

# Event generation with exponential scaling in multiplicity using AmpliCol

Rikkert Frederix<sup>a</sup> and Timea Vitos<sup>b,c</sup>

<sup>a</sup>*Department of Physics, Lund University,  
Sölvegatan 14A, SE-223 62, Lund, Sweden*

<sup>b</sup>*Department of Physics and Astronomy, Uppsala University,  
Box 516, 751 20, Uppsala, Sweden*

<sup>c</sup>*Department of Theoretical Physics, ELTE Eötvös Loránd University,  
Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary*

*E-mail:* [rikkert.frederix@fysik.lu.se](mailto:rikkert.frederix@fysik.lu.se), [timea.vitos@physics.uu.se](mailto:timea.vitos@physics.uu.se)

**ABSTRACT:** Efficient generation of LHC events is hindered by the rapidly rising cost of evaluating QCD matrix elements with increasing multiplicity. We build on a recently proposed two-step strategy in which unweighted events are first generated using the leading-colour (LC) approximation and then reweighted to full-colour (FC) accuracy, utilising the LC integration efficiency while recovering the exact FC prediction. In this work we extend the method to general Standard Model processes and present AMPLICOL, a standalone implementation designed for LHC collisions. We benchmark multi-jet,  $t\bar{t}$ +jets,  $ZZ$ +jets, and Drell-Yan+jets production, measuring the time required to obtain a fixed number of unweighted events at FC accuracy. Across all processes, the runtime exhibits a stable exponential scaling with multiplicity, far milder than the factorial growth of conventional matrix-element generators. This demonstrates that the AMPLICOL code enables efficient event generation at multiplicities that are otherwise computationally prohibitive.

**KEYWORDS:** Automation, Specific QCD Phenomenology

**ARXIV EPRINT:** [2601.19483](https://arxiv.org/abs/2601.19483)

---

**Contents**

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The AMPLICOL approach</b>	<b>2</b>
2.1	Overview of the two-step event generation	2
2.2	Colour singlets	3
<b>3</b>	<b>Results</b>	<b>4</b>
<b>4</b>	<b>Discussion and outlook</b>	<b>6</b>
<b>A</b>	<b>Secondary unweighting efficiency</b>	<b>7</b>

---

**1 Introduction**

Accurate simulations of high-energy collisions are indispensable for both precision measurements and searches for new physics at the LHC [1]. As experimental analyses increasingly rely on multi-jet final states and complex event topologies, the demands placed on matrix-element generators continue to grow. Yet the computational cost of evaluating QCD amplitudes with many external partons remains a fundamental obstacle [2], even at leading-order (LO) accuracy: the colour algebra proliferates, the integrand becomes sharply structured, and the phase-space integration slows to a crawl. Even with decades of progress — recursive amplitude construction [3–8], importance sampling [9–13], improved phase-space mappings [14–23], stochastic sampling of discrete quantum numbers [5, 6, 14, 18, 24], utilising hardware acceleration [25–36], and machine-learning substitutes [37–50]—the scaling of the matrix-element evaluation still limits the multiplicities that can be reached in practice.

A promising strategy, explored in a recent work [22], is to decouple integration efficiency from colour accuracy. Leading-colour (LC) amplitudes capture the dominant kinematic structure of QCD matrix elements while avoiding the factorial growth of the full-colour (FC) expansion. This observation motivates a two-step workflow: generate unweighted events using the LC integrand, and subsequently restore FC accuracy by reweighting each event with the exact colour-summed matrix element. In earlier work, this approach was shown to yield substantial gains for pure-QCD processes, with LC amplitudes providing both a tractable integrand and a reweight factor that remains close to unity across the entire phase space.

In this paper we extend this two-step strategy to general Standard Model (SM) processes and present AMPLICOL, a standalone implementation of the LC→FC event-generation framework. The extension to realistic LHC processes introduces new challenges — notably the treatment of colour-singlet particles and multiple quark lines, both of which break the one-to-one correspondence between colour orderings and efficient phase-space mappings defined in ref. [22]. We address these issues through a multi-channel setup. With these ingredients, AMPLICOL can generate LC unweighted samples efficiently for a broad class of processes and subsequently reweight them to FC accuracy with high secondary unweighting efficiency.

To quantify the performance of this approach, we study four representative LHC processes — multi-jet,  $t\bar{t}$  + jets,  $ZZ$  + jets, and Drell-Yan+jets production — across increasing jet multiplicities. For each process we measure the wall-time required to obtain  $10^5$  unweighted FC events on a single CPU core. Conventional matrix-element generators scale factorially with multiplicity, but as we will present later, our results show a clear exponential scaling. We also analyse the scaling of the reweighting step, which grows faster than the LC integration and is expected to become the limiting factor at very high multiplicities.

The remainder of the paper is organised as follows. Section 2 summarises the two-step methodology and describes the extensions required for processes with colour singlets. Section 3 presents timing benchmarks and scaling fits for the selected processes. We conclude in section 4 and outline future improvements and prospects for integration with existing event-generation frameworks. Appendix A provides additional diagnostics on the secondary unweighting efficiency and effective sample size.

## 2 The AMPLICOL approach

The methodology adopted in the AMPLICOL program follows closely that presented in ref. [22] and is based on a two-step approach for event generation. In the following we present a short summary of its main points. We refer the interested reader to the mentioned reference for further details of the method.

### 2.1 Overview of the two-step event generation

The main hurdle in the computation of tree-level cross sections is the highly complex and peaked matrix element  $|\mathcal{M}^2|$  in the integrand,

$$\sigma_{\text{FC}} \simeq \int |\mathcal{M}|^2 d\Phi = \int \sum_{i,j} \mathcal{A}_i C_{ij} \mathcal{A}_j^* d\Phi, \tag{2.1}$$

where we have suppressed the parton density functions, flux factor and additional factors coming from colour and helicity averages. The integral can be rewritten as a double sum over colour-ordered amplitudes  $\mathcal{A}_i$  and corresponding colour factors  $C_{ij}$  [51–54], as denoted by the second equality sign, in which the  $i, j$  indices label colour orderings of the external QCD particles. The only terms contributing at LC accuracy in this colour sum are those with identical colour order,  $i = j$  and so the LC-truncated cross section becomes

$$\sigma_{\text{LC}} \simeq \int \sum_i \mathcal{A}_i C_{ii} \mathcal{A}_i^* d\Phi = \sum_i C_{ii} \int |\mathcal{A}_i|^2 d\Phi. \tag{2.2}$$

The two-step event generation proposed in ref. [22] starts from generating unweighted events using this LC approximation in the first step, treating each (unique) element in the sum as a separate integration channel. In the second step, these LC events are passed through a reweighting algorithm, in which each event weight is multiplied with the reweight factor

$$r^{\text{LC} \rightarrow \text{FC}} = \frac{|\mathcal{M}|^2}{\sum_i C_{ii} |\mathcal{A}_i|^2}, \tag{2.3}$$

which is the correction factor for obtaining the FC weight. This correction factor is evaluated using the specific kinematics and helicity of each event, in this way obtaining a FC-accurate

(weighted) event sample. Finally, a secondary unweighting is performed in order to obtain equal-weight events.

The main result of ref. [22] is that this two-step procedure results in efficient event generation for all-QCD (i.e., multi-jet) processes, because the single elements in the sum of eq. (2.2) are simple enough that the phase-space parametrisation introduced in ref. [55] captures the structure of the integrand well if the building blocks are ordered according to the colour ordering of  $|\mathcal{A}_i|^2$ . This results in competitive primary unweighting efficiencies in the LC event generation. Moreover, the reweight factor defined in eq. (2.3) is close to a constant resulting in secondary unweighting efficiencies in the LC $\rightarrow$ FC reweighting typically well above 50%.

For this work, we have extended the AMPLICOL code developed for ref. [22] to allow for any tree-level process generation in the SM, and focus on presenting results for the runtime to generate  $10^5$  unweighted events (at FC accuracy) for standard-candle SM processes relevant to the LHC.

## 2.2 Colour singlets

The main obstacle in the extension of the AMPLICOL program from QCD-only processes to the complete SM is the treatment of colour singlets in the generation of the LC events, i.e., the first step of the two-step procedure. The reason for this difficulty is that the order of the  $2 \rightarrow 3$  phase-space blocks introduced in ref. [55] should be in a one-to-one correspondence with a single colour ordering of the LC amplitudes. In that way, the elements of the sum of eq. (2.2) can be treated as completely independent channels for which the  $2 \rightarrow 3$  phase-space blocks form an efficient phase-space parametrisation, as shown in ref. [22]. Colour singlets do not directly fit into this picture: a LC approximation does not single out a unique “ordering” for colour singlets when there are multiple quark lines and/or multiple colour singlets. Hence, a phase-space parametrisation should be used that takes all the possible orderings for the colour singlets into account. For this we adopt a multi-channel approach with weight optimisation, following ref. [12], among all the channels that differ only by the unique positions of the colour singlets in the phase-space ordering.

Similarly to the colour singlets, for processes with more than two quark lines there is also no longer a unique mapping between the colour ordering and the ordering used in the phase-space generation. This is due to the fact that interchanging two quark lines is no longer equivalent to cyclically permuting the colour ordering beyond two-quark-line processes. While we also have implemented a multi-channel procedure to deal with three-quark-line processes, in practice our implementation is not very efficient: including the three-quark line processes seriously deteriorates the results presented below. In the light that contributions with three or more quark lines are subdominant in the computation of cross sections, we refrain from including them, and leave optimising the AMPLICOL code regarding these type of processes for future work.

### 3 Results

We use the above method for the event generation of four benchmark LHC processes at center-of-mass energy of  $\sqrt{s} = 14$  TeV:

$$\begin{aligned}
 pp &\rightarrow nj, \\
 pp &\rightarrow t\bar{t} + (n-2)j, \\
 pp &\rightarrow ZZ + (n-2)j, \\
 pp &\rightarrow e^+e^- + (n-1)j,
 \end{aligned}
 \tag{3.1}$$

with the multiplicity number  $n$  varied in the range  $n \in [2, 7]$ . In all processes we use a 5-flavour scheme and set the following cuts on the jets, whenever applicable:

$$p_T(j) > 30 \text{ GeV}, \quad \eta(j) < 6.0, \quad \Delta R(j_1, j_2) > 0.4.
 \tag{3.2}$$

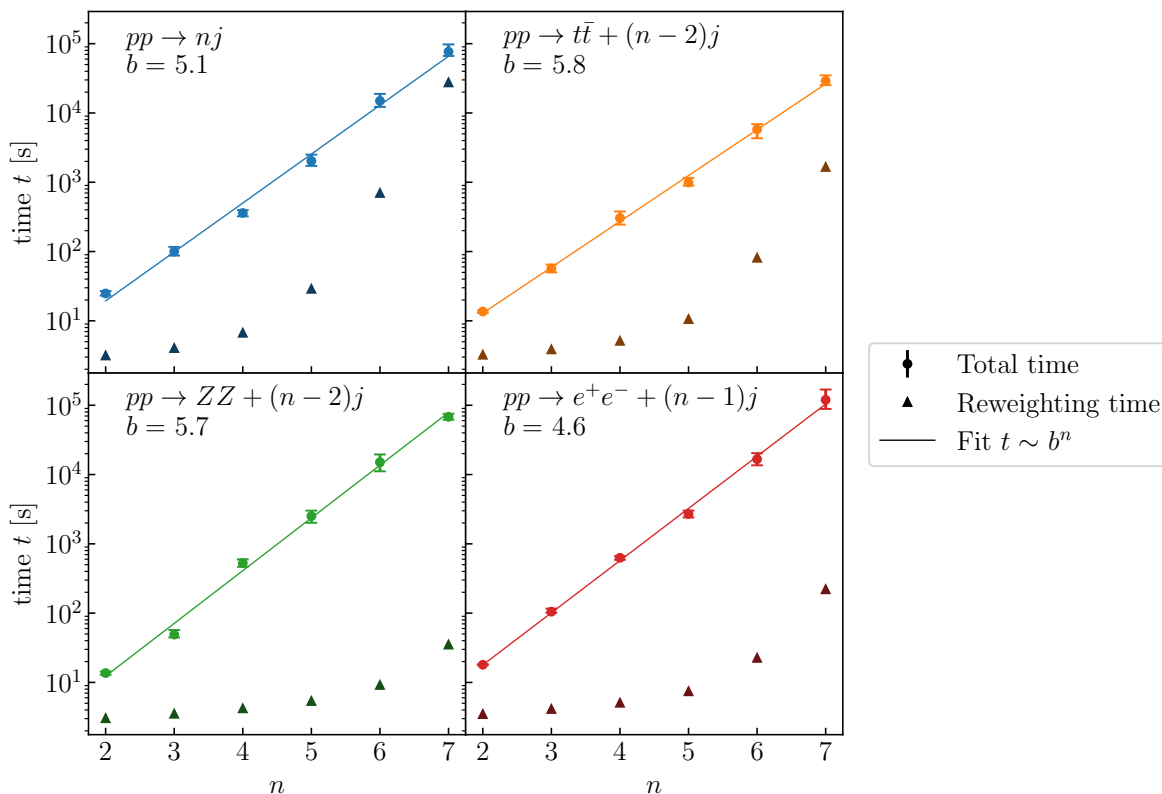
In the Drell-Yan processes, a cut on the leptonic invariant mass of  $m(e^+e^-) > 50$  GeV is used in order to avoid the photon-mediated singularity, and no further cuts on the leptons are placed. We set the renormalisation and factorisation scales equal to the  $Z$ -boson mass, and use the NNPDF23nlo set [56] for the initial-state parton distributions. Wherever applicable, results for single phase-space points and integrated cross sections are cross-checked against the matrix-element generator MadGraph5\_aMC@NLO [19]. We perform all the computations on a single core of an Intel i7-8700K CPU running at 3.70GHz and extract the evaluation time, focusing on how the runtime scales with the multiplicity number  $n$ .

The total time  $t_{\text{tot}}$  for the generation of unweighted events at FC accuracy in the two-step approach consists of

$$t_{\text{tot}} = t_{\text{int}} + t_{\text{rwgt}}.
 \tag{3.3}$$

Here, the integration time  $t_{\text{int}}$  is for the generation of unweighted LC-accurate events. This includes the setup of the integration grids, finding the upper bound of the event weights, and the primary unweighting of the generated events. At most a fraction of 0.1% of the cross section, the so-called overweight fraction (see e.g. ref. [48]), is truncated to improve the unweighting efficiency, but typically this number is between 0.05 to 0.07% for the results presented in this work. The time for reweighting plus secondary unweighting  $t_{\text{rwgt}}$  is required to adjust the weights such that they capture the FC corrections. One must take into consideration the loss of events in the secondary unweighting step and generate the LC event number accordingly, such that the final number of FC events is the required event sample. We consider this fractional loss in the timings of the total time. In the following we compare these various times for the different processes.

In figure 1 we show the main results of this work: the time to generate  $N = 10^5$  unweighted events at FC accuracy on a single CPU core. It is shown for the four benchmark processes for varying jet multiplicity. The plot includes the total time  $t_{\text{tot}}$  required to generate these events (circles with error bars) and also separately the reweighting time  $t_{\text{rwgt}}$  for each of the processes (triangles). The error bars on the total time are such that they cover the envelope of the total times of 10 separate runs (with different random seeds); the central value is the



**Figure 1.** Computational time (in seconds) for the four benchmark processes. Shown is the total computation time (circle with error bar) for  $10^5$  unweighted events at FC accuracy (including the generation time of the LC events, the reweighting time, and accounting for the secondary unweighting loss), and separately the reweighting time (triangle). An exponential curve is fitted to the total timings and portrayed as a line in the plots, with the base  $b$  of the exponential indicated in each of the four sub-plots.

average. In addition, a fit of an exponential curve is made to the total time with respect to the multiplicity  $n$ . The fit indicates an exponential growth for all processes with a base of around 5 (with a slight variance between process types, each shown in the figure).

It is remarkable that an exponential fit covers the total timing results so well. Given that the contribution to the timing from the reweighting is always subdominant — even for  $pp \rightarrow 7j$  it is only about a third of the total time, we can conclude that the timing is dominated by the generation of events at LC accuracy. At LC the complexity of the matrix elements scales polynomially ( $\sim n^3$ ) within a given integration channel and for a single helicity configuration, which means that it scales like  $\sim 2^n$  due to the fact that AMPLICOL performs an explicit sum over the helicity configurations for each phase-space point. The remaining factor  $2.5^n$ - $3^n$  comes from the phase-space integration and unweighting: we have explicitly checked that in the AMPLICOL code the matrix elements need to be evaluated about 2.5–3 times more often in order to generate the same number of LC unweighted events when one increases the multiplicity by one. There is also a small increase in the runtime with multiplicity due to the decrease of the secondary unweighting efficiency, i.e. in the LC $\rightarrow$ FC

reweighting, but this effect is subdominant because it stays approximately within the 60–80% range as long as there are at least 4 coloured particles in the process, see appendix A.

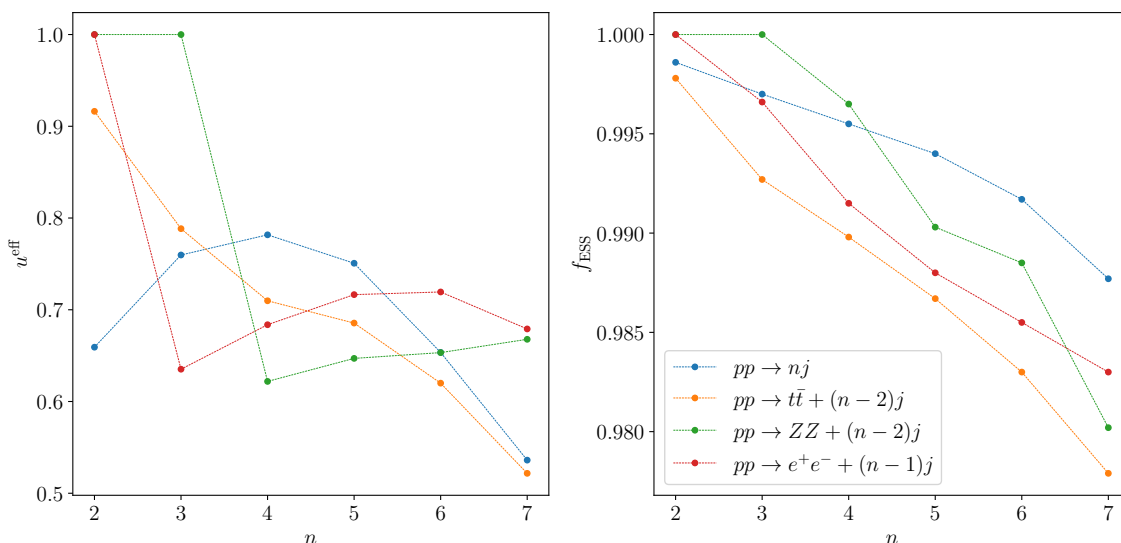
In the current implementation the reweight time  $t_{\text{rwt}}$  scales faster-than-exponentially with particle multiplicity, as can be clearly seen from the triangles in figure 1. Therefore, for large  $n$ , the time needed for reweighting will become the dominant part of the runtime. From the results in the figure, we expect this to happen when the number of coloured particles in the process (initial plus final state) is equal to ten. If matrix element calculations are needed for such large multiplicities, improvements to the reweight part of AMPLICOL code would be very beneficial. One possibility would be to reweight to next-to-leading colour only [22, 57]; but also ideas along the line of machine learning the reweight factors [58] and using other colour bases to prevent the beyond-exponential growth of FC matrix elements [59] are being investigated.

## 4 Discussion and outlook

In this work we have presented the program AMPLICOL, which performs the two-step event generation of high-multiplicity tree-level processes, with an exponential increase in computation time up for the multiplicities considered rather than the factorial growth with conventional event generators. We illustrated the computation time for four benchmark processes with various multiplicities. We have shown that with the AMPLICOL code it is possible to generate leading order events for  $2 \rightarrow n$  processes for  $n$  up to 7 (for  $pp \rightarrow \text{jets}$ ,  $pp \rightarrow t\bar{t} + \text{jets}$  and  $pp \rightarrow ZZ + \text{jets}$ ), or 8 (for  $pp \rightarrow e^+e^- + \text{jets}$ ) on a single CPU core with reasonable speed — for the most complicated of these it takes 1–2 days to generate  $10^5$  unweighted events. In the near future, the program will be interfaced with the MADGRAPH\_AMC@NLO user-friendly matrix-element generator [19], improving the automated computation of multi-jet processes, and uncovering multiplicities which are currently out of reach within that framework. Moreover, the AMPLICOL code has a relatively simple structure, and the adaptation to run on multiple cores, or a GPU, should be relatively straight-forward, potentially improving the wall-time for event generation significantly.

For all the processes presented in this work, the flavour combinations that include three (or more) quark lines have not been included. The reason is not that the code is not capable of including these sub-processes, but rather that the LC event generation efficiency deteriorates significantly when these are included, resulting in much longer runtimes. It is expected that these contributions are more complicated than sub-processes with at most two quark lines: due to various possible orderings of the quark lines at LC, they are no longer in one-to-one correspondence (up to cyclic permutations) with a single, unique ordering used for phase-space generation. We have tried including a simple multi-channel approach, similar to what works for colour singlets where the same problem arises, but without sufficient success. We leave improving the efficiency for processes with multiple quark lines for future work.

For multiplicities beyond the ones presented in this work, the reweight time to improve the accuracy from LC to FC, will become the dominant part of the runtime. In that case using techniques beyond what is currently used in AMPLICOL will be beneficial. Possibilities that are already being investigated are using machine learning to approximate the LC→FC reweight factors [58], and using multiplet bases to tame the factorial growth in the computation of the FC matrix elements [59].



**Figure 2.** The unweighting efficiency  $u^{\text{eff}}$  for increasing jet multiplicity (left) and effective sample size ratio  $f_{\text{ESS}}$  (right) for the four benchmark processes considered.

Finally, we expect that efficiency gains observed when using machine learning to improve phase-space sampling, for example within the MADNIS framework [45–47], can also be adopted in the AMPLICOL code.

## Acknowledgments

The work of R.F. is supported by the Swedish Research Council under project number 202004423. The work of T.V. is supported by the Swedish Research Council under project number VR:2023-00221.

## A Secondary unweighting efficiency

In our secondary unweighting procedure, i.e. in the reweighting of the LC events to FC accuracy, we follow the usual procedure of first finding the maximum weight among a sample of events  $w_{\text{max}}$  and then sampling from the pool of events with the acceptance probability

$$u_i^{\text{eff}} = \frac{w_i}{w_{\text{max}}}, \tag{A.1}$$

where  $w_i$  is proportional to the reweight factor,  $r^{\text{LC} \rightarrow \text{FC}}$ . The fraction of events retained in this procedure is the secondary unweighting efficiency.<sup>1</sup> In figure 2 (left) we plot the secondary unweighting efficiency for the processes considered. This efficiency is approximately 60–80% for all processes, indicating on average a sample size of only 1.3–1.7 times larger at LC accuracy to generate the required number of FC events.

<sup>1</sup>Contrary to the primary unweighting, we allow for no overweight fraction since we are dealing with relatively small samples of weighted events.

Even though the secondary unweighting efficiency is large, one may consider keeping the weighted events instead. A way of assessing the statistical power of a weighted sample of size  $N$  with weights  $w_i$  is by evaluating the Kish effective sample size (ESS) ratio [60], defined by

$$f_{\text{ESS}} = \frac{1}{N} \frac{\left(\sum_{i=1}^N w_i\right)^2}{\sum_{i=1}^N w_i^2}. \quad (\text{A.2})$$

This measure evaluates the spread of the weights, reaching a value of 1 for a sample of unweighted events (equal weights). A few outliers from an otherwise almost equal-weight event sample do not impact this effective size greatly, while one outlier will greatly decrease the unweighting efficiency defined in eq. (A.1). We show the ESS ratios for the four benchmark processes for varying multiplicity  $n$  in the right-hand plot of figure 2. Even for the highest multiplicities at  $n = 7$ , all of these processes remain at an ESS ratio of beyond 97.5%, with the multi-jet processes scaling the best with the multiplicity increase. This suggests that keeping the weighted events and thereby reducing the effective sample size by less than 2.5% significantly outperforms the unweighting of the events, except in cases where post-processing of the events (parton shower, hadronisation, detector simulation, etc.) is particularly time consuming.

**Data Availability Statement.** This article has no associated data or the data will not be deposited.

**Code Availability Statement.** This article has no associated code or the code will not be deposited.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License (CC-BY4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] J.M. Campbell et al., *Event generators for high-energy physics experiments*, *SciPost Phys.* **16** (2024) 130 [[arXiv:2203.11110](#)] [[INSPIRE](#)].
- [2] HEP SOFTWARE FOUNDATION collaboration, *A Roadmap for HEP Software and Computing R&D for the 2020s*, *Comput. Softw. Big Sci.* **3** (2019) 7 [[arXiv:1712.06982](#)] [[INSPIRE](#)].
- [3] F.A. Berends and W.T. Giele, *Recursive Calculations for Processes with  $n$  Gluons*, *Nucl. Phys. B* **306** (1988) 759 [[INSPIRE](#)].
- [4] F. Caravaglios and M. Moretti, *An algorithm to compute Born scattering amplitudes without Feynman graphs*, *Phys. Lett. B* **358** (1995) 332 [[hep-ph/9507237](#)] [[INSPIRE](#)].
- [5] F. Caravaglios, M.L. Mangano, M. Moretti and R. Pittau, *A new approach to multijet calculations in hadron collisions*, *Nucl. Phys. B* **539** (1999) 215 [[hep-ph/9807570](#)] [[INSPIRE](#)].
- [6] P. Draggiotis, R.H.P. Kleiss and C.G. Papadopoulos, *On the computation of multigluon amplitudes*, *Phys. Lett. B* **439** (1998) 157 [[hep-ph/9807207](#)] [[INSPIRE](#)].
- [7] R. Britto, F. Cachazo and B. Feng, *New recursion relations for tree amplitudes of gluons*, *Nucl. Phys. B* **715** (2005) 499 [[hep-th/0412308](#)] [[INSPIRE](#)].

- [8] C. Duhr, S. Hoeche and F. Maltoni, *Color-dressed recursive relations for multi-parton amplitudes*, *JHEP* **08** (2006) 062 [[hep-ph/0607057](#)] [[INSPIRE](#)].
- [9] G.P. Lepage, *A New Algorithm for Adaptive Multidimensional Integration*, *J. Comput. Phys.* **27** (1978) 192 [[INSPIRE](#)].
- [10] G.P. Lepage, *Adaptive multidimensional integration: VEGAS enhanced*, *J. Comput. Phys.* **439** (2021) 110386 [[arXiv:2009.05112](#)] [[INSPIRE](#)].
- [11] A. van Hameren, *PARNI for importance sampling and density estimation*, *Acta Phys. Polon. B* **40** (2009) 259 [[arXiv:0710.2448](#)] [[INSPIRE](#)].
- [12] R. Kleiss and R. Pittau, *Weight optimization in multichannel Monte Carlo*, *Comput. Phys. Commun.* **83** (1994) 141 [[hep-ph/9405257](#)] [[INSPIRE](#)].
- [13] T. Ohl, *Vegas revisited: Adaptive Monte Carlo integration beyond factorization*, *Comput. Phys. Commun.* **120** (1999) 13 [[hep-ph/9806432](#)] [[INSPIRE](#)].
- [14] C.G. Papadopoulos, *PHEGAS: A phase space generator for automatic cross-section computation*, *Comput. Phys. Commun.* **137** (2001) 247 [[hep-ph/0007335](#)] [[INSPIRE](#)].
- [15] F. Krauss, R. Kuhn and G. Soff, *AMEGIC++ 1.0: A matrix element generator in C++*, *JHEP* **02** (2002) 044 [[hep-ph/0109036](#)] [[INSPIRE](#)].
- [16] F. Maltoni and T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, *JHEP* **02** (2003) 027 [[hep-ph/0208156](#)] [[INSPIRE](#)].
- [17] W. Kilian, T. Ohl and J. Reuter, *WHIZARD: Simulating Multi-Particle Processes at LHC and ILC*, *Eur. Phys. J. C* **71** (2011) 1742 [[arXiv:0708.4233](#)] [[INSPIRE](#)].
- [18] T. Gleisberg and S. Hoeche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039 [[arXiv:0808.3674](#)] [[INSPIRE](#)].
- [19] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [[arXiv:1405.0301](#)] [[INSPIRE](#)].
- [20] O. Mattelaer and K. Ostrolenk, *Speeding up MadGraph5\_aMC@NLO*, *Eur. Phys. J. C* **81** (2021) 435 [[arXiv:2102.00773](#)] [[INSPIRE](#)].
- [21] E. Bothmann et al., *Efficient phase-space generation for hadron collider event simulation*, *SciPost Phys.* **15** (2023) 169 [[arXiv:2302.10449](#)] [[INSPIRE](#)].
- [22] R. Frederix and T. Vitos, *Leading-colour-based unweighted event generation for multi-parton tree-level processes*, *JHEP* **12** (2024) 201 [[arXiv:2409.12128](#)] [[INSPIRE](#)].
- [23] SHERPA collaboration, *Event generation with Sherpa 3*, *JHEP* **12** (2024) 156 [[arXiv:2410.22148](#)] [[INSPIRE](#)].
- [24] M.L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003) 001 [[hep-ph/0206293](#)] [[INSPIRE](#)].
- [25] J. Kanzaki, *Monte Carlo integration on GPU*, *Eur. Phys. J. C* **71** (2011) 1559 [[arXiv:1010.2107](#)] [[INSPIRE](#)].
- [26] K. Hagiwara et al., *Fast calculation of HELAS amplitudes using graphics processing unit (GPU)*, *Eur. Phys. J. C* **66** (2010) 477 [[arXiv:0908.4403](#)] [[INSPIRE](#)].
- [27] K. Hagiwara et al., *Fast computation of MadGraph amplitudes on graphics processing unit (GPU)*, *Eur. Phys. J. C* **73** (2013) 2608 [[arXiv:1305.0708](#)] [[INSPIRE](#)].

- [28] E. Bothmann et al., *Many-gluon tree amplitudes on modern GPUs: A case study for novel event generators*, *SciPost Phys. Codeb.* **2022** (2022) 3 [[arXiv:2106.06507](#)] [[INSPIRE](#)].
- [29] S. Carrazza, J. Cruz-Martinez, M. Rossi and M. Zaro, *MadFlow: automating Monte Carlo simulation on GPU for particle physics processes*, *Eur. Phys. J. C* **81** (2021) 656 [[arXiv:2106.10279](#)] [[INSPIRE](#)].
- [30] E. Bothmann et al., *A portable parton-level event generator for the high-luminosity LHC*, *SciPost Phys.* **17** (2024) 081 [[arXiv:2311.06198](#)] [[INSPIRE](#)].
- [31] J.M. Cruz-Martinez, G. De Laurentis and M. Pellen, *Accelerating Berends-Giele recursion for gluons in arbitrary dimensions over finite fields*, *Eur. Phys. J. C* **85** (2025) 590 [[arXiv:2502.07060](#)] [[INSPIRE](#)].
- [32] A. Valassi, S. Roiser, O. Mattelaer and S. Hageboeck, *Design and engineering of a simplified workflow execution for the MG5aMC event generator on GPUs and vector CPUs*, *EPJ Web Conf.* **251** (2021) 03045 [[arXiv:2106.12631](#)] [[INSPIRE](#)].
- [33] A. Valassi et al., *Developments in Performance and Portability for MadGraph5\_aMC@NLO*, *PoS ICHEP2022* (2022) 212 [[arXiv:2210.11122](#)] [[INSPIRE](#)].
- [34] A. Valassi et al., *Speeding up Madgraph5 aMC@NLO through CPU vectorization and GPU offloading: towards a first alpha release*, in the proceedings of the *21th International Workshop on Advanced Computing and Analysis Techniques in Physics Research: AI meets Reality*, Bari, Italy, October 24–28 (2022) [[arXiv:2303.18244](#)] [[INSPIRE](#)].
- [35] S. Hageboeck et al., *Madgraph5\_aMC@NLO on GPUs and vector CPUs Experience with the first alpha release*, *EPJ Web Conf.* **295** (2024) 11013 [[arXiv:2312.02898](#)] [[INSPIRE](#)].
- [36] A. Valassi et al., *Madgraph on GPUs and vector CPUs: Towards production. The 5-year journey to the first LO release CUDACPP v1.00.00*, *EPJ Web Conf.* **337** (2025) 01021 [[arXiv:2503.21935](#)] [[INSPIRE](#)].
- [37] J. Bendavid, *Efficient Monte Carlo Integration Using Boosted Decision Trees and Generative Deep Neural Networks*, [arXiv:1707.00028](#) [[INSPIRE](#)].
- [38] M.D. Klimek and M. Perelstein, *Neural Network-Based Approach to Phase Space Integration*, *SciPost Phys.* **9** (2020) 053 [[arXiv:1810.11509](#)] [[INSPIRE](#)].
- [39] I.-K. Chen, M.D. Klimek and M. Perelstein, *Improved neural network Monte Carlo simulation*, *SciPost Phys.* **10** (2021) 023 [[arXiv:2009.07819](#)] [[INSPIRE](#)].
- [40] C. Gao, J. Isaacson and C. Krause, *i-flow: High-dimensional Integration and Sampling with Normalizing Flows*, *Mach. Learn. Sci. Tech.* **1** (2020) 045023 [[arXiv:2001.05486](#)] [[INSPIRE](#)].
- [41] E. Bothmann et al., *Exploring phase space with Neural Importance Sampling*, *SciPost Phys.* **8** (2020) 069 [[arXiv:2001.05478](#)] [[INSPIRE](#)].
- [42] C. Gao et al., *Event Generation with Normalizing Flows*, *Phys. Rev. D* **101** (2020) 076002 [[arXiv:2001.10028](#)] [[INSPIRE](#)].
- [43] R. Winterhalder et al., *Targeting multi-loop integrals with neural networks*, *SciPost Phys.* **12** (2022) 129 [[arXiv:2112.09145](#)] [[INSPIRE](#)].
- [44] S. Badger et al., *Machine learning and LHC event generation*, *SciPost Phys.* **14** (2023) 079 [[arXiv:2203.07460](#)] [[INSPIRE](#)].
- [45] T. Heimel et al., *MadNIS — Neural multi-channel importance sampling*, *SciPost Phys.* **15** (2023) 141 [[arXiv:2212.06172](#)] [[INSPIRE](#)].

- [46] T. Heimel et al., *The MadNIS reloaded*, *SciPost Phys.* **17** (2024) 023 [[arXiv:2311.01548](#)] [[INSPIRE](#)].
- [47] T. Heimel, O. Mattelaer, T. Plehn and R. Winterhalder, *Differentiable MadNIS-Lite*, *SciPost Phys.* **18** (2025) 017 [[arXiv:2408.01486](#)] [[INSPIRE](#)].
- [48] E. Bothmann et al., *Efficient many-jet event generation with Flow Matching*, [arXiv:2506.18987](#) [[INSPIRE](#)].
- [49] T. Janßen, R. Poncelet and S. Schumann, *Sampling NNLO QCD phase space with normalizing flows*, *JHEP* **09** (2025) 194 [[arXiv:2505.13608](#)] [[INSPIRE](#)].
- [50] L. Beccatini, F. Maltoni, O. Mattelaer and R. Winterhalder, *Amplitude Surrogates for Multi-Jet Processes*, [arXiv:2512.11036](#) [[INSPIRE](#)].
- [51] J.E. Paton and H.-M. Chan, *Generalized veneziano model with isospin*, *Nucl. Phys. B* **10** (1969) 516 [[INSPIRE](#)].
- [52] M.L. Mangano, S.J. Parke and Z. Xu, *Duality and Multi-Gluon Scattering*, *Nucl. Phys. B* **298** (1988) 653 [[INSPIRE](#)].
- [53] V. Del Duca, L.J. Dixon and F. Maltoni, *New color decompositions for gauge amplitudes at tree and loop level*, *Nucl. Phys. B* **571** (2000) 51 [[hep-ph/9910563](#)] [[INSPIRE](#)].
- [54] F. Maltoni, K. Paul, T. Stelzer and S. Willenbrock, *Color Flow Decomposition of QCD Amplitudes*, *Phys. Rev. D* **67** (2003) 014026 [[hep-ph/0209271](#)] [[INSPIRE](#)].
- [55] E. Byckling and K. Kajantie, *Reductions of the phase-space integral in terms of simpler processes*, *Phys. Rev.* **187** (1969) 2008 [[INSPIRE](#)].
- [56] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244 [[arXiv:1207.1303](#)] [[INSPIRE](#)].
- [57] R. Frederix and T. Vitos, *The colour matrix at next-to-leading-colour accuracy for tree-level multi-parton processes*, *JHEP* **12** (2021) 157 [[arXiv:2109.10377](#)] [[INSPIRE](#)].
- [58] J.M. Villadamigo et al., *FASTColor — Full-color Amplitude Surrogate Toolkit for QCD*, [arXiv:2509.07068](#) [[INSPIRE](#)].
- [59] O. Bolinder, R. Frederix and M. Sjodahl, *All-gluon amplitudes with off-shell recursion in multiplet bases*, *JHEP* **10** (2025) 168 [[arXiv:2507.22636](#)] [[INSPIRE](#)].
- [60] L. Kish, *Survey Sampling*, John Wiley & Sons, Inc., New York, U.S.A. (1965).