daughters' $E_{L,R}^{\text{max}}$, to be used in the subsequent branching of the granddaughters, by

$$E_{L,R}^{\text{max}} = \frac{1}{2} E_M^{\text{max}} \left[1 + \frac{t_{L,R}}{t_M} + \beta(E_M^{\text{max}}) \cos \theta_M \left(1 - \frac{t_{L,R}}{t_M} \right) \right]. \quad (3.32)$$

This corresponds to taking the minimum and maximum values of z_M in eq. (3.4). For the initial hard scattering we simply take $E_M^{\text{max}} = E_{CM}$. This definition of E^{max} is indeed invariant under truncation, since it only depends on E_{CM} and the values of $\cos \theta$ and t of a particle's direct ancestors, i.e. only on those branches that cannot get truncated unless the particle itself is truncated first.

With truncation, we now have two different ways of using the **GenEvA** algorithm in the context of a complete event generation framework, as illustrated in figure 14. One

⁵As shown by eq. (3.4), the energies of two sisters depend on each others virtualities as well as their mother's energy and $\cos \theta$. Thus, truncating the aunt of a particle changes its mother's energy.

method is to run the **GenEvA** algorithm to $t_{\text{cut}} \gg \Lambda_{\text{IR}}^2$ and truncate it to $\{t_{\text{cut}}, n_{\text{max}}\}$, which means that $(n \leq n_{\text{max}})$ -parton events are unaffected, while $(n > n_{\text{max}})$ -parton events get truncated to n_{max} -parton phase space with a variable, event-specific t_{match} scale. The events can then be passed to an external showering/hadronization program to start a new shower at t_{cut} for $n < n_{\text{max}}$ and the variable t_{match} scale for $n = n_{\text{max}}$. This uses **GenEvA** more-or-less as a traditional phase space generator. Alternatively, we can run the **GenEvA** algorithm all the way down to $t_{\text{cut}} = \Lambda_{\text{IR}}^2$, interface with a hadronization routine, and then after the fact truncate the shower history back up to $\{t_{\text{match}} \gg \Lambda_{\text{IR}}^2, n_{\text{max}}\}$ to obtain a point in n -body phase space which now already has a complete shower attached to it.

Because truncation preserves probability, both of these methods will give statistically identical results, however, the second method allows multiple different truncation schemes on the same already hadronized (and even detector-simulated) event. Of course, in practice an external showering program can be tuned to data, whereas the **GenEvA** shower only guarantees leading-logarithmic accuracy, but the possibility to improve perturbative theoretical distributions after detector simulation is an intriguing possibility, and for this reason it might be useful to develop an analytically calculable but realistic parton shower.

4. The Jacobian factor

In the previous section, we saw that the analytic parton shower covers all of phase space. However, the parton shower distributes events in terms of the shower variables $\{t, \cos \theta, \phi\}$ as in eq. (3.1), and there is a Jacobian factor from converting these to the variables describing Lorentz-invariant phase space.

The complete parton shower space $d\Sigma_n$ for a shower history with n final state particles has two pieces. First, there are the variables $d\Phi_2^{(0)}$ of the initial hard scattering that yields two on-shell (and for us massless) final states A and B . Second, there is the product of the individual $d\Sigma_2$ splitting variables that describe the subsequent $1 \rightarrow 2$ parton shower branches. This gives

$$d\Sigma_n = d\Phi_2^{(0)}(p_{\text{CM}}; p_A^{(0)}, p_B^{(0)}) \prod_{i=1}^{n-2} d\Sigma_2(M_i \rightarrow L_i R_i), \quad (4.1)$$

where $p_{A,B}^{(0)}$ are the four-momenta of the particles A, B before running the parton shower, and i runs over all parton shower branches. We define that the way the parton shower takes the momenta of the A and B particles off shell is by leaving the angles $d\Omega_2$ unchanged. Therefore, the phase space for the initial on-shell final state particles is related to that of the ultimate off-shell, showered particles via

$$d\Phi_2^{(0)}(p_{\text{CM}}; p_A^{(0)}, p_B^{(0)}) = \frac{1}{\lambda(t_{\text{CM}}; t_A, t_B)} d\Phi_2(p_{\text{CM}}; p_A, p_B), \quad (4.2)$$

where the ordinary phase space factor λ is given in eq. (3.5).

To convert the natural shower variables $d\Sigma_n$ to Lorentz-invariant n -body phase space $d\Phi_n$, we can use the well-known method to recursively decompose n -body phase space in

terms of multiple copies of 2-body phase space:

$$\begin{aligned}
 d\Phi_n(p_{\text{CM}}; p_1, \dots, p_n) &= (2\pi)^4 \delta^4\left(p_{\text{CM}} - \sum_{i=1}^n p_i\right) \prod_{i=1}^n \frac{d^4 p_i}{(2\pi)^3} \delta(p_i^2 - m_i^2) \theta(p_i^0) \\
 &= d\Phi_{n-1}(p_{\text{CM}}; p_1, \dots, p_{n-2}, Q) \frac{dQ^2}{2\pi} d\Phi_2(Q; p_{n-1}, p_n) \\
 &= d\Phi_{n-1}(p_{\text{CM}}; p_1, \dots, p_{n-2}, Q) \frac{1}{J(t_Q; t_{n-1}, t_n)} dt_Q d\cos\theta_Q d\phi_Q,
 \end{aligned} \tag{4.3}$$

where the 2-body Jacobian factor is given by

$$J(t_M; t_L, t_R) = \frac{64\pi^3}{\lambda(t_M; t_L, t_R)}. \tag{4.4}$$

The variables $t_Q \equiv Q^2$, $\cos\theta_Q$, and ϕ_Q in the last line of eq. (4.3) are exactly the same variables used in $d\Sigma_2$ in eq. (3.1). Their integration ranges are such that for any $d\Phi_2(p_M; p_L, p_R)$ we have

$$-1 \leq \cos\theta_M \leq 1, \quad 0 \leq \phi_M \leq 2\pi, \quad 0 \leq \sqrt{t_L} + \sqrt{t_R} \leq \sqrt{t_M}, \tag{4.5}$$

identical to eq. (3.2). Lorentz-invariant phase space can thus be written suggestively as

$$d\Phi_n(p_{\text{CM}}; p_1, \dots, p_n) = d\Phi_2(p_{\text{CM}}; p_A, p_B) \prod_{i=1}^{n-2} \frac{1}{J(t_{M_i}, t_{L_i}, t_{R_i})} d\Sigma_2(M_i \rightarrow L_i R_i), \tag{4.6}$$

where A and B can be thought of as the particles originating from the hard interaction and t_{M_i} , t_{L_i} and t_{R_i} denote the virtualities of the mother and daughter particles in the i -th phase space recursion.

Though both the parton shower and Lorentz-invariant phase space can be built recursively, there is one important difference between the parton shower $d\Sigma_n$ and Lorentz-invariant phase space $d\Phi_n$ regarding identical particles. A parton shower distinguishes all final state particles by a series of “ L ” and “ R ” labels, regardless of whether those particles have the same quantum numbers. In contrast, Lorentz-invariant phase space does not distinguish between identical particles. Thus, an additional factor has to be included in the Jacobian to account for this difference. For a parton shower history with n final state particles that contains n_1, n_2, \dots, n_k sets of identical particles, with $n_1 + n_2 + \dots + n_k = n$, we have to divide the Jacobian by an extra factor of $n_1! n_2! \dots n_k!$.

Putting all this information together, the total Jacobian relating n -body parton shower space to Lorentz-invariant n -body phase space is

$$J(\Sigma_n) \equiv \left| \frac{\partial(\Sigma_n)}{\partial(\Phi_n)} \right| = \frac{1}{\lambda(t_{\text{CM}}; t_A, t_B)} \frac{1}{n_1! n_2! \dots n_k!} \prod_{i=1}^{n-2} J(t_{M_i}, t_{L_i}, t_{R_i}). \tag{4.7}$$

This factor is a crucial piece of the master formula eq. (2.12).

5. The overcounting factor

5.1 Overview

Different parton shower histories can give rise to the same set of final state four-momenta, so the parton shower covers phase space multiple times. Therefore, if we wish to reweight a parton shower event to a matrix element, which only depends on the phase space Φ of the external particles but knows nothing about the underlying parton shower history Σ , we have to account for the fact that the mapping $\Phi(\Sigma)$ from parton shower space Σ to phase space Φ is not one-to-one. This is achieved by the overcounting factor $\hat{\alpha}(\Sigma)$ in eq. (2.12).

Since there is only a fixed number of parton shower trees that can yield a given set of external particles, there is also only a fixed number $n(\Phi)$ of points $\Sigma_i(\Phi)$ that map to a given point Φ . Certainly the easiest choice for $\hat{\alpha}(\Sigma)$ is thus

$$\hat{\alpha}(\Sigma) = \frac{1}{n[\Phi(\Sigma)]}, \quad (5.1)$$

which only requires the knowledge of $n(\Phi)$ for $\Phi = \Phi(\Sigma)$, but not the precise form of each $\Sigma_i(\Phi)$. Furthermore, we have defined our parton shower without any dead zones, such that for a given set of n final state particles, each compatible parton shower tree fully covers the corresponding n -body phase space. Hence, $n(\Phi)$ only depends on the particle types in Φ and thus equals the number of compatible parton shower trees, a quantity that could be calculated in advance. On the other hand, the reweighting efficiency of this choice will usually be poor, because if two points $\Sigma_{1,2}(\Phi)$ have $\mathcal{P}(\Sigma_1) \gg \mathcal{P}(\Sigma_2)$ then eq. (2.12) yields weights $w(\Sigma_1) \ll w(\Sigma_2)$ that are very different.

The optimal choice in terms of reweighting efficiency is to have the same weight $w[\Sigma_i(\Phi)]$ for all i and a given Φ . This is achieved by

$$\hat{\alpha}(\Sigma) = \frac{\mathcal{P}(\Sigma)J(\Sigma)}{\sum_i \mathcal{P}(\Sigma_i)J(\Sigma_i)}, \quad (5.2)$$

which just reduces eq. (2.12) back to eq. (2.7). Unfortunately, calculating the sum in eq. (5.2) requires the construction of all possible parton shower histories $\Sigma_i[\Phi(\Sigma)]$ and the computation of their probabilities $\mathcal{P}(\Sigma_i)$. Since the number of parton shower histories Σ_i for n final state particles grows like $n!$, this quickly becomes computationally expensive.

In general, we can assign

$$\hat{\alpha}(\Sigma) = \frac{\alpha(\Sigma)}{\sum_i \alpha(\Sigma_i)}, \quad (5.3)$$

such that $\sum_i \hat{\alpha}(\Sigma_i) = 1$ automatically, and choose $\alpha(\Sigma)$ such that $\sum_i \alpha(\Sigma_i)$ is more readily calculable. The choices $\alpha(\Sigma) = 1$ and $\alpha(\Sigma) = \mathcal{P}(\Sigma)J(\Sigma)$ recover eq. (5.1) and eq. (5.2), respectively. The better $\alpha(\Sigma)$ approximates $\mathcal{P}(\Sigma)J(\Sigma)$ the more uniform the total weights will be, resulting in a better reweighting efficiency. At the same time, it is desirable to have a fast way to compute $\sum_i \alpha(\Sigma_i)$, such that the computational time used to correct the overcounting issue is less than (or at least comparable with) the computational time needed to evaluate the matrix element $\sigma(\Phi)$. In the next subsection we show that, with

a suitable choice of $\alpha(\Sigma)$, the ALPHA algorithm of refs. [77, 78] can be applied to calculate the sum $\sum_i \alpha(\Sigma_i)$, reducing the $n!$ problem to a 2^n problem, which is a significant gain.

Before going on, we want to point out that the overcounting factor $\hat{\alpha}(\Sigma)$ has appeared in other Monte Carlo programs, most notably **MadEvent**. Indeed, our solution to the phase space overcounting by the parton shower was inspired by the technique used in ref. [14] to slice phase space into complementary integration ranges. Given a set of n Feynman diagrams

$$|\mathcal{M}|^2 = \left| \sum_{i=1}^n \mathcal{M}_i \right|^2, \quad (5.4)$$

MadEvent performs independent phase space integrations over n different functions

$$f_i = w_i |\mathcal{M}|^2 \quad \text{with} \quad w_i = \frac{|\mathcal{M}_i|^2}{\sum_i |\mathcal{M}_i|^2}, \quad (5.5)$$

such that the sum over the n integrations $\sum_i w_i |\mathcal{M}|^2 = |\mathcal{M}|^2$ gives the desired integrated squared-amplitude. The “overcounting factor” w_i is not gauge invariant, but it does have the singularity structure of the i -th Feynman diagram, allowing a phase space integration that is aware of the relevant poles. **GenEvA** does essentially the same, but now $\hat{\alpha}(\Sigma)$ will capture the singularity and color factor structure of a given parton shower history.

5.2 The ALPHA algorithm

We want to calculate a quantity that corresponds to a sum over all possible graphs. This same issue arises in the computation of quantum amplitudes which are sums over all possible Feynman diagrams. Therefore, we can use known techniques to efficiently calculate the sum $\sum_i \alpha(\Sigma_i)$ by constructing a toy “Lagrangian” whose “generating functional” provides the desired “amplitude” as in figure 15.

One simple algorithm for doing this is the ALPHA algorithm [77, 78], which is the core algorithm used in the **ALPGEN** event generator [13]. The ALPHA algorithm essentially works by reducing common subexpressions in the sum over all graphs. In a more familiar language, it exploits the fact that the calculation of a tree-level amplitude (which for us is the exact result) can be reduced to recursively solving equations of motion in the presence of external sources.

Consider the following Lagrangian:

$$L(\{\phi_a\}) = \frac{1}{2} C_{ab} \phi_a \phi_b + \frac{\lambda}{6} D_{abc} \phi_a \phi_b \phi_c, \quad (5.6)$$

where λ is a dummy coupling constant and ϕ_a are simple fields with no space-time dependence (i.e. field theory in $0+0$ dimensions). The generating functional $Z(\{J_a\})$ of this theory is⁶

$$Z(\{J_a\}) = -L(\{\phi_a\}) + \phi_a C_{ab} J_b, \quad (5.7)$$

where the fields ϕ_a satisfy the variational condition (the equations of motion)

$$\frac{\delta L}{\delta \phi_a} = C_{ab} J_b. \quad (5.8)$$

⁶The factor of C_{ab} is included in the source term such that J_a is a source for ϕ_a , not a source for $C_{ab}^{-1} \phi_b$.

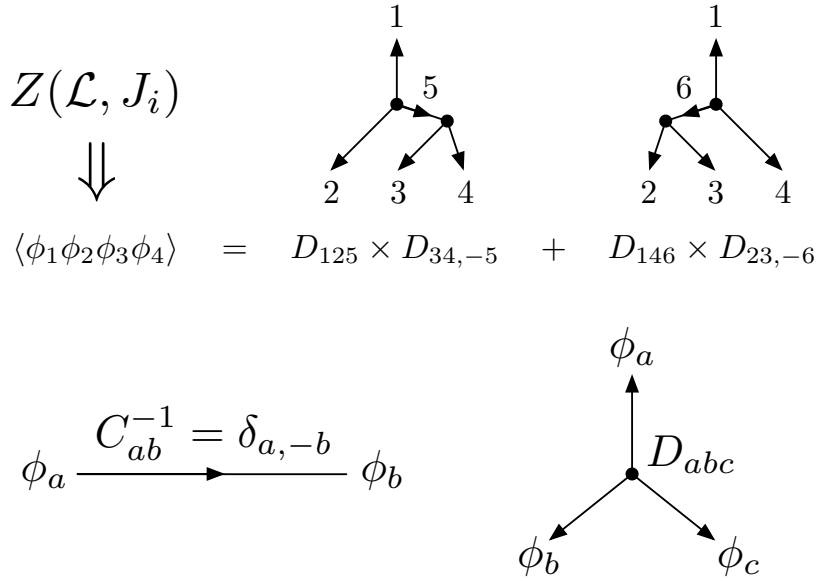


Figure 15: Schematic picture of the ALPHA algorithm. Given a quantity that can be described as the sum over all diagrams, one can create a Lagrangian \mathcal{L} whose generating functional Z yields the appropriate sum over “Feynman diagrams”. The ALPHA algorithm gives an efficient way of computing these sums as long as the fields ϕ_i only appear linearly in \mathcal{L} , hence the need for the off-diagonal propagator $C_{ab}^{-1} = \delta_{a,-b}$. In the above example, particle 1 could correspond to a photon, 2 to a quark, 3 to a gluon, and 4 to an anti-quark. The 125 and 146 vertices could correspond to the electromagnetic couplings, the 34, -5 vertex to the $\bar{q} \rightarrow \bar{q}g$ splitting function, and the 23, -6 vertex to the $q \rightarrow qg$ splitting function.

The n -body correlation function is equal to

$$\langle \phi_1 \phi_2 \cdots \phi_n \rangle = \lim_{J_i \rightarrow 0} \frac{\partial^n Z(\{J_a\})}{\partial J_1 \partial J_2 \cdots \partial J_n}, \quad (5.9)$$

and is identical to the sum over all Feynman diagrams (or in our case, the sum over all parton shower histories) with propagators C_{ab}^{-1} , vertices D_{abc} , and n external legs.

To calculate Z at tree level, we can solve eq. (5.8) for the fields ϕ_a . We start by expanding ϕ_a in a series in λ

$$\phi_a = \phi_a^0 + \lambda \phi_a^1 + \cdots + \lambda^{n-2} \phi_a^{n-2}, \quad (5.10)$$

where ϕ_a^i is the coefficient of ϕ_a at order λ^i , and the series need only go to order λ^{n-2} , because an n -point correlator needs only $n - 2$ three-point interactions. The equation of motion following from eq. (5.8) is

$$\phi_a = J_a - \frac{\lambda}{2} C_{ab}^{-1} D_{bcd} \phi_c \phi_d, \quad (5.11)$$

which we can easily solve recursively order by order in λ using eq. (5.10). The result is

$$\begin{aligned} \phi_a^0 &= J_a, \\ \phi_a^k &= -\frac{1}{2}C_{ab}^{-1}D_{bcd}(\phi_c^0\phi_d^{k-1} + \phi_c^1\phi_d^{k-2} + \dots + \phi_c^{k-1}\phi_d^0) = -\frac{1}{2}C_{ab}^{-1}D_{bcd} \sum_{i+j=k-1} \phi_c^i\phi_d^j, \end{aligned} \quad (5.12)$$

which can then be plugged back into eq. (5.7).

In the special case that \mathcal{L} is only linear in the ϕ_a fields, the expression for the correlator in eq. (5.9) simplifies. For n external fields $\phi_1, \phi_2, \dots, \phi_n$:

$$\langle \phi_1 \cdots \phi_n \rangle = Z(J_1 = 1, J_2 = 1, \dots, J_n = 1, J_{\text{other}} = 0). \quad (5.13)$$

We will see below that the ‘‘Lagrangian’’ relevant for **GenEvA** is linear in the ϕ_a , so eq. (5.13) does indeed hold.

5.3 A shower ‘‘Lagrangian’’

To build a Lagrangian useful for the **GenEvA** algorithm, we introduce a separate field ϕ for every particle in a parton shower history (both for internal and external fields) and label each field by the specific four momentum p_a and additional quantum numbers f_a (flavor, spin, color, etc.) of the particle it represents,

$$\phi = \phi_{p_a, f_a} \equiv \phi_a. \quad (5.14)$$

We stress that p_a and f_a are just labels; the field itself does not carry any momentum or quantum numbers. Note also that as far as the parton shower is concerned, all particles are distinguishable by their ‘‘ L ’’ and ‘‘ R ’’ labels. In addition, we define all momenta as outgoing and introduce two separate fields for the same particle going into a vertex and coming out of a vertex, with the convention that if ϕ_a represents an outgoing particle with outgoing momentum p_a then ϕ_{-a} represents the same particle incoming with outgoing momentum $-p_a$. The reason to have two fields for every particle is in order to use the result of eq. (5.13) which requires \mathcal{L} to be linear in ϕ_a .

Assume we have a parton shower history with n external (incoming or outgoing) particles with momenta P_i , where i runs from 1 to n . We again define all P_i as outgoing, such that overall momentum conservation reads $\sum_i P_i = 0$. For example, for a final state parton shower with $n - 1$ outgoing partons we include an ‘‘outgoing’’ center-of-mass particle with momentum

$$P_n = -\sum_{i=1}^{n-1} P_i. \quad (5.15)$$

The crucial point is that the momenta P_i span all possible momenta p_a that can occur in any parton shower history with the same set of external particles. Explicitly, we introduce $2^n - 2$ different fields ϕ_a corresponding to the momenta

$$p_a = \sum_{i=1}^n c_a^i P_i \quad \text{with} \quad c_a^i = 0 \text{ or } 1, \quad (5.16)$$

excluding the case where all c_a^i are 0 or 1, which would give $p_a = 0$. The index a runs over $1 \leq |a| \leq 2^{n-1} - 1$, and by definition we take $c_{-a}^i = 1 - c_a^i$ such that

$$p_{-a} = \sum_{i=1}^n (1 - c_a^i) P_i = 0 - \sum_{i=1}^n c_a^i P_i = -p_a, \quad (5.17)$$

as defined above. For simplicity, we also define $c_a^i = \delta_a^i$ for $1 \leq a \leq n$ such that the first n fields ϕ_a represent the external particles of the history.⁷ Hence, while the number of parton shower histories grows like $n!$, the number of ϕ_a fields needed to calculate the sum over parton shower histories only grows like 2^n , and is therefore computationally tractable for reasonable numbers of external particles.

To include the additional quantum numbers f_a in the above discussion, we can replace $p_a \rightarrow (p, f)_a$ and $P_i \rightarrow (P, f)_i$ everywhere, supplementing eq. (5.16) with appropriate quantum number information (e.g. to include flavors). We only need to include a field ϕ_a if there is at least one way to perform the specific sum over f . For example, if we have two outgoing quarks with momenta P_1 and P_2 then we do not need a field with momentum label $p_{12} = P_1 + P_2$ because there is no particle that can decay into two quarks. On the other hand, we can add two quarks and an antiquark of the same flavor by first adding the quark and the antiquark to give a gluon which can be added to the remaining quark.⁸

The addition of momenta is always unique. However, including additional quantum numbers, the final result of the sum over f can be ambiguous. For instance, including QED radiation and ignoring color, adding a quark and an antiquark can give a gluon or a photon. In this case, we either have to introduce a separate field ϕ for each final result, or alternatively, include the additional quantum number, color in this case, that breaks the ambiguity.

With the above conventions, momentum conservation in the parton shower requires

$$C_{ab} = C(a)\delta_{a,-b}, \quad C_{ab}^{-1} = \frac{1}{C(a)}\delta_{a,-b}, \quad D_{abc} = 0 \quad \text{for} \quad p_a + p_b + p_c \neq 0. \quad (5.18)$$

The first condition on the propagator means that an outgoing particle with p_a can only connect to its incoming version with outgoing momentum $p_{-a} = -p_a$, while the second condition enforces momentum conservation at each vertex in the parton shower history. Overall momentum conservation now implies that Z can only be linear in the sources for the external fields so eq. (5.13) holds.⁹ Note that this relies crucially on the fact that we introduced separate fields for incoming and outgoing particles, such that C_{ab} would be off-diagonal.

To obtain the desired sum over all parton shower histories we can now simply use eq. (5.13) to plug the expansion of the fields in eq. (5.10) back into eq. (5.7) and extract

⁷Except for the trivial case $n = 2$ where $p_1 = P_1$ and $p_{-1} = P_2$.

⁸In practice, the required fields ϕ_a are constructed recursively by always combining two previously constructed fields, so adding flavors is trivial and requires no additional computational cost.

⁹To see this, assume that Z is quadratic or higher in one of the sources for the external fields with momentum P_i . This implies that there exists a correlator which has two or more copies of one of the original external fields. By construction, this cannot be the case, because it would require at least one field with momentum label containing $2P_i$ for some i .

the order λ^{n-2} piece:

$$\begin{aligned} \sum_i \alpha(\Sigma_i) &\equiv \langle \phi_1 \cdots \phi_n \rangle \\ &= -\frac{1}{2} \sum_{i+j=n-2} C(a) \phi_a^i \phi_{-a}^j - \frac{1}{6} \sum_{i+j+k=n-3} D_{abc} \phi_a^i \phi_b^j \phi_c^k + C(a) \phi_a^0 \phi_{-a}^{n-2}, \end{aligned} \quad (5.19)$$

where a sum over a, b, c is understood and we also used eq. (5.18) for C_{ab} . The expressions for the fields ϕ_a^k in eq. (5.19) are obtained from eqs. (5.12) and (5.18) and setting $J_{1,\dots,n} = 1$ and all other sources to zero,

$$\begin{aligned} \phi_a^0 &= \begin{cases} 1 & \text{for } 1 \leq a \leq n, \\ 0 & \text{otherwise,} \end{cases} \\ \phi_a^k &= -\frac{1}{2C(a)} D_{-a,bc} \sum_{i+j=k-1} \phi_b^i \phi_c^j. \end{aligned} \quad (5.20)$$

Once $C(a)$ and D_{abc} are specified, then eq. (5.19) together with eq. (5.20) is our final result for $\sum_i \alpha(\Sigma_i)$. To obtain $\alpha(\Sigma)$ for a particular parton shower history (i.e. the numerator in eq. (5.3)) we can simply multiply the appropriate factors of $1/C(a)$ and D_{abc} for the given parton shower history. We will discuss the precise form of C_a and the nonzero D_{abc} next.

5.4 Choice of overcounting function

As discussed in section 5.1, ideally $\alpha(\Sigma)$ should equal the probability for generating a given history, but if we want to use the ALPHA algorithm outlined above, we have to be in a situation where that probability can be cast into a propagator $C(a)^{-1}$ and a local vertex D_{abc} , which can only depend on the momenta and quantum numbers associated with the indices a, b, c . As we have seen in section 3, the splitting of a mother into two daughters requires information from the grandmother, which cannot be encoded in the local vertex D_{abc} . Also, from the computational point of view, calculating the probability for a history requires numerically expensive Sudakov factors. Therefore, the compromise we will take is to choose

$$C(a) = 1 \quad (5.21)$$

and have all nonzero D_{abc} equal to a Jacobian-improved splitting function.

We stress again that we are free to make any choice for $C(a)$ and the nonzero D_{abc} . For example, choosing $C(a) = 1$ and $D_{abc} = 1$ the algorithm simply counts the number of parton shower histories, which is equivalent to using eq. (5.1) and provides a cross-check on the implementation. The choice of overcounting ‘‘Lagrangian’’ only determines the efficiency of the program and not the resulting distributions.

By definition, if D_{abc} is nonzero then the flavors f_a, f_b, f_c form a valid splitting with $p_a + p_b + p_c = 0$. In a parton shower, the mother particle is treated differently from the daughters, so to determine which of the three particles acts as the mother of the splitting, we calculate the invariant masses $t_{a,b,c} = p_{a,b,c}^2$. Since the parton shower is ordered in t , we know that the largest of $t_{a,b,c}$ belongs to the mother. Assuming for simplicity that this is t_a , we use

$$D_{abc} = f_{f_a \rightarrow f_b f_c}(t_a, \cos \theta_a, \phi_a) J(t_a; t_b, t_c). \quad (5.22)$$

where $J(t_a; t_b, t_c)$ is the Jacobian defined in eq. (4.4) and $f_{f_a \rightarrow f_b f_c}$ is the splitting function defined in section 3.2.¹⁰ Since this choice captures all the singularities of QCD, we can expect the corresponding $\alpha(\Sigma)$ to be reasonably close to $\mathcal{P}(\Sigma)J(\Sigma)$.

If $t_a = E_{\text{CM}}^2$, then the splitting corresponds to the initial hard interaction, in which case we use

$$D_{abc}^{\text{CM}} = J(t_a; t_b, t_c). \tag{5.23}$$

We can also add extra matrix element information to D_{abc}^{CM} such as angular information or charges. In practice this is only important if there are two or more different hard-interaction vertices that can contribute to the same final state, such as $e^+e^- \rightarrow u\bar{u}d\bar{d}$, which can have both a $e^+e^- \rightarrow u\bar{u}$ and $e^+e^- \rightarrow d\bar{d}$ core.

6. The matrix element

The last missing ingredient in calculating the weight for a given parton shower history is the desired distribution $\sigma[\Phi(\Sigma)]$. The appropriate choice for this distribution is entirely determined by physics considerations, and one should implement the best possible theoretical calculations available. Since this work focuses on the generation of the phase space, we will only consider two simple theoretical distributions as examples and discuss more elaborate choices in the companion paper [1].

We briefly discuss phase space issues for tree-level (LO) matrix elements $\sigma(\Phi)$. We then consider tree-level matrix elements improved to leading-logarithmic accuracy (LO/LL). This corresponds to the type of calculation one does in parton shower/matrix element merging [25–37]. To arrive at a simple expression for the LO/LL cross section, we will use the notion of $\sigma_i(\Phi)$ and the alternative form of the event weight in eq. (2.13).

6.1 Tree-level matrix elements

The simplest theoretical distributions one could use in an event generator are given by fixed-order tree-level matrix elements. These results depend only on the scalar products of the four-momenta of the external particles and are defined entirely by Lorentz-invariant phase space variables Φ . While such tree-level calculations give at best a rudimentary description of the true QCD distributions, we will use them to validate our phase space generator in section 7.1 and test its efficiency in section 7.3.

We want to touch on two additional subtleties having to do with phase space vetoes. Recall that at tree level, there are singularities in QCD matrix elements when two partons get close in $t_{ij} = (p_i + p_j)^2$. In full QCD, these singularities are regulated by Sudakov factors, but a tree-level event generator has to cut off these t_{ij} singularities by hand. In the context of the **GenEvA** algorithm, the natural phase space cut variable is virtuality.

¹⁰There is a technical issue that the splitting functions are functions of E_M^{start} , so strictly speaking there is not enough local information to even determine the correct splitting functions, because in the shower algorithm we set E_M^{start} to the maximum energy of the grandmother. However, regardless of which E_M^{start} we choose to define D_{abc} , the singularities will still be correct. In practice, we calculate E_M^{start} assuming that the mother’s sister is unbranched and that the grandmother’s rest frame is the CM frame.

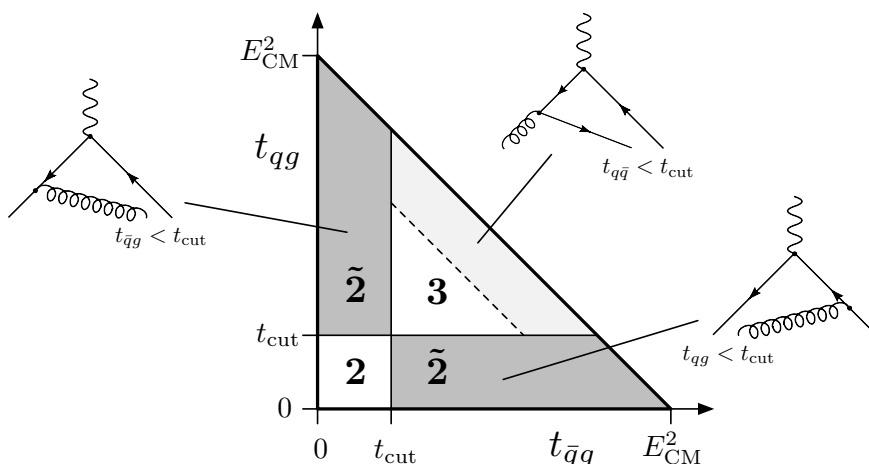


Figure 16: 3-body phase space for $e^+e^- \rightarrow q\bar{q}g$. Region **2** corresponds to the 2-jet region because the parton shower will not populate regions with both t_{qg} and $t_{\bar{q}g}$ less than t_{cut} . However, the parton shower can “accidentally” distribute events in the $\tilde{\mathbf{2}}$ regions where either t_{qg} or $t_{\bar{q}g}$ is less than t_{cut} . These regions must be vetoed because they occupy a QCD singularity away from a shower singularity. While there is no QCD singularity associated with $t_{q\bar{q}} < t_{\text{cut}}$, it is convenient to also veto that region for simplicity (e.g. when comparing to MadEvent in section 7.1). The remaining part of phase space defines the 3-jet region **3**.

After all, if we run the shower down to t_{cut} , then the shower never encounters the singular region of QCD.

However, when we run the shower, it is possible for two particles to “accidentally” have virtualities smaller than t_{cut} , and the matrix element will return a singular answer. This sickness also appears in the overcounting factor, because the ALPHA algorithm will encounter a value of D_{abc} that cannot be generated by a shower with a t_{cut} ending scale. Consider 3-body phase space for $e^+e^- \rightarrow q\bar{q}g$ as in figure 16. There are regions of phase space for which a large angle emission from $q \rightarrow qg$ can put the gluon closer to the anti-quark than t_{cut} . The $\bar{q} \rightarrow \bar{q}g$ splitting cannot generate those same four-momenta because it is below the cutoff scale, so there is no “partner” shower history for the event. Moreover, the shower generated those four-momenta far away from any QCD singularity, while the four-momenta are indeed in the singular region, so the event weight will be very large.

The solution, of course, is to veto events from the shower for which the minimum t_{ij} between singularity-producing partons is less than t_{cut} . Note that we need not veto events for which the q and \bar{q} are closer than t_{cut} , because in the context of $e^+e^- \rightarrow q\bar{q}g$, there is no $q\bar{q}$ singularity.¹¹ If for simplicity we do not wish to impose a flavor-dependent t_{cut} scale, then we can always veto events where any pair of particles (singular or not) are closer than t_{cut} . Denoting the fixed-order tree-level matrix elements by $\sigma^{\text{tree}}(\Phi)$, this is equivalent to taking

$$\sigma^{\text{LO}}(\Phi) = \sigma^{\text{tree}}(\Phi) \prod_{ij} \theta(t_{ij} > t_{\text{cut}}). \tag{6.1}$$

¹¹There is a singularity in $e^+e^- \rightarrow q\bar{q}q'\bar{q}'$, of course.

Using eq. (2.12), the corresponding event weight is

$$w^{\text{LO}}(\Sigma) = \frac{\sigma^{\text{tree}}[\Phi(\Sigma)] \hat{\alpha}(\Sigma)}{\mathcal{P}(\Sigma) J(\Sigma)} \prod_{ij} \theta(t_{ij} > t_{\text{cut}}), \quad (6.2)$$

which is what we will use in section 7.1.

The second subtlety has to do with the fact that the scale t_{match} is ambiguous for n_{max} -body phase space. Recall from section 3.5 that to create fully exclusive events with no dead zones, we want to start a parton shower approximation at t_{match} . When n_{max} -body phase space is generated, the scale t_{match} will vary on an event-by-event basis to ensure that the attached parton shower will populate all of $(n > n_{\text{max}})$ -body phase space. However, since the same n_{max} four-momenta can be created by different shower histories during the phase space generation, they can be associated with several different t_{match} scales.

As far as phase space vetoes are concerned, the variable t_{match} can simply be ignored, because even after truncating to $\{t_{\text{cut}}, n_{\text{max}}\}$, the shower does populate all of n_{max} -body phase space down to t_{cut} . However, if we start a parton shower approximation at t_{match} , we do have to be mindful that events having the same n_{max} partonic four-momenta will start the shower at different scales. Far from being a bug, this t_{match} ambiguity will be exploited in the next subsection to create the leading-logarithmically improved event sample. As we discuss in the companion paper [1], this ambiguity is a physical ambiguity that has to be resolved with calculations.

6.2 Logarithmically-improved matrix elements

It is well known that fixed-order calculations do not give an adequate description of true QCD distributions in singular regions of phase space. This is because fixed-order calculations only reproduce the power singularities that occur when intermediate propagators go on shell. However, in addition to these power singularities there are logarithmically enhanced terms, which start to show up one higher order in perturbation theory. These logarithmically enhanced terms have the form $[\alpha_s \log^2(t/E_{\text{CM}}^2)]^n$. Thus, the effective expansion parameter in the singular regions of phase space is not α_s , but rather $\alpha_s \log^2(t/E_{\text{CM}}^2)$. These double-logarithmic terms can be resummed to all orders in perturbation theory, giving an effective expansion parameter $\alpha_s \log(t/E_{\text{CM}}^2)$, which is only enhanced by a single logarithm. However, such resummed calculations are more difficult to implement, since they not only depend on the four-momenta of the external particles, but also on a choice of scales which determine the exact form of the resummation.

Resummed calculations are crucial if one wants to merge partonic calculations with parton shower algorithms to produce fully exclusive events, as required for meaningful comparisons with experimental collider data. The parton shower automatically includes the resummation of the double-logarithmic terms via the Sudakov factor discussed in eq. (3.10), and therefore has double-logarithmic sensitivity to the scale at which it is started. Thus, the absence of resummed partonic calculations leads to a double-logarithmic dependence on the unphysical matching scale t_{match} between the partonic calculation and the parton shower. This issue is discussed in much more detail in the companion paper [1].

Most implementations of resummed calculations in event generators define the relevant scales in terms of the external four-momenta. An example is the so-called CKKW prescription [25] to merge fixed order calculations with parton showers. In this prescription, one first generates an event based on tree-level matrix elements and then uses a flavor-aware k_T algorithm on the external particles to find the most likely parton shower history that could have generated this event. The resummed partonic result is then obtained by reweighting the fixed-order event by the appropriate no-branching probabilities associated with this specific parton shower history.¹²

There are two main issues with this approach and other similar ones. First, this kind of algorithm only allows partonic calculations which are the simple product of a tree-level matrix element and Sudakov factors. In general, the resummed expression can contain various terms with differing logarithmic structures. A recent study [70, 71] of the production of highly energetic partons in soft-collinear effective theory (SCET) [66–69] has shown that the logarithmic structure is obtained by a sequence of matching and running calculations, which leads to a sum over terms, each with a different resummation of logarithms. These different terms can be identified with the different possible shower histories, which is due to the fact that the effective theory and the parton shower are both derived from the soft-collinear limit of QCD.

Second, reweighting tree-level calculations by Sudakov factors may result in inefficient event generators. This is because the product of many Sudakov factors can give numerically very different results for different events depending on the precise scales found by the k_T algorithm or any other jet identification procedure. Thus, if the initial set of events was distributed only with the correct power-singularity structure, the resulting events after including logarithmic information can have widely differing event weights, leading to a poor statistical efficiency. Of course, one can easily avoid this issue by accounting for both the power and logarithmic singularities at the time of initial event generation.

In **GenEvA** both of these issues are solved automatically, because the underlying phase space generator is itself a parton shower. First, we can assign different weights for different parton shower histories, allowing an easy implementation of the sum over different logarithmically resummed terms. Second, the events are directly generated based on the singularity structure of QCD and already include the leading-logarithmic effects. Thus, reweighting the parton shower to resummed partonic expressions will actually give much more uniform weights than reweighting to fixed-order matrix elements. We will come back to this issue in section 7.3 when discussing the efficiency of **GenEvA** and focus for now on how to implement resummed partonic calculations in **GenEvA**.

A particularly useful expression for resummed partonic distributions is given in the companion paper [1]. Starting with eq. (2.13), note that the parton shower by itself is creating events with $w(\Sigma) = 1$. Therefore, the internal **GenEvA** parton shower distributes events with an effective Σ -dependent “matrix element”

$$\sigma_i^{\text{shower}}[\Phi(\Sigma)] = \mathcal{P}(\Sigma)J(\Sigma) \equiv Q(\Sigma)\Delta(\Sigma), \quad (6.3)$$

¹²CKKW also corrects for the running of α_s .

where $Q(\Sigma)$ is the product of a hard scattering matrix element, splitting functions, and Jacobians, and $\Delta(\Sigma)$ is the product of Sudakov factors. Starting from the fixed-order tree-level matrix elements $\sigma^{\text{tree}}(\Phi)$, an LO/LL sample can be defined by (see eq. (5.9) in ref. [1])

$$\sigma_i^{\text{LO/LL}}[\Phi(\Sigma)] = \sigma^{\text{tree}}[\Phi(\Sigma)] \frac{Q(\Sigma)}{\sum_i Q[\Sigma_i(\Phi)]} \Delta(\Sigma), \quad (6.4)$$

where the sum runs over all shower histories $\Sigma_i(\Phi)$ that yield the same phase space point Φ . Therefore, the formula for the LO/LL event weight is quite simple¹³

$$w^{\text{LO/LL}}(\Sigma) = \frac{\sigma_i^{\text{LO/LL}}[\Phi(\Sigma)]}{\mathcal{P}(\Sigma)J(\Sigma)} = \frac{\sigma^{\text{tree}}[\Phi(\Sigma)]}{\sum_i Q[\Sigma_i(\Phi)]}. \quad (6.5)$$

Although eq. (6.4) includes a Sudakov factor, the LO/LL event weight no longer has any Sudakov information in it, because the internal **GenEvA** shower already distributes events with the correct double-logarithmic information. Since the Sudakov factors $\Delta(\Sigma)$ contained in $\mathcal{P}(\Sigma)$ in the LO weight in eq. (6.2) come at some computational cost, creating this LO/LL merged sample is computationally even more efficient than the LO result alone.

Note the difference between this result and CKKW. Here, we are using every shower history to define an appropriate Sudakov factor. The relative contribution of each Sudakov factor is given by

$$\frac{Q(\Sigma)}{\sum_i Q[\Sigma_i(\Phi)]}. \quad (6.6)$$

In those regions of phase space that are dominated by one shower history, one $Q(\Sigma) \rightarrow \infty$ and therefore one Sudakov dominates. This corresponds to the approach of ref. [26]. In CKKW [25], the chosen Sudakov factor always corresponds to the dominant history, whereas in eq. (6.4), away from the singular regions, an average Sudakov is used, more closely mimicking the SCET calculation [70, 71].

It would be interesting to study which approach better corresponds to experimental data, though we emphasize that the **GenEvA** algorithm can of course distribute events that are morally equivalent to the CKKW procedure, by simply choosing the partonic cross section

$$\sigma^{\text{CKKW}}(\Phi) = \sigma^{\text{tree}}(\Phi) \Delta[\Sigma_{\text{dom}}(\Phi)], \quad (6.7)$$

where $\Sigma_{\text{dom}}(\Phi)$ is the dominant shower history for the phase space point Φ .

6.3 Further logarithmic possibilities

GenEvA also allows a numerical treatment of an interesting single-logarithmic issue. In the LO matrix element eq. (6.1), we vetoed events for which two partons were “accidentally” closer than t_{cut} . This made sense, because we were not using the Sudakov information

¹³Since we use the **ALPHA** algorithm as described in section 5 to compute the sum over parton shower histories, the precise splitting functions $Q(\Sigma)$ in eq. (6.4) are those used in eq. (5.22). The actual splitting functions $\tilde{Q}(\Sigma)$ used in the shower slightly differ from these by the choice of E^{max} . Therefore, in practice, eq. (6.5) contains an additional factor $Q(\Sigma)/\tilde{Q}(\Sigma)$.

in the LO calculation. However, imagine the case that the tree-level matrix element were identically equal to the product of splitting functions

$$\sigma^{\text{tree}}[\Phi(\Sigma)] \rightarrow \sum_i Q[\Sigma_i(\Phi)], \quad (6.8)$$

such that the event weight in eq. (6.5) is always one. Because the parton shower preserves probability, the $t_{ij} < t_{\text{cut}}$ veto will now decrease the total cross section for $e^+e^- \rightarrow n$ jets from the Born value. Putting in the correct expression for $\sigma^{\text{tree}}(\Phi)$, this will result in a single-logarithmic decrease in the cross section. While this effect is formally beyond LO/LL order, it is numerically important for high-multiplicity matrix elements.

As we argued in figure 16, events that violate the t_{cut} constraint should be thought of as actually contributing to lower-dimensional phase space. Imagine we take an event sample and truncate to $\{t_{\text{cut}}, n_{\text{max}}\}$. As explained earlier, this gives $(n < n_{\text{max}})$ -parton events with $t_{\text{match}} = t_{\text{cut}}$, and n_{max} -parton events for which t_{match} changes from event to event. Even after truncation, there will be some n -parton events with $3 \leq n \leq n_{\text{max}}$ left that violate the t_{cut} constraint. Instead of vetoing these “rogue” events, we can further truncate them to some scale $t' > t_{\text{match}}$ until the t_{cut} violation disappears. This will take all rogue n -parton events and make them into $(< n)$ -parton events, recovering the cross section that would otherwise be lost to the t_{cut} veto. We will refer this additional truncation up to t' as “truncation-prime”.

Remember that these events were originally truncated to $\{t_{\text{cut}}, n_{\text{max}}\}$, so we have to restart the shower at the originally determined t_{match} scale (and not t'), otherwise we would double-count existing regions of phase space from the ordinary n -parton events. If the probability of getting a given n -body shower history Σ_n after truncation is $\mathcal{P}^{\text{trunc}}(\Sigma_n)$, then the probability $\mathcal{P}^{\text{trunc}'}(\Sigma_n)$ to get the same shower history after truncation-prime is given schematically by

$$\mathcal{P}^{\text{trunc}'}(\Sigma_n) = \mathcal{P}^{\text{trunc}}(\Sigma_n) + \sum_{i=1}^{n_{\text{max}}-n} \int_{\tilde{\mathbf{n}}} d(\Sigma_{n+i}/\Sigma_n) \mathcal{P}^{\text{trunc}}(\Sigma_{n+i}), \quad (6.9)$$

where $d(\Sigma_{n+i}/\Sigma_n)$ represents an integral over those $(n+i)$ -body shower variables that are not shared by the n -body shower, and $\tilde{\mathbf{n}}$ represents those regions of $(n+i)$ -body phase space that really look like n -jet events (e.g. the $\tilde{\mathbf{2}}$ regions in figure 16). Note that for $n = n_{\text{max}}$, truncation-prime does not affect the event probability.

If we now use the same event weight as in eq. (6.5) for the resulting event sample after truncation-prime, then this is equivalent to choosing

$$\sigma_i^{\text{LO/LL}'}[\Phi(\Sigma)] = \sigma^{\text{tree}}[\Phi(\Sigma)] \frac{Q(\Sigma)}{\sum_i Q[\Sigma_i(\Phi)]} \Delta(\Sigma) \frac{\mathcal{P}^{\text{trunc}'}(\Sigma)}{\mathcal{P}^{\text{trunc}}(\Sigma)}. \quad (6.10)$$

In other words, we can do the integrations in eq. (6.9) numerically by truncating the shower to t' . This $\mathcal{P}^{\text{trunc}'} / \mathcal{P}^{\text{trunc}}$ enhancement is formally beyond leading-logarithmic order, so it is safe to include it. Numerically, it helps yield more stable total cross sections as t_{cut} is varied. The main downside to using the numerical $\mathcal{P}^{\text{trunc}'} / \mathcal{P}^{\text{trunc}}$ method is that it

breaks the reversibility of further truncation. That is, once truncation-prime is applied, the probability for obtaining a set of four-momenta is proportional to $\mathcal{P}^{\text{trunc}'}$ and thus extremely convoluted, so one can no longer use the argument at the end of section 3.5 that reweighting can occur after hadronization.

7. Results

In this section, we show results for the process $e^+e^- \rightarrow n$ jets using the partonic cross sections defined in the previous section. As input for the fixed-order tree-level result $\sigma^{\text{tree}}(\Phi)$, we use **MadGraph** [79] to obtain numerical Fortran **HELAS** [80] routines for $e^+e^- \rightarrow n$ partons,¹⁴ where $2 \leq n \leq 6$, and a “parton” is either a gluon or a u , d , s , or c quark (neglecting quark masses). In this section, we always use the center-of-mass energy and phase space cut

$$E_{\text{CM}} = 1000 \text{ GeV}, \quad \sqrt{t_{\text{cut}}} = 100 \text{ GeV}, \quad (7.1)$$

unless otherwise noted.

We first validate the **GenEvA** algorithm by using the simple LO cross section from eq. (6.1) and comparing the output of **GenEvA** with that of **MadEvent** [14]. We then show some rudimentary results using the LO/LL cross sections in eqs. (6.4) and (6.10), with a more thorough discussion given in the companion paper [1]. Finally, **MadEvent** offers a useful benchmark to compare the efficiency of the **GenEvA** algorithm. We will see that **GenEvA**’s efficiency is at or above **MadEvent** levels, and that **GenEvA** itself is more efficient at distributing logarithmically-improved LO/LL results than pure tree-level LO results.

7.1 Comparison with MadEvent

Since **GenEvA** uses the same **MadGraph**/**HELAS** matrix-element engine as **MadEvent**, both programs should and do give identical results. For the purpose of this comparison we are not attaching any additional parton shower emissions, and simply compare the partonic cross sections for a fixed number of final state partons.

We first consider the results for the total cross sections, which are shown in table 1 for the n -parton samples combining all flavor subprocesses as well as for various individual flavor subprocesses. One can clearly see that the two programs agree within statistical uncertainties. Notice that while in **MadEvent** the relative uncertainty for each subprocess is roughly the same, in **GenEvA** the relative uncertainty scales as the square root of the cross section for the given subprocess. This is because **GenEvA** automatically produces a sample containing the sum of all subprocesses in a single run, with the ratio of the various subprocesses roughly given by the QCD symmetry structures. This means that the most important channels automatically get the most statistics and thus the smallest relative uncertainties, and no extra integration time is spent on less important channels. In **MadEvent** the individual subprocesses are generated separately and combined afterwards.

¹⁴For simplicity, we are only including the diagrams with $e^+e^- \rightarrow \gamma^* \rightarrow$ partons. An intermediate Z -boson could easily be included.

process	MadEvent	GenEvA	process	MadEvent	GenEvA
LO 3 (fb)	216.71 ± 0.21	216.77 ± 0.22	LO 5 (ab)	2542 ± 3	2543 ± 3
$u\bar{u}g$	86.62 ± 0.13	86.60 ± 0.18	$u\bar{u}ggg$	912 ± 2	912 ± 2
$d\bar{d}g$	21.75 ± 0.07	21.55 ± 0.10	$d\bar{d}ggg$	227.5 ± 0.9	228.3 ± 0.8
$s\bar{s}g$	21.63 ± 0.06	21.73 ± 0.10	$u\bar{u}d\bar{d}g$	33.8 ± 0.2	34.3 ± 0.4
$c\bar{c}g$	86.71 ± 0.13	86.70 ± 0.18	$u\bar{u}u\bar{u}g$	25.6 ± 0.2	25.7 ± 0.3
LO 4 (fb)	36.44 ± 0.04	36.49 ± 0.04	LO 6 (ab)	67.9 ± 0.3	68.0 ± 0.2
$u\bar{u}gg$	14.00 ± 0.03	14.00 ± 0.02	$u\bar{u}gggg$	22.41 ± 0.09	22.29 ± 0.12
$d\bar{d}gg$	3.504 ± 0.013	3.511 ± 0.011	$u\bar{u}u\bar{u}gg$	1.117 ± 0.006	1.14 ± 0.03
$u\bar{u}d\bar{d}$	0.175 ± 0.001	0.180 ± 0.003	$u\bar{u}u\bar{u}u\bar{u}$	0.005 ± 0.001 ⁻	0.005 ± 0.001
$u\bar{u}u\bar{u}$	0.132 ± 0.001	0.132 ± 0.002	$u\bar{u}d\bar{d}s\bar{s}$	0.019 ± 0.001 ⁻	0.020 ± 0.005

Table 1: Comparison of total cross sections between **GenEvA** and **MadEvent**. Both programs agree within statistical uncertainties (0.001⁻ indicates an error too small to be reported by **MadEvent**). In **MadEvent**, the relative uncertainty on each sample is roughly constant, whereas in **GenEvA**, less populated channels are allowed to have higher relative uncertainty. This happens because the ratio of various subprocesses is given in **GenEvA** by the QCD symmetry structures, and no extra integration time is spent determining the ratio between different channels.

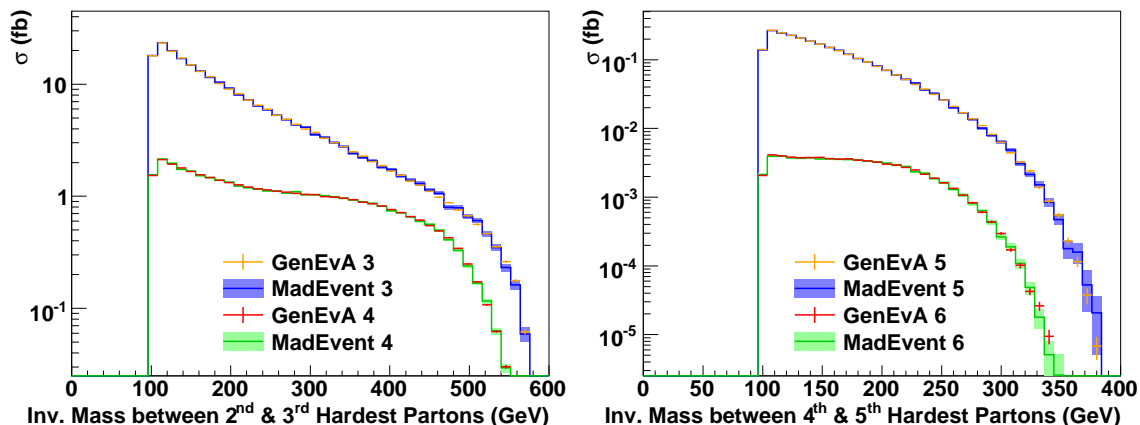


Figure 17: Comparison between **MadEvent** and **GenEvA** for 3-, 4-, 5-, and 6-parton tree-level matrix elements. Left panel: The pairwise invariant mass of the 2nd and 3rd hardest partons for 3- and 4-parton matrix elements. Right panel: The pairwise invariant mass of the 4th and 5th hardest partons for 5- and 6-parton matrix elements. The two programs agree within the statistical uncertainties, validating the **GenEvA** algorithm since both programs use the same **HELAS/MadGraph** matrix elements. Note the hard cutoff at $\sqrt{t_{\text{cut}}} = 100$ GeV.

Next, we compare the differential distributions generated by both programs for a few simple partonic observables. We have checked a large range of partonic observables and in all cases the distributions obtained from **GenEvA** and **MadEvent** agree perfectly within the statistical uncertainties, verifying that the **GenEvA** phase space generator is working properly. As an illustration, we show in figure 17 the distribution of the pairwise invariant mass between the 2nd and 3rd hardest partons for 3- and 4-parton matrix elements (left panel) and between the 4th and 5th hardest partons for 5- and 6-parton matrix elements (right panel). Here, hardest is defined by the parton with the largest energy. We will

discuss the different efficiencies of the two programs later on in section 7.3.

7.2 Logarithmically-improved results

The main usefulness of an LO/LL partonic calculation is for creating fully merged LO/LL samples interfaced with a parton shower. We do this in the companion paper [1] and here only give partonic results, which by themselves have only limited usefulness, but show the effect of Sudakov suppression.

From eq. (6.4), we see that the LO/LL partonic calculation depends on a Sudakov factor, which in turn depends on the t_{match} scale of the event after truncation. For an n -body matrix element with fixed n , there are two different cases, as discussed in section 3.5: If $n = n_{\text{max}}$, then $t_{\text{match}} > t_{\text{cut}}$ is an event-specific scale corresponding to the last emission and the result should be regarded as an inclusive sample. If $n < n_{\text{max}}$, we always have $t_{\text{match}} = t_{\text{cut}}$ and the result should be regarded as exclusive. For $n < n_{\text{max}}$, we also have the choice of whether to include the truncation-prime correction factor in eq. (6.10). These three possibilities are labeled as¹⁵

$$\begin{aligned}
 \text{LO/LL inc. } n &\equiv \text{eq. (6.4) with } n = n_{\text{max}} , \\
 \text{LO/LL exc. } n &\equiv \text{eq. (6.4) with } n < n_{\text{max}} , \\
 \text{LO/LL}' \text{ exc. } n &\equiv \text{eq. (6.10) with } n < n_{\text{max}} .
 \end{aligned}
 \tag{7.2}$$

Differential distributions for these three possibilities are shown in figure 18 for $n = 3$, i.e. $e^+e^- \rightarrow q\bar{q}g$. Again, the hardest parton is defined as the parton with the largest energy. When the invariant mass between the 1st and 2nd hardest partons is at the lower threshold, the three partons are well separated, with a large emission scale for the gluon. Since the LO/LL inc. sample only captures the logarithms between E_{CM} and the gluon emission scale, it contains no Sudakov suppression in this part of phase space, and hence asymptotes to the LO result. The LO/LL exc. sample has an additional overall Sudakov suppression from leading-logarithmic running from the gluon emission scale down to the fixed matching scale $t_{\text{match}} = t_{\text{cut}}$. The upper endpoint is the 2-jet region with a low gluon-emission scale, so all three LO/LL samples are Sudakov suppressed compared to the LO sample. The LO/LL' exc. sample has roughly the same shape as the LO/LL exc. sample but the overall scale is increased, because the $\tilde{\mathbf{n}}$ integrations it contains restore some of the lost cross section from phase space vetoes. The reason it asymptotes to the LO/LL inc. sample at the upper endpoint is because in this region it is most likely for two partons to be “accidentally” closer than t_{cut} . The identical effect can be seen in the right panel when the invariant mass between the 1st and 3rd hardest partons is small. Note that the Sudakov suppression from the LO result to the LO/LL results is formally double-logarithmic, while the increase from the LO/LL exc. to the LO/LL' exc. result due to the inclusion of the $\tilde{\mathbf{n}}$ integrations is formally single-logarithmic.

¹⁵In the companion paper [1], we only show LO/LL inc. and LO/LL' exc. results and will therefore remove the prime symbol from the LO/LL notation.

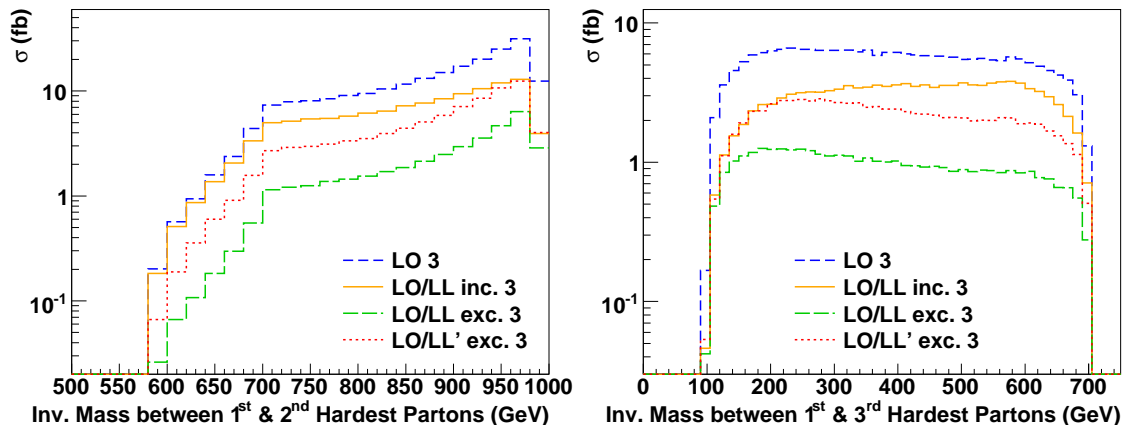


Figure 18: The effect of Sudakov factors on 3-parton matrix elements $e^+e^- \rightarrow q\bar{q}g$. In the LO/LL inc. sample, the Sudakov suppression only captures the logarithms between E_{CM} and the gluon emission scale. In the LO/LL exc. sample, there are additional no-branching probabilities from the gluon emission scale down to the fixed $t_{\text{match}} = t_{\text{cut}}$ scale, giving effectively an “exclusive” 3-jet result with an overall Sudakov suppression. The LO/LL’ exc. sample includes the \tilde{n} correction factor from eq. (6.10), while the LO/LL exc. sample does not. Left panel: Invariant mass between 1st and 2nd hardest partons. The LO and LO/LL inc. samples agree near the lower endpoint, because this is the 3-jet region where Sudakov suppression is not active. The upper endpoint is the 2-jet region, and the \tilde{n} integrations in the LO/LL’ exc. sample restore lost cross section compared to the LO/LL exc. sample, asymptoting to the LO/LL inc. result. Right panel: Invariant mass between 1st and 3rd hardest partons, which shows similar effects. In the companion paper [1], we will combine the LO/LL partonic results to create a merged LO/LL sample.

7.3 The efficiency of GenEvA

We argued in section 1.3 that **GenEvA** has several advantages over more traditional algorithms, so it is interesting to see how its efficiency compares to other event generators. We begin by addressing the speed of **GenEvA** for fixed-order calculations by comparing it to **MadEvent** [14], and then show how **GenEvA** is optimized to generate resummed partonic results.

Comparing these two programs is not straightforward, since the different methods used for phase space generation make it difficult to design a fair comparison. First, **GenEvA** is written in C++, whereas **MadEvent** is written in Fortran. Thus different levels of optimization in the compilers can affect the results. However, given that both programs use the same Fortran **MadGraph**/HELAS matrix elements and because evaluating the matrix element is the slowest part of **GenEvA**, the comparison is reasonable though not optimal. Second, in a single run **GenEvA** directly produces a fully inclusive sample, containing a sum over all subprocesses and multiplicities, whereas **MadEvent** generates events for each subprocess and multiplicity separately and combines them at the end of the run. Thus, **MadEvent** is producing many events that are thrown out at the end of the day. This is a source of inefficiency that could easily be ameliorated by first calculating the approximate cross sections for the various subprocesses and then generating only the events required. In our speed comparisons, we therefore give timing information for n -parton samples with fixed n and

	η_{eff}	T_{eff} (msec)	$T_{0.9}$ (msec)
GenEvA LO 3	0.789	0.57	0.62
GenEvA LO/LL inc. 3	0.965	0.47	< 0.47
MadEvent 3	0.982	2.6	< 2.6
MadEvent $u\bar{u}g$	0.994	3.0	< 3.0
GenEvA LO 4	0.525	1.7	2.2
GenEvA LO/LL inc. 4	0.713	1.3	1.5
MadEvent 4	0.809	11.1	11.4
MadEvent $u\bar{u}gg$	0.752	5.4	5.7
GenEvA LO 5	0.390	10.0	15
GenEvA LO/LL inc. 5	0.557	8.6	10.8
MadEvent 5	0.843	62	64
MadEvent $u\bar{u}ggg$	0.833	27	27
GenEvA LO 6	0.298	160	250
GenEvA LO/LL inc. 6	0.396	150	230
MadEvent 6	0.809	1900	2300
MadEvent $u\bar{u}gggg$	0.784	330	350

Table 2: Comparison of the speed and efficiency between **GenEvA** and **MadEvent**. η_{eff} is the statistical efficiency defined in eq. (2.17), and T_{eff} is the time required to create one statistical event. $T_{0.9}$ is the time to create one statistical event in a partially unweighted sample with $\eta_{\text{eff}}(w_0) = 0.9$, as defined in eq. (2.19). Note that **GenEvA** performs better on Sudakov-improved (LO/LL) results than on tree-level (LO) matrix elements. **GenEvA** is in all cases faster than **MadEvent**, in most channels by a factor of a few, despite the fact that **GenEvA**'s efficiencies are lower since it is a fixed-grid algorithm.

summed over all subprocesses for both **GenEvA**¹⁶ and **MadEvent**, as well as for **MadEvent** samples with only one flavor subprocess.

The efficiency results are shown in table 2, where 3-, 4-, 5-, and 6-parton matrix elements are compared. Shown are the statistical efficiencies η_{eff} defined in eq. (2.17), as well as the time T_{eff} needed to create one statistically equivalent event. Because **GenEvA** is a fixed-grid algorithm, its statistical efficiency η_{eff} is lower than that of **MadEvent**, where grid optimization is used. However, the time needed to create a statistically equivalent weighted event sample is smaller for **GenEvA** than **MadEvent**, presumably because of the speed of the parton-shower-based phase space algorithm. To assess this efficiency/speed tradeoff, we also report $T_{0.9}$, the time needed to create one statistically equivalent event in a partially unweighted event sample with $\eta_{\text{eff}}(w_0) = 0.9$ as defined in eq. (2.19). That is, we unweight just those events with weights below a threshold weight w_0 , with w_0 chosen such that the final event sample has a 90% statistical efficiency. By this $T_{0.9}$ measure, **GenEvA** is still faster than **MadEvent** in all channels, often by a factor of a few.

We do not show comparisons of the unweighting efficiencies between **GenEvA** and

¹⁶To obtain an n -parton sample with fixed n in **GenEvA** we simply set $n = n_{\text{max}}$ and discard events with fewer than n final state partons.

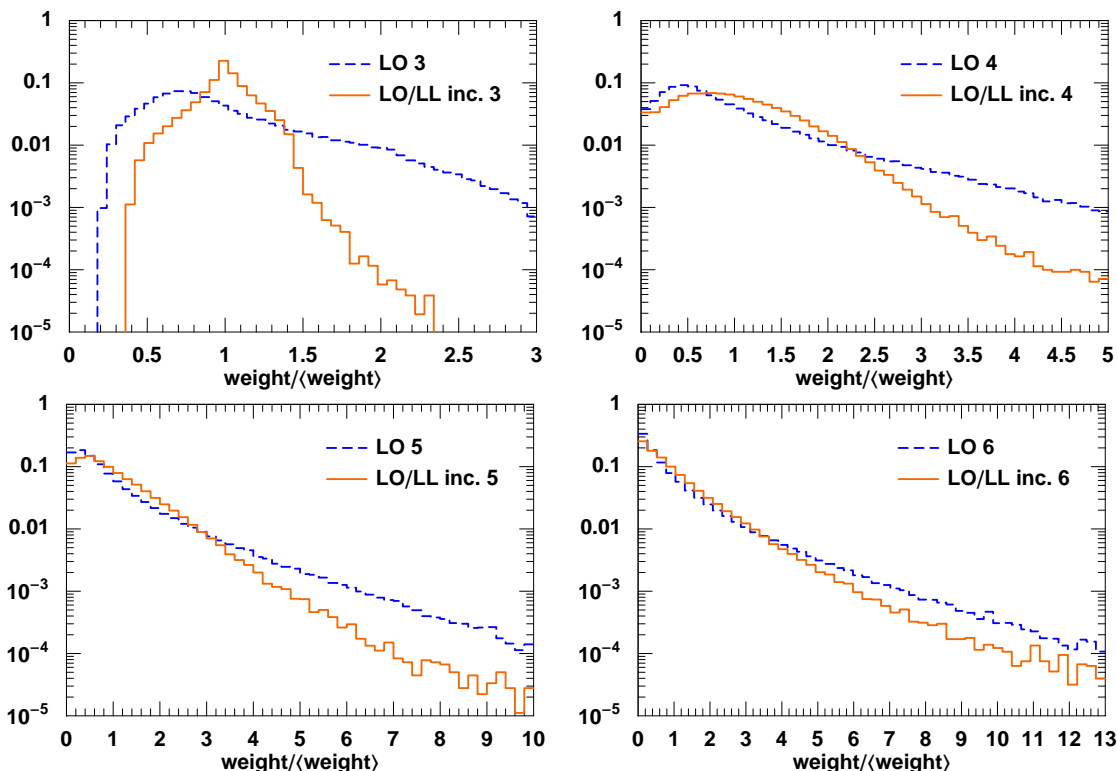


Figure 19: Comparison of the weight distributions between **GenEvA** LO and LO/LL samples for 3- to 6-parton matrix elements. Because the **GenEvA** phase space generator already includes leading-logarithmic information, the LO/LL weight distributions are more strongly peaked at their average weights. More importantly, there is a considerable suppression of the high-weight tail when going from the LO to LO/LL samples. As expected, the effect of leading logarithms is less relevant for the higher-multiplicity samples, as phase space suppression keeps the double-logarithms from growing too large.

MadEvent, because there is no reasonable way to do so. As argued in section 2.3, unweighting efficiencies are sensitive to the exponential tails of the weight distribution. Depending on how long one runs the two programs, one may or may not sample these tails, giving wildly varying unweighting efficiencies during one run and between different runs. Furthermore, it is completely sufficient for all practical purposes to have *partially* unweighted samples with a reasonably high statistical efficiency.

As we discussed above, the fact that **GenEvA** is using a parton shower as a phase space generator makes it ideal to distribute events according to logarithmically-improved partonic calculations. This is because the parton shower already contains the correct double-logarithmic behavior, such that the resulting weights are expected to be uniform, giving rise to a high statistical efficiency. To see this, consider the event weight given in eq. (6.5). Since the splitting functions reproduce the tree-level matrix elements of QCD in the collinear limit, the denominator has exactly the same singular behavior as the numerator, with no logarithmic differences.

This increase in efficiency for LO/LL event generation can be seen in a few different

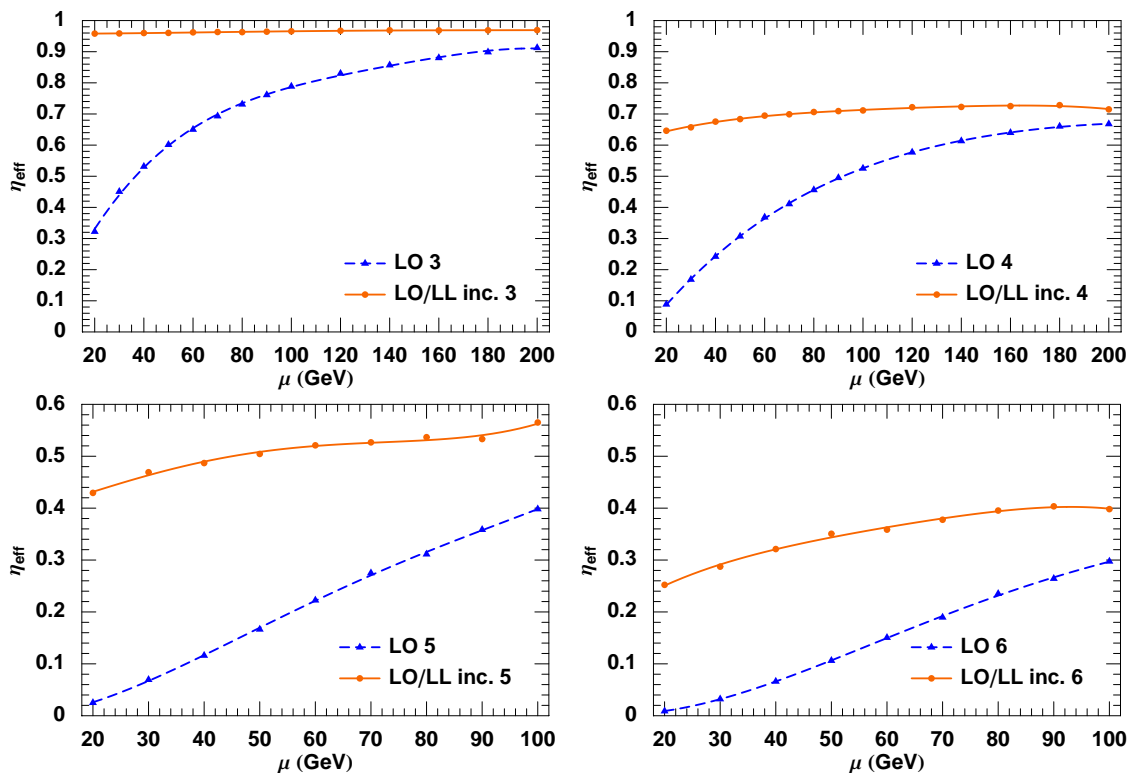


Figure 20: Comparison of the statistical efficiencies η_{eff} between GenEvA LO and LO/LL samples for 3- to 6-parton matrix elements. As the matching scale $\mu = \sqrt{t_{\text{match}}}$ is lowered, the efficiency of the LO/LL sample stays roughly constant, while the efficiency of the LO sample drops dramatically.

ways, the simplest being that in table 2, the three different efficiency/speed measures are all better for the GenEvA LO/LL samples than for the GenEvA LO samples. In figure 19, we show the distribution of event weights between the LO/LL and LO samples. For $n = 3$ especially, the weight distribution is far more peaked near the average weight in the LO/LL sample compared to the LO sample, showing that the inclusion of leading-logarithmic information softens the high-weight tails. Even for $n = 6$, where Sudakov logarithms play a much smaller role in determining the differential distributions, there is still a factor of 5 decrease in the number of high-weight events. The change in efficiency is shown even more dramatically in figure 20, which compares the statistical efficiency between the LO/LL and LO samples as the matching scale μ is changed. For the LO/LL samples, the efficiency is nearly constant as μ varies, while the LO samples exhibit a sharp decrease in efficiency for $\mu \rightarrow 0$.

Thus, GenEvA is faster at distributing resummed results than tree-level results. We note that the efficiency of GenEvA could improve further with the use of a more realistic internal parton shower. For example, tree-level matrix elements include the effect of color coherence which is absent in the virtuality-ordered shower currently used in GenEvA. If a p_{\perp} -ordered shower could be implemented as discussed in section 8.3, then the leading color coherence effects would already be captured by the phase space generator, which should lead to more uniform weights.

8. Conclusions

We have shown how to construct a phase space generator from a parton shower. The resulting **GenEvA** algorithm uses a very different strategy compared to more conventional approaches. Most available generators use a non-recursive, multi-channel, adaptive-grid algorithm to distribute events in phase space directly according to a desired distribution, which are often tree-level matrix elements. In contrast, **GenEvA** uses a parton shower to generate events, and in a second step reweights the events to the desired distribution.

Though the **GenEvA** algorithm uses a fixed grid with no adaptation, the recursive nature of the parton shower allows for an incredibly fast generation of phase space points. Since the parton shower has the same singularity structure as QCD, the effectively fixed grid used in the **GenEvA** algorithm is presumably very close to the final adapted grid in traditional phase space generators. As a result, the **GenEvA** algorithm is able to compete with and likely even outperform traditional generators.

The main advantage of the **GenEvA** algorithm arises when one goes beyond fixed-order calculations. It is well known that leading-logarithmic terms have to be taken into account in order to obtain realistic QCD predictions. In particular, when merging matrix element calculations with subsequent parton shower evolution, leading-logarithmic information is essential to avoid strong dependence on the unphysical matching scale. Unfortunately, analytic expressions for logarithmically improved, fully differential distributions are prohibitively difficult to obtain for a large number of external particles, since the required scales in the logarithms depend in a complicated way on all kinematic invariants in the event.

For this reason, traditional generators often implement leading-logarithmic improvements after the initial phase space generation is complete, and use the known kinematics of the event in order to calculate the required scales in the logarithms. In contrast, the parton shower used to generate phase space points in **GenEvA** automatically contains the leading logarithmic behavior of QCD by construction, and for this reason **GenEvA** is in fact more efficient in distributing logarithmically improved distributions than fixed-order calculations and thus has the potential to outperform traditional techniques.

To conclude, we discuss some potential extensions to the **GenEvA** algorithm to make it useful for LHC physics.

8.1 Towards hadronic collisions

Currently, the **GenEvA** algorithm is only designed to work for e^+e^- collisions with no initial state photon radiation. The technical reason for building **GenEvA** for e^+e^- first is that the final state parton shower with fixed center-of-mass energy is easier to understand and verify. However, the name **GenEvA** is obviously inspired by the site of the upcoming LHC experiment at CERN. We anticipate extending the **GenEvA** algorithm to hadronic collisions, and one might wonder whether there are any additional technical hurdles to overcome.

To build **GenEvA** for e^+e^- collisions, there were three technical problems that needed to be solved. First, in order to generate events that conserve energy-momentum, parton showers need to adjust the kinematics of the parton splittings after they are generated, which typically makes it impossible to calculate the probability analytically. Second, different

parton shower histories can give rise to the same points in Lorentz-invariant phase space, such that a parton shower covers phase space multiple times. Third, one needs a way to “truncate” a high-multiplicity event generated by the parton shower to a lower-multiplicity event for which the probability is still known analytically.

In this paper we explained in detail how these technical issues are solved in the **GenEvA** algorithm. The first issue is solved by employing the analytic parton shower developed in ref. [52]. The second issue is solved by introducing the overcounting factor which multiplies the generated event weight. Finally, the double-branch probability used in this paper enforces a global evolution, which allows for a simple truncation algorithm.

These technical solutions should have straightforward generalizations to hadronic collisions. While we know of no analytically calculable initial state shower, building one should not be too difficult as it is well known that phase space can be recursively described both in terms of an s -channel “final state” tree and a t -channel “initial state” tree. The trick to creating the analytic final state shower was to use $\cos\theta$ as the primary variable instead of z , and an analogous trick should be possible for the initial state shower. The issue of double-counting will require figuring out how to deal with interference between initial state and final state shower configurations, but as a last resort, one can always do a brute force sum over parton shower histories to solve the double-counting issue. Truncation should be more-or-less identical to the final state shower, because many initial state showers are based on backwards evolution, and therefore can have a notion of well-ordering.

Hence, the main new issue in hadronic collisions is the proper treatment of parton distribution functions (PDFs). However, apart from issues of efficiency, this is really not an issue, because parton distribution functions simply introduce two new variables x_1 and x_2 and define a new choice of partonic calculation

$$d\sigma = f_1(x_1, Q_1^2) f_2(x_2, Q_2^2) dx_1 dx_2 d\hat{\sigma}_{12 \rightarrow X} \equiv \sigma'(\Phi') d\Phi', \quad (8.1)$$

where Φ' includes $x_{1,2}$ in addition to the ordinary phase space variables, and $\sigma'(\Phi')$ is the fully differential cross section including PDF information. The point of reweighting is that no matter how events are selected in Φ' , events can be distributed according to $\sigma'(\Phi')$, and it is the user’s responsibility to, for example, use NLO PDFs with NLO calculations, use the proper scales for $Q_{1,2}^2$, and so on.

8.2 Heavy resonances

At the LHC, the most interesting Standard Model backgrounds to Beyond the Standard Model signals come not from QCD itself but from heavy resonances like W , Z , and t created in association with QCD radiation. Because **GenEvA** is based on a parton shower, one might wonder whether it is possible to include heavy resonances in the **GenEvA** algorithm.

For this purpose, it is useful to think of **GenEvA** not as a parton shower but as simply a recursive phase space generator, and a resonances corresponds to just a funny choice of “splitting function”. Because events will be reweighted at the end of the day to the desired distribution, all what is important is that the “splitting functions” have the right singularity structure, which for the purposes of heavy resonances is the Breit-Wigner line shape.

What will be particularly bizarre about the **GenEvA** algorithm with heavy resonances is that there will be “Sudakov factors” associated with resonances, but as long as the resonances are narrow enough, this Sudakov factor will either be unity if $t_{\text{cut}} > m^2$ or zero if $t_{\text{cut}} < m^2$, meaning that a resonance is either on shell or it is decayed. This also means that even if $t_{\text{cut}} > m^2$ and a resonance is treated as on shell for the matrix element calculation, if the **GenEvA** shower is restarted at t_{cut} , the Breit-Wigner distribution will be recovered for the resonance with the rest of the shower history compensating for the change in energy-momentum.

Once the logistics for handling mass thresholds with heavy resonance is established, the treatment of not-so-heavy resonances like the bottom quark will be straightforward, and one would use the ordinary massive Altarelli-Parisi splitting functions for b and c quarks. The top quark is an interesting case, because in **GenEvA**, the top would both decay via $t \rightarrow Wb$ but also radiate through $t \rightarrow tg$. Both of these processes need to coexist in **GenEvA** in order to cover all of multiplicity, flavor, and phase space, and this dual nature of the top may be interesting for understanding Monte Carlo effects in top-quark mass measurements.

8.3 Choice of evolution variable

For the **GenEvA** algorithm to be relevant in the LHC era, it will have to have a smooth interface with a hadronization/underlying event model. Regardless of the phase space strategy, every parton-based Monte Carlo program eventually has to interface to some external program that is responsible for hadronizing the initial and final state partons into the hadrons observed in the experiment and also to deal with secondary interactions, beam remnants, pile up, and other nonperturbative effects present in any hadronic collision.

In figure 14 we argued that because **GenEvA** is based on a parton shower and because there already exists a smooth interface between the showering regime and the hadronization regime in existing programs, then **GenEvA** should inherit the same smooth interface. However, as mentioned already, the internal **GenEvA** shower is optimized for speed and is not really meant to provide the best possible modeling of small angle QCD emissions. For example, the **GenEvA** algorithm currently evaluates α_s at a fixed scale and does not implement color coherence. Also, it uses virtuality as the evolution variable and not p_{\perp} as modern showers do. While it is in principle possible to construct an analytic parton shower which includes all these effects, this might not be the most efficient strategy.

As discussed in section 3.5, instead of using the internal shower in the showering regime, **GenEvA** can rely on an external parton shower as long as the external parton shower starts evolving from the event-defined matching scale t_{match} . From a physics point of view, it is important that the double-logarithmic dependence on t_{match} cancel between the partonic and showering regimes, but this is difficult if the evolution variables differ in the two regimes. In this paper, we have been using virtuality as the matching variable, so the current version of the **GenEvA** algorithm can therefore only be smoothly interfaced with a virtuality-ordered parton shower. How then, can **GenEvA** interface with more modern p_{\perp} -ordered showers?

This turns out to be an algorithmic as opposed to a conceptual problem, though. All **GenEvA** requires to map out phase space is a parton shower with no dead zones for which an analytic formula exists for calculating the probability for generating a given point in phase space. As a phase space generator, a p_{\perp} -ordered shower is no better or worse than a virtuality-ordered shower, because at the level of phase space, p_{\perp} -ordering is equivalent to virtuality-ordering with an additional veto to eliminate certain unwanted shower histories.

The main problem with a p_{\perp} -ordered shower is that there is no “local” way to determine whether a parton shower history is p_{\perp} -ordered or not. With a virtuality-ordered shower, every way of combining four-momenta that is consistent with flavor gives a valid shower history, allowing **GenEvA** to use the numerically efficient **ALPHA** algorithm [77, 78] to “count” the number of parton shower histories that give identical final state four-momenta. It is impossible to use the **ALPHA** algorithm to count p_{\perp} -ordered shower histories (or even virtuality-ordered shower histories with an angular-ordering veto), because whether or not a history is p_{\perp} -ordered (or angular-ordered) requires information about previous branches.

It would be straightforward to rewrite **GenEvA** to use a p_{\perp} -ordered shower to map out phase space as long as one could find a way to efficiently count p_{\perp} -ordered shower histories. One idea is to generate the full p_{\perp} -ordered expression for $\sum_i \hat{\alpha}(\Sigma_i)$, and then use some variant of the **O’Mega** algorithm [81] to generate numerically efficient code to calculate this sum. In other words, we can treat $\sum_i \hat{\alpha}(\Sigma_i)$ as a “matrix element” by itself, and construct an expression for it offline.

One additional subtlety comes from the new interleaved evolution introduced in **Pythia** [6]. In order for reweighting to make sense, **GenEvA** assumes that any external showering program will not affect the existing **GenEvA** shower history. In interleaved evolution, the primary interaction is not independent of secondary ones and the underlying event competes with the hard scattering for the momentum fraction of the incoming partons. Again, this is an algorithmic and not a conceptual problem. One could always write a more complicated version of **GenEvA** around an interleaved shower, and as long as the probability to obtain a given (interleaved) shower history is known, the **GenEvA** algorithm will be more or less unchanged. Practically speaking, though, the matching scale can probably be taken sufficiently large such that **Pythia** interleaved evolution starting from the **GenEvA** cutoff will give a reasonable description of the hadronic regime.

8.4 Outlook

Once the technical issues involving hadronic collisions, heavy resonances, and evolution variables are solved, there is a final technical question about how the **GenEvA** algorithm will implement more advanced matrix elements, such as those that involve complicated loop calculations [82], or those based on effective theories [71, 70, 72]. As currently written, the **GenEvA** algorithm assumes that there is a unique answer for the value of $\sigma(\Phi)$, but depending on the analytic form of $\sigma(\Phi)$, it may be more convenient to use Monte Carlo methods not only to sample phase space but also to determine $\sigma(\Phi)$. While the specific implementation will affect the performance of the **GenEvA** algorithm, the idea of additional “hidden” Monte Carlo variables being used to determine $\sigma(\Phi)$ is still compatible with the

GenEvA algorithm's use of reweighting. GenEvA can in principle implement any desired differential cross section as long as it is well-defined on paper.

Combined with the GenEvA framework [1], the GenEvA algorithm has the potential to significantly improve our understanding of Standard Model backgrounds. The ability to conceive of an event simultaneously as a shower history and as a fixed-body phase space point allows one to answer questions that are difficult to address in either language individually. The efficiency of the GenEvA algorithm makes it a worthy competitor to existing methods, and the automatic inclusion of leading-logarithmic effects raises the bar for what the experimental collaborations can expect from the Monte Carlo community 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 a DOE OJI award and an LDRD grant from LBNL. JT is supported by a fellowship from the Miller Institute for Basic Research in Science.

References

- [1] C.W. Bauer, F.J. Tackmann and J. Thaler, *GenEvA (I): a new framework for event generation*, *JHEP* **12** (2008) 010 [[arXiv:0801.4026](#)].
- [2] V.N. Gribov and L.N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438 [*Yad. Fiz.* **15** (1972) 781].
- [3] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, *Nucl. Phys.* **B 126** (1977) 298.
- [4] 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].
- [5] T. Sjöstrand et al., *High-energy-physics event generation with PYTHIA 6.1*, *Comput. Phys. Commun.* **135** (2001) 238 [[hep-ph/0010017](#)].
- [6] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026 [[hep-ph/0603175](#)].
- [7] 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](#)].
- [8] 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](#)].

- [9] S. Gieseke et al., *HERWIG++ 2.0 release note*, [hep-ph/0609306](#).
- [10] L. Lönnblad, *ARIADNE version 4: a program for simulation of QCD cascades implementing the color dipole model*, *Comput. Phys. Commun.* **71** (1992) 15.
- [11] 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](#).
- [12] T. Gleisberg et al., *SHERPA 1.0, a proof-of-concept version*, *JHEP* **02** (2004) 056 [[hep-ph/0311263](#)].
- [13] 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](#)].
- [14] F. Maltoni and T. Stelzer, *MadEvent: automatic event generation with MadGraph*, *JHEP* **02** (2003) 027 [[hep-ph/0208156](#)].
- [15] COMPHEP collaboration, E. Boos et al., *CompHEP 4.4: automatic computations from Lagrangians to events*, *Nucl. Instrum. Meth.* **A534** (2004) 250 [[hep-ph/0403113](#)].
- [16] F. Krauss, R. Kuhn and G. Soff, *AMEGIC++ 1.0: a matrix element generator in C++*, *JHEP* **02** (2002) 044 [[hep-ph/0109036](#)].
- [17] W. Kilian, T. Ohl and J. Reuter, *WHIZARD: simulating multi-particle processes at LHC and ILC*, [arXiv:0708.4233](#).
- [18] A. Kanaki and C.G. Papadopoulos, *HELAC: a package to compute electroweak helicity amplitudes*, *Comput. Phys. Commun.* **132** (2000) 306 [[hep-ph/0002082](#)].
- [19] C.G. Papadopoulos, *PHEGAS: a phase space generator for automatic cross-section computation*, *Comput. Phys. Commun.* **137** (2001) 247 [[hep-ph/0007335](#)].
- [20] A. Cafarella, C.G. Papadopoulos and M. Worek, *Helac-Phegas: a generator for all parton level processes*, [arXiv:0710.2427](#).
- [21] R.K. Ellis and J. Campbell, *MCFM - Monte Carlo for FeMtobarn processes*, <http://mcfm.fnal.gov>.
- [22] Z. Nagy, *NLOJet++*, <http://nagyz.web.cern.ch/nagyz/Site/NLOJet++.html>.
- [23] P. Aurenche et al., *The PHOX family*, http://wwwlapp.in2p3.fr/lapth/PHOX_FAMILY/main.html.
- [24] M. Bähr et al., *VBFNLO - NLO parton level Monte Carlo for vector boson fusion*, <http://www-itp.particle.uni-karlsruhe.de/~vbfnlweb/>.
- [25] S. Catani, F. Krauss, R. Kuhn and B.R. Webber, *QCD matrix elements + parton showers*, *JHEP* **11** (2001) 063 [[hep-ph/0109231](#)].
- [26] L. Lönnblad, *Correcting the colour-dipole cascade model with fixed order matrix elements*, *JHEP* **05** (2002) 046 [[hep-ph/0112284](#)].
- [27] F. Krauss, *Matrix elements and parton showers in hadronic interactions*, *JHEP* **08** (2002) 015 [[hep-ph/0205283](#)].
- [28] M.L. Mangano, *The so-called MLM prescription for ME/PS matching* (2004), talk presented at the *Fermilab ME/MC Tuning Workshop*, October 4 2004.
- [29] S. Mrenna and P. Richardson, *Matching matrix elements and parton showers with HERWIG and PYTHIA*, *JHEP* **05** (2004) 040 [[hep-ph/0312274](#)].

- [30] A. Schälicke and F. Krauss, *Implementing the ME+PS merging algorithm*, *JHEP* **07** (2005) 018 [[hep-ph/0503281](#)].
- [31] N. Lavesson and L. Lönnblad, *W + jets matrix elements and the dipole cascade*, *JHEP* **07** (2005) 054 [[hep-ph/0503293](#)].
- [32] S. Höche et al., *Matching parton showers and matrix elements*, [hep-ph/0602031](#).
- [33] 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](#)].
- [34] 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](#)].
- [35] N. Lavesson and L. Lönnblad, *Merging parton showers and matrix elements — back to basics*, *JHEP* **04** (2008) 085 [[arXiv:0712.2966](#)].
- [36] Z. Nagy and D.E. Soper, *Parton showers with quantum interference*, *JHEP* **09** (2007) 114 [[arXiv:0706.0017](#)].
- [37] Z. Nagy and D.E. Soper, *Parton showers with quantum interference: leading color, spin averaged*, *JHEP* **03** (2008) 030 [[arXiv:0801.1917](#)].
- [38] 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](#)].
- [39] J.C. Collins, *Subtraction method for NLO corrections in Monte Carlo event generators for lepton production*, *JHEP* **05** (2000) 004 [[hep-ph/0001040](#)].
- [40] 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](#)].
- [41] 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](#)].
- [42] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029 [[hep-ph/0204244](#)].
- [43] 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](#)].
- [44] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040 [[hep-ph/0409146](#)].
- [45] 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](#)].
- [46] 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](#)].
- [47] 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](#)].
- [48] 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](#)].
- [49] D.E. Soper, *Next-to-leading order QCD calculations with parton showers. II: soft singularities*, *Phys. Rev. D* **69** (2004) 054020 [[hep-ph/0306268](#)].

- [50] Z. Nagy and D.E. Soper, *Matching parton showers to NLO computations*, *JHEP* **10** (2005) 024 [[hep-ph/0503053](#)].
- [51] 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](#)].
- [52] C.W. Bauer and F.J. Tackmann, *Gaining analytic control of parton showers*, *Phys. Rev. D* **76** (2007) 114017 [[arXiv:0705.1719](#)].
- [53] M.H. Seymour, *Matrix element corrections to parton shower algorithms*, *Comput. Phys. Commun.* **90** (1995) 95 [[hep-ph/9410414](#)].
- [54] M.H. Seymour, *A simple prescription for first order corrections to quark scattering and annihilation processes*, *Nucl. Phys. B* **436** (1995) 443 [[hep-ph/9410244](#)].
- [55] M. Bengtsson and T. Sjöstrand, *Coherent parton showers versus matrix elements: implications of PETRA - PEP data*, *Phys. Lett. B* **185** (1987) 435.
- [56] M. Bengtsson and T. Sjöstrand, *A comparative study of coherent and noncoherent parton shower evolution*, *Nucl. Phys. B* **289** (1987) 810.
- [57] G. Miu, *A matrix-element-based modification of the parton shower*, [hep-ph/9804317](#).
- [58] G. Miu and T. Sjöstrand, *W production in an improved parton shower approach*, *Phys. Lett. B* **449** (1999) 313 [[hep-ph/9812455](#)].
- [59] E. Norrbin and T. Sjöstrand, *QCD radiation off heavy particles*, *Nucl. Phys. B* **603** (2001) 297 [[hep-ph/0010012](#)].
- [60] G.P. Lepage, *A new algorithm for adaptive multidimensional integration*, *J. Comput. Phys.* **27** (1978) 192.
- [61] S. Kawabata, *A new version of the multidimensional integration and event generation package BASES/SPRING*, *Comput. Phys. Commun.* **88** (1995) 309.
- [62] A. Pukhov et al., *CompHEP: a package for evaluation of Feynman diagrams and integration over multi-particle phase space. User's manual for version 33*, [hep-ph/9908288](#).
- [63] R. Kleiss, W.J. Stirling and S.D. Ellis, *A new Monte Carlo treatment of multiparticle phase space at high-energies*, *Comput. Phys. Commun.* **40** (1986) 359.
- [64] P.D. Draggotis, A. van Hameren and R. Kleiss, *SARGE: an algorithm for generating QCD antennas*, *Phys. Lett. B* **483** (2000) 124 [[hep-ph/0004047](#)].
- [65] 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].
- [66] C.W. Bauer, S. Fleming and M.E. Luke, *Summing Sudakov logarithms in $B \rightarrow X_s \gamma$ in effective field theory*, *Phys. Rev. D* **63** (2000) 014006 [[hep-ph/0005275](#)].
- [67] 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](#)].
- [68] C.W. Bauer and I.W. Stewart, *Invariant operators in collinear effective theory*, *Phys. Lett. B* **516** (2001) 134 [[hep-ph/0107001](#)].
- [69] 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](#)].

- [70] C.W. Bauer and M.D. Schwartz, *Improving jet distributions with effective field theory*, *Phys. Rev. Lett.* **97** (2006) 142001 [[hep-ph/0604065](#)].
- [71] C.W. Bauer and M.D. Schwartz, *Event generation from effective field theory*, *Phys. Rev. D* **76** (2007) 074004 [[hep-ph/0607296](#)].
- [72] M.D. Schwartz, *Resummation and NLO matching of event shapes with effective field theory*, *Phys. Rev. D* **77** (2008) 014026 [[arXiv:0709.2709](#)].
- [73] A.H. Mueller, *On the multiplicity of hadrons in QCD jets*, *Phys. Lett.* **B 104** (1981) 161.
- [74] G. Marchesini and B.R. Webber, *Simulation of QCD jets including soft gluon interference*, *Nucl. Phys.* **B 238** (1984) 1.
- [75] G. Gustafson, *Dual description of a confined color field*, *Phys. Lett.* **B 175** (1986) 453.
- [76] G. Gustafson and U. Pettersson, *Dipole formulation of QCD cascades*, *Nucl. Phys.* **B 306** (1988) 746.
- [77] F. Caravaglios and M. Moretti, *An algorithm to compute Born scattering amplitudes without Feynman graphs*, *Phys. Lett.* **B 358** (1995) 332 [[hep-ph/9507237](#)].
- [78] F. Caravaglios, M.L. Mangano, M. Moretti and R. Pittau, *A new approach to multi-jet calculations in hadron collisions*, *Nucl. Phys.* **B 539** (1999) 215 [[hep-ph/9807570](#)].
- [79] T. Stelzer and W.F. Long, *Automatic generation of tree level helicity amplitudes*, *Comput. Phys. Commun.* **81** (1994) 357 [[hep-ph/9401258](#)].
- [80] H. Murayama, I. Watanabe and K. Hagiwara, *HELAS: HELicity Amplitude Subroutines for Feynman diagram evaluations*, KEK-91-11.
- [81] M. Moretti, T. Ohl and J. Reuter, *O'Mega: an optimizing matrix element generator*, [hep-ph/0102195](#).
- [82] Z. Bern, L.J. Dixon and D.A. Kosower, *On-shell methods in perturbative QCD*, *Ann. Phys. (NY)* **322** (2007) 1587 [[arXiv:0704.2798](#)].