



Unbiased elimination of negative weights in Monte Carlo samples

Jeppe R. Andersen¹, Andreas Maier^{2,a}

¹ Department of Physics, Institute for Particle Physics Phenomenology, University of Durham, Durham DH1 3LE, UK

² Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany

Received: 28 September 2021 / Accepted: 27 April 2022 / Published online: 12 May 2022
© The Author(s) 2022

Abstract We propose a novel method for the elimination of negative Monte Carlo event weights. The method is process-agnostic, independent of any analysis, and preserves all physical observables. We demonstrate the overall performance and systematic improvement with increasing event sample size, based on predictions for the production of a W boson with two jets calculated at next-to-leading order perturbation theory.

1 Introduction

The LHC physics programme has entered an era of precision measurements. Such measurements demand equally precise theory predictions. To match this requirement, it is most often necessary to include at least next-to-leading order (NLO) perturbative corrections. For a number of processes, even higher fixed-order corrections, i.e. NNLO and NNNLO, have to be taken into account. In order to consistently combine virtual corrections, usually computed in $d = 4 - 2\epsilon$ dimensions, and real corrections, obtained via numerical integration over four-dimensional momenta, one introduces subtraction terms which render each component finite, see for example [1,2].

Since one is usually interested in distributions, additional effects from parton showers and hadronisation have to be accounted for. Methods such as MC@NLO [3] and POWHEG [4] can be employed for matching parton showers and NLO fixed-order predictions. Depending on the phase-space region of interest, further resummation may be required. Examples include transverse momentum [5] and high-energy resummation [6,7]. High-accuracy predictions for multi-jet observables are then obtained by merging exclusive event samples using e.g. the MEPS@NLO [8] or UNLOPS [9,10] approaches. Finally, the detector response to the generated events is simulated [11].

Thanks to continuous improvements, modern general purpose event generators [12–15] render the inclusion of many of these corrections feasible for a large range of scattering processes. However, the improved accuracy comes at a significant cost in computing time. What is more, each of the individual calculational steps outlined above also increases the cost of all subsequent steps. The reason for this is that typically a number of negative-weight counterevents are generated, for instance for unitarisation or in order to prevent double counting. This implies that a much larger number of events has to be processed to reach the same statistical significance as for the case of purely positive-weight event samples. This problem is particularly pronounced in the final detector simulation step, which can take hours of CPU time for each event.

In general, the number of events that have to be forwarded to the detector simulation can be reduced by unweighting the event sample. Taking W as the largest absolute event weight in the sample, an event with weight w_i is rejected with probability $\mathbb{P} = 1 - |w_i|/W$ and otherwise assigned the new weight $w_i \rightarrow \text{sgn}(w_i)W$. If all event weights are positive, an unweighted sample of N events has uniform weight $w_i = W$, which is ideal in the sense that the Monte Carlo uncertainty estimate for the cross section, $\langle\sigma\rangle = \sqrt{\sum_{i=1}^N w_i^2}$, is minimal among all samples with N events and the same predicted cross section $\sigma = \sum_{i=1}^N w_i$. Conversely, a given precision goal can be reached with the smallest number of events.

This ideal case of completely uniform weights is no longer achieved as soon as negative weights appear. For a fractional negative weight contribution

$$r_- = -\frac{\sum_{w_i < 0} w_i}{\sum_i |w_i|}, \quad (1)$$

the number of events required for a given precision goal increases to

$$N(r_-) = \frac{N(0)}{(1 - 2r_-)^2} \quad (2)$$

^a e-mail: andreas.martin.maier@desy.de (corresponding author)

relative to the required number $N(0)$ in the absence of negative weights, see also [16].

Consequently, there have been a number of recent efforts to reduce the fraction r_- of the negative weight contribution. One avenue that is being explored is to optimise the event generation itself [16–20]. An alternative approach is to remove negative weights from the generated samples, while taking care that observables are not affected [21–23]. The algorithm proposed in [21] has been shown to be very effective in a highly non-trivial example. However, there are two caveats. First, the algorithm requires the choice of an auxiliary distribution, despite being process-agnostic in nature. Second, while the algorithm has been shown to produce sound results in practice, a proof of the correctness has only been given for a specific set of observables related to the chosen distribution.

In the following, we present a novel method, dubbed *cell resampling*, which eliminates negative event weights locally in phase space, is independent of the underlying process, and preserves all physical observables. Following the earlier proposal [21], we insert an additional resampling step before unweighting and detector simulation into the event generation chain outlined above. This step only affects the event weights, leaving all other properties untouched. Barring exact cancellations between weights, the number of events remains the same as well. We describe our method in Sect. 2, and apply it to a calculation of W boson production with two jets at next-to-leading order in Sect. 3. While we mostly focus on the application to fixed-order QCD generation of weighted event samples, the approach is largely independent of the details of the event generation, and we comment on possible generalisations whenever we have to make specific choices. We conclude with a summary and an outlook in Sect. 4.

2 Cell resampling

Our main goal is to eliminate negative event weights without affecting predictions for observables. We only consider observables that can be both predicted faithfully by a given event generator and measured in a real-world experiment involving detectors with a finite resolution.

To measure an observable \mathcal{O} we first select a phase-space region \mathcal{D} that is large enough to be resolved in a real-world experiment, where the resolution is typically limited not only by the detector, but also by the available statistics. The measured value of the observable is then obtained by counting events within this region. Obviously, this will always yield a non-negative result.

On the theory side, we predict the value of the observable from a Monte Carlo event sample with both positive and negative weights by summing up the weights w_i of the events i contained in the selected region \mathcal{D} :

$$\mathcal{O} = \sum_{i \in \mathcal{D}} w_i. \quad (3)$$

Since this corresponds to a Monte Carlo estimate of the integrated physical cross section over the selected phase space region, the result has to be non-negative given enough statistics.

We now aim to modify the event weights in such a way that negative weights are eliminated. At the same time, predictions for any observables of the kind discussed above should be preserved. To this end, we focus on a single event with negative weight, our *seed*, and consider a small solid sphere in phase space that is centred around said event and contains no other events. We call this sphere a *cell*. We then gradually increase the radius of the cell until the sum over all weights of events contained inside the cell is non-negative.

The key point is that with sufficiently many generated events, this cell can be made arbitrarily small. In particular, it can be made so small that it cannot be resolved anymore in a real-world experiment. At this point, we may freely redistribute the weights of the events within the cell without affecting any observables. One possible choice is to replace all event weights by their absolute value and rescale them such that the sum of weights is preserved. That is, we perform the replacement

$$w_i \rightarrow \frac{\sum_{j \in \mathcal{C}} w_j}{\sum_{j \in \mathcal{C}} |w_j|} |w_i| \quad (4)$$

for all weights w_i of events i inside the cell \mathcal{C} . This accomplishes the original intent behind the generation of negative-weight counterevents: any overestimation of the cross section from positive-weight events is cancelled *locally* in phase space.

After applying the above steps successively to each negative-weight event we are left with a physically equivalent sample of all positive-weight events. While it is in principle possible to apply the procedure to unweighted events, this would needlessly inflate cell sizes and systematically discard all events inside a cell, since their weights would sum up to zero. We therefore assume weighted input events. If desired, efficient unweighting can be performed after the elimination of negative weights.

Note that the method is completely agnostic to any details of the event generation, including the underlying process, and does not refer to any properties of specific observables. The only prerequisite is sufficiently high statistics, so that single cells are not resolved by any observable that can be measured in a real-world experiment.

In practice, it may not be feasible to generate and resample enough events to reach this goal. A particular concern is posed by cells containing events in different histogram bins of some distribution. Small enough cells close to bin boundaries can even turn out to be beneficial by smoothing out statistical jitter due to bin migration. However, as cells grow

larger, the resampling will start to smear out characteristic features, such as resonance peaks.

To ensure that smearing effects remain negligible compared to other sources of uncertainty we can impose an upper limit on the cell sizes. The cost we have to pay is that some cells may not contain sufficient positive weight events to cancel the contribution from events with negative weight. This implies that we no longer eliminate all of the negative event weights. Still, since at least part of the negative seed event weight is absorbed by surrounding events, a subsequent unweighting will reduce the fraction of negative-weight events compared to the original sample. The size of cells with negative accumulated weight is related to the extent in phase space of counter-terms, and as such is related to the usual problems with bin-to-bin migration in higher-order calculations.

For the sake of a streamlined presentation, we will allow cells to grow arbitrarily large for the remainder of this section. We will come back to the discussion of a cell size limit when analysing the practical performance of our method in Sect. 3.

2.1 Distances in phase space

A central ingredient in cell resampling is the definition of a distance function in phase space. While it may be possible to achieve even better results with functions tailored to specific processes and observables, our aim here is to define a universal distance measure.

There are two main requirements. First, let us consider events which are close to each other according to our distance function. Such events will very likely be part of the same cells and redistribute weights among each other. It is therefore absolutely essential that they have similar experimental signatures or only differ in properties that are not predicted by the event generator. Otherwise it is possible that only some of these events contribute to a given observable and the total weight of this subset and therefore the prediction for the observable is changed by the resampling. Second, it is desirable that events with a large distance according to the chosen function are easily distinguishable experimentally. In short, the distance function should reflect both the experimental sensitivity and the limitations of the event generator.

One consequence of these requirements is that the distance function has to be infrared safe. Adding further soft particles to an event should only change distances by a small amount. Specifically, final-state particles with vanishing four-momentum are undetectable and should not affect distances at all. Furthermore, the distance function has to have limited sensitivity to collinear splittings as well as soft radiation. In general, this can be achieved by basing the distance function on suitable defined infrared-safe physics objects instead of the elementary particles in the final state. The precise definition of these physics objects may depend on

the details of the theory prediction. In the following, we consider only pure QCD corrections and ensure infrared safety by considering final-state jets instead of partons. It is natural to adopt the same jet clustering that was used for the event generation.

Keeping the general requirements in mind, we now define a concrete distance function for use with fixed-order Monte Carlo generators, which we use in the following. Note that our choice is by no means unique and we do not claim that our definition is optimal. A particularly promising alternative is the “energy mover’s distance” introduced in [24] and generalised to include flavour information in [25]. We leave a detailed comparison between different distance functions to future work.

2.1.1 Definition of the distance function

As a first step, we cluster the partons in each event into jets. We then group outgoing particles into sets according to their types, i.e. according to all discrete observable properties such as flavour and charge. Here, we are using the word “particle” in a loose sense, designating a jet – or, more generally, any infrared-safe physics object – as a single particle. Particles which only differ by their four-momenta (including possible differences in their invariant masses) end up inside the same set. For the time being, we do not differentiate between polarisations or various types of jets, but it is straightforward to add such distinctions.

The distance d between an event e with T particle type sets $\mathcal{S} = \{s_1, s_2, \dots, s_T\}$ and an event e' with sets $\mathcal{S}' = \{s'_1, s'_2, \dots, s'_T\}$ is then given by the sum of the distances between matching sets, i.e.

$$d(e, e') = \sum_{t=1}^T d(s_t, s'_t), \tag{5}$$

where s_t, s'_t contain all particles of type t that occur in the respective event. If there are no particles of type t in a given event, then the corresponding set is empty. This ensures that all events in the sample have the same number of particle type sets.

Next, we have to define the distance between a set s_t , containing P particles of type t with four-momenta p_1, \dots, p_P and a set s'_t with Q particles of the same type with four-momenta q_1, \dots, q_Q . Without loss of generality we can assume $Q \leq P$. To facilitate the following steps we first add $P - Q$ particles with vanishing momenta to the set Q , i.e. we define the momenta

$$q_{Q+1} = \dots = q_P = 0. \tag{6}$$

Naturally, the distance between s_t and s'_t must not depend on the labelling of the momenta. We therefore consider all permutations of the momenta q_1, \dots, q_P and sum the pairwise

distances to the momenta p_1, \dots, p_P for each permutation. The minimum defines the distance between s and s' :

$$d(s_t, s'_t) = \min_{\sigma \in S_P} \sum_{i=1}^P d(p_i, q_{\sigma(i)}). \tag{7}$$

Here, S_P denotes the symmetric group, i.e. the group of all permutations of P elements.

While this distance function is obviously invariant under relabelling and when adding vanishing momenta, its calculation quickly becomes prohibitively expensive when the number of momenta P in the sets grows large. More precisely, the computational cost scales as $\mathcal{O}(P!) = \mathcal{O}(P^{P+\frac{1}{2}})$. For events obtained through fixed-order calculations, P is typically small enough and the poor scaling is not a concern. However, after parton showering, large values of $P \approx 10$ can be reached. In such cases, we use an alternative distance function \tilde{d} which is easier to compute. We first define a unique labelling by ordering the momenta according to their norm

$$\|p_i\| \equiv d(p_i, 0), \quad \|q_i\| \equiv d(q_i, 0). \tag{8}$$

We then search for the nearest neighbour of p_1 in Q , i.e. the momentum $q_{NN(1)} \in Q$ that minimises the distance to p_1 :

$$d(p_1, q_{NN(1)}) \leq d(p_1, q_i) \quad \forall q_i \in s'_t. \tag{9}$$

In the next step, we find the nearest neighbour $q_{NN(2)}$ of p_2 , *excluding* the momentum $q_{NN(1)}$ and add $d(p_2, q_{NN(2)})$ to the total distance. We iteratively define further exclusive nearest neighbours $NN(j)$ by removing previous ones,

$$s'_{t,\bar{j}} = s'_t \setminus \{q_{NN(1)}, \dots, q_{NN(j-1)}\}, \tag{10}$$

and finding the element $q_{NN(j)}$ such that

$$d(p_j, q_{NN(j)}) \leq d(p_j, q_i) \quad \forall q_i \in s'_{t,\bar{j}}. \tag{11}$$

Finally, we add up all these distances. To arrive at a symmetric distance function, we compare to the total distance obtained by exchanging s_t and s'_t . The smaller of the two distances defines $\tilde{d}(s_t, s'_t)$:

$$\tilde{d}(s_t, s'_t) = \min \left(\sum_{i=1}^P d(p_i, q_{NN(i)}), \sum_{i=1}^P d(q_i, p_{NN(i)}) \right). \tag{12}$$

Since this minimises over only two out of all possible sets of pairings between momenta p_i and q_j , $\tilde{d}(s_t, s'_t)$ scales as $\mathcal{O}(P^2)$ instead of $\mathcal{O}(P^{P+\frac{1}{2}})$. For the same reason, it is bounded from below by $d(s_t, s'_t)$. However, for nearby events we expect that the exclusive nearest neighbours $NN(i)$ defined above coincide with the *actual* nearest neighbours. In this case the distance functions \tilde{d} and d will yield the same result. Let us finally remark that \tilde{d} also remains unchanged when adding vanishing momenta to the sets s_t and s'_t .

The last missing ingredient is the distance $d(p, q)$ between two momenta p and q . One can in principle

choose different distance functions for different particles types (e.g. to include only transverse momentum for neutrinos), but for simplicity we will in this study apply the same distance for all types. Note that the Minkowski distance $\sqrt{(p-q)_\mu (p-q)^\mu}$ is *not* a suitable distance. Such a distance would be insensitive to lightlike differences between momenta, which does not match the reality of experimental measurements. Since, for a given particle mass, only three momentum components are independent, our distance measure is based on the difference in *spatial* momentum. Anticipating the fact that experimental measurements are more sensitive to the perpendicular momentum components we further add a rescaled difference in transverse momenta:

$$d(p, q) = \sqrt{\sum_{i=1}^3 (p_i - q_i)^2 + \tau^2 (p_\perp - q_\perp)^2}. \tag{13}$$

τ is a tunable parameter.

We emphasise again that the presented distance function is purely intended for use with fixed-order QCD, and we are only concerned about sensitivity to those observables that can be predicted faithfully. More sophisticated theory predictions, like those obtained from parton showers, will require further refinements. For example, sensitivity to jet substructure could be introduced through a specialised metric for the distance between two jets in place of equation (13). We will leave such explorations to future work.

2.1.2 Alternative norms

The choice of a momentum distance function is not unique and one could alternatively consider other coordinate systems, such as light-cone coordinates, or general p-norms

$$d_p(p, q) = \left(\sum_{i=1}^3 (p_i - q_i)^p \right)^{\frac{1}{p}}. \tag{14}$$

One interesting possibility yet to be explored is to take inspiration from jet distance measures and employ a distance in rapidity or azimuthal angle, which has to be combined with a transverse momentum distance in a meaningful way. We leave such explorations to further study.

It is, in any case, crucial to not violate the triangle inequality

$$d(p, q_1 + q_2) \leq d(p, q_1) + d(p, q_2). \tag{15}$$

To illustrate this point, let us consider the *square* of the spatial norm

$$d^2(p, q) = \sum_{i=1}^3 (p_i - q_i)^2, \tag{16}$$

a set s with one particle momentum p_1 , and a set s' with one particle momentum q_1 . For simplicity, let us assume one-

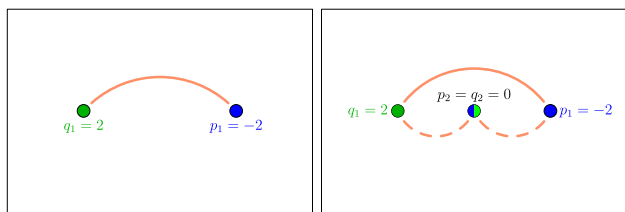


Fig. 1 Distances between two sets s, s' , each of which contains one particle (left panel) or two particles (right panel). Blue dots represent particles in set s , green dots stand for particles in set s' . The possible pairings used in the distance calculation are indicated by lines. The square norm d^2 erroneously results in a shorter distance for the pairing indicated by the dashed line

dimensional momenta with $p_1 = 2$ and $q_1 = -2$ in arbitrary units. Since there is only one possible pairing of momenta, the event distance is $d_2(s, s') = d(p_1, q_1) = 4$ using the spatial $p = 2$ norm and $d^2(s, s') = d^2(p_1, q_1) = 16$ when using the square.

We now add one particle with vanishing momentum to each s and s' , reminding ourselves that this change is not experimentally detectable and therefore must not affect the distance between the sets. This scenario is shown in Fig. 1. According to Eq. (7) the set distance is now given by

$$d(s, s') = \min[d(p_1, q_1) + d(p_2, q_2), d(p_1, q_2) + d(p_2, q_1)], \tag{17}$$

$$p_2 = q_2 = 0, \tag{18}$$

with $d(p_2, q_2) = 0$ for either choice. For the spatial norm both permutations yield the same result and $d_2(s, s') = 4$ as before. However, if we use the square d^2 instead, we find that the triangle inequality is violated and $d^2(s, s') = 8$ instead of 16.

2.1.3 Example for an event distance

To illustrate one possible distance measure, let us consider a simple yet non-trivial example, where we calculate the distance between two events e and e' . e has two outgoing photons and a jet, and e' two photons and no jets. The distance between the two events is therefore

$$d(e, e') = d(s_\gamma, s'_\gamma) + d(s_j, s'_j). \tag{19}$$

The jet set s_j for the event e contains a single jet with momentum p_j , whereas the corresponding set s'_j for e' is empty. Likewise, s_γ contains the photons in e with momenta p_1, p_2 and s'_γ the photons in e' with momenta p'_1, p'_2 . The distance between these two sets is

$$d(s_\gamma, s'_\gamma) = \min[d(p_1, p'_1) + d(p_2, p'_2), d(p_1, p'_2) + d(p_2, p'_1)], \tag{20}$$

where $d(p, q) = \sqrt{\sum_{i=1}^3 (p_i - q_i)^2 + \tau^2 (p_\perp - q_\perp)^2}$ is the distance introduced in Eq. (13).

To calculate the distance $d(s_j, s'_j)$ between the jet sets, we first add an auxiliary ‘‘jet’’ with vanishing momentum $p'_j = 0$ to s'_j , so that both sets each contain a single jet. The set distance is then

$$d(s_j, s'_j) = d(p_j, p'_j) = \sqrt{\sum_{i=1}^3 p_{ji}^2 + \tau^2 p_{j\perp}^2}, \tag{21}$$

where p_j denotes the momentum of the physical jet.

2.1.4 Distances in high-dimensional spaces

When considering a fixed number of generated events in phase spaces of increasing dimension, the typical distance between two events, and therefore the characteristic cell size, will grow very quickly. Conversely, when trying to achieve a given cell size, the required number of events will become prohibitively large very soon. We might be worried that for processes with many final-state particles cells unavoidably become so large that they can be resolved experimentally. This is not the case. The reason is that experimental resolution is limited by statistics in addition to detector limitations. In fact, the sizes of resolved regions of phase space will increase in the same way as the cell sizes with the number of dimensions.

It is instructive to consider how the final cell size in a high-dimensional phase space translates to differences in real-world observables, which are typically low-dimensional distributions. To obtain some quantitative insight, we can model a cell as an n -ball with radius R in a n -dimensional Euclidean space. For a sufficiently small radius, events within the cell will be approximately uniformly distributed. The exceptions are the cell seed at the centre and the last event added at a distance R to the cell. It is well known that for large n , events tend to be close to the surface, and the mean distance of an event to the seed is indeed given by

$$\bar{d} = \frac{1}{V_n} \int dV_n r = \frac{n}{n+1} R, \tag{22}$$

where V_n is the volume of the n -ball. However, when predicting one-dimensional distributions, we project the events onto a line, which we can take as the n -th coordinate axis. For the mean distance in this direction we obtain

$$\bar{d}_n = \frac{1}{V_n} \int dV_n r_n = \frac{\Gamma(\frac{n}{2} + 1)}{\Gamma(\frac{1}{2})\Gamma(\frac{n+3}{2})} R = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \tag{23}$$

The difference in one-dimensional distributions is much smaller than the cell size, especially in high-dimensional phase spaces.

2.2 Nearest-neighbour search and locality-sensitive hashing

A very appealing feature of cell resampling is that the cell radii automatically become smaller with increasing number of events. Anticipating the discussion of the practical application with the distance measure in Sect. 3, we indeed find a significant reduction in cell size, as shown in Fig. 2. Due to large fluctuations in the weights of the cell seeds, and correspondingly in the cell sizes, we consider the median of the cell radii instead of the arithmetic mean.

From this point of view, cell resampling should be applied to event samples that are as large as possible. However, the main motivation behind the resampling is to reduce the computational cost of later steps. This implies that the CPU time needed for resampling should be much smaller. On the one side, various event generation steps are usually performed on an event-by-event basis, and the asymptotic generation time scales linearly with the number n of events. On the other side, the number of negative-weight events and therefore the number of cells are also proportional to n . To achieve the same asymptotic $\mathcal{O}(n)$ runtime scaling, the time to construct a single cell has to be constant, i.e. independent of the sample size. However, while obviously asymptotically inferior, overall $\mathcal{O}(n \log^k n)$ scaling for some integer k will presumably be sufficient for real-world sample sizes.

To reiterate, a cell is constructed as follows. We choose a seed, i.e. one of the remaining negative-weight events. As long as the sum of weights inside the cell is negative, we subsequently find the closest event to the seed outside the cell according to our distance function and add it to the cell. Finally, we reweight all events inside the cell.

The selection of a negative-weight event can be achieved in constant time with the help of a pregenerated array containing these events, which can be created in $\mathcal{O}(n)$ steps. If we want to construct cells in a specific order, an additional

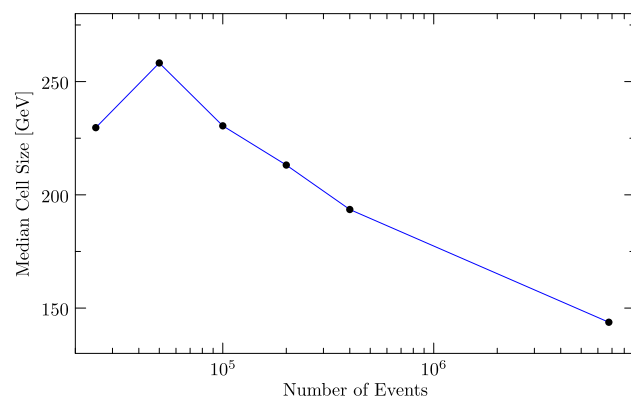


Fig. 2 Median cell radius as a function of sample size according to the distance measure defined in Sect. 2.1.1 with $\tau = 0$. See Sect. 3 for details on the event generation

$\mathcal{O}(n \log n)$ sorting step is required. The impact of different seed selection strategies is discussed in appendix A.

The biggest challenge is to find nearby events in constant or, at worst, logarithmic time. A naive nearest-neighbour search requires $\mathcal{O}(n)$ operations for each cell, leading to an overall $\mathcal{O}(n^2)$ scaling. Nearest-neighbour search is a well-studied problem in computer science. In modern particle physics, one of the most prominent applications is jet clustering [26]. However, jet clustering requires searches in two dimensions and the algorithms used there are not suited for the present high-dimensional problem. For approximate searches in high-dimensional spaces, *locality-sensitive hashing* (LSH) [27,28] has been found to be highly efficient. In the following we give a short outline while describing one way in which LSH can be used for finding events that are close to a given cell.

First, we relax the condition that cells have to be kept spherical by subsequently adding nearest neighbours. We will contend ourselves with cells that are small in each phase-space direction and, most importantly, shrink with increasing statistics. This can be achieved by an approximate nearest-neighbour search.

In order to being able to apply standard techniques, we map the events onto points in a high-dimensional Euclidean space. A crucial observation is that this map does not have to preserve distances, neither in the absolute nor in the relative sense. It is sufficient that events that are nearby in phase space are mapped onto points that are nearby in Euclidean space with high probability. To this end, we recall that, in each event e , we group particles into sets $s_{e,t}$ according to their type t . To each of these sets, we now add particles with vanishing four-momentum until for each particle type all sets in all events have the same number of elements,¹ i.e.

$$|s_{e,t}| = |s_{e',t}| \quad \text{for all events } e, e' \text{ and types } t. \quad (24)$$

We then interpret each set $s_{e,t}$ as a tuple $\tau_{e,t} = (p_{e,t,1}, \dots, p_{e,t,P_t})$ of P_t four-momenta, which we sort lexicographically according to the momentum components. Finally, for the event e the coordinates of the resulting point V_e in Euclidean space correspond to the spatial momentum components in order. The first three components $i = 1, 2, 3$ are obtained from the first spatial momentum in the tuple for the first particle type, i.e. $(V_e)_i = (p_{e,1,1})_i$. The next three components correspond to the second momentum in the tuple for the first particle type, and after reaching the last momentum in a tuple we proceed to the tuple for the next particle type. The total dimension of the Euclidean space is therefore

¹ Sets corresponding to different particle types may still have different cardinalities.

$$D = 3 \sum_{t=1}^T |s_{e,t}|, \tag{25}$$

which is independent of the chosen event e . Events e and e' which are nearby in phase space will most likely have a similar number of outgoing particles of each type with similar momenta. Therefore, the coordinates of V_e and $V_{e'}$ will be similar and the points will be close to each other.

We have now recast the problem to finding approximate nearest neighbours between the points V_e in a D -dimensional Euclidean space. To facilitate this task, we consider *random projections* onto lower dimensional sub spaces. This is a well-established technique motivated by the Johnson-Lindenstrauss Lemma [29], which states that it is always possible to find projections that nearly preserve the distances between the points. Specifically, we define projections Π_h onto hyperplanes $h = 1, \dots, H$, each fixed by a randomly chosen unit normal vector as discussed in [30,31]. H should be chosen in such a way that it grows logarithmically with the number of points, i.e. the number of events. We then perform a second projection onto the x axis, i.e. we take the first coordinate of each projected vector:

$$V_e \mapsto c_{e,h} = (\Pi_h V_e)_1. \tag{26}$$

For each $h = 1, \dots, H$, we sort the coordinates $c_{e,h}$ obtained from all events e in the sample and divide them into *buckets* of size B . B may increase at worst logarithmically with the number of events, otherwise the described algorithm for cell creation would no longer fulfil our time complexity constraint, as shown later. Note that events e, e' with nearby coordinates $c_{e,h}, c_{e',h}$ will be inside the same bucket with high probability.

This effectively defines H *hash functions* f_h , each of which maps a given event onto a single integer number, namely the index of the bucket associated with the coordinate $c_{e,h}$:

$$f_h : e \mapsto \text{bucket}[c_{e,h}] = \text{bucket}[(\Pi_h V_e)_1]. \tag{27}$$

These hash functions are *locality sensitive*. Nearby events are likely to be mapped onto the same number.

We use the hash functions to create H *hash tables*, where we store the values of $f_h(e)$ for all events e . When creating a cell, we look up the hashed values $f_h(e)$ for the seed event e . We then consider all events that lie in the same bucket as e in one of the H projected coordinates, i.e. at most HB events. In practice, the number will be much lower, as nearby events will share more than one bucket with e . Starting from the events sharing the most buckets with e , we identify nearest neighbour candidates.² We select candidates with weights

w_i , until we fulfil

$$\sum_i w_i \geq -aw_e, \tag{28}$$

where w_e is the (negative) seed weight and a an arbitrary constant with $a \geq 1$. In our implementation, we choose $a = 2$.

For each of the nearest neighbour candidate, we then compute the actual distance to the seed e using the metric defined in Sect. 2.1. We proceed to add nearest neighbours according to the actual distance to the cell, as we did when using naive nearest-neighbour search, until the sum of weights inside the cell is no longer negative.

Let us now analyse the asymptotic time complexity of the LSH-based algorithm. To create the hash table, we have to compute H projected coordinates for each of the n events, where $H = \mathcal{O}(\log n)$. Once the H hyperplanes are fixed, each coordinate can be computed independently in constant time. In total we compute $\mathcal{O}(Hn) = \mathcal{O}(n \log n)$ coordinates. For each h , we have to sort the coordinates of n events, which gives a time complexity of $\mathcal{O}(Hn \log n) = \mathcal{O}(n \log^2 n)$. The partitioning into buckets is again linear in the number of events, and therefore asymptotically negligible compared to the sorting.

Looking up the values of the hash functions for a single seed event requires $\mathcal{O}(H) = \mathcal{O}(\log n)$ operations. To select the candidates we have to probe all events in H buckets of size B , with time complexity $\mathcal{O}(HB) = \mathcal{O}(\log^2 n)$. Since the number of events needed to compensate the seed weight does not increase with n , only a constant number of distance functions have to be calculated after selecting candidates. In total, the number of cells grows linearly with n and the creation of each cell requires $\mathcal{O}(\log^2 n)$ constant-time steps, which results in an overall time complexity of $\mathcal{O}(n \log^2 n)$.

Finally, we consider the practical performance of our implementation for various sample sizes. Figure 3 compares the median cell size, measured by the largest distance between the seed and any other event inside the cell, and scaling of the computing time to the naive linear nearest-neighbour search. Heuristically, we have chosen $H = 15 \log n$ hash functions and a fixed bucket size of $B = 1000$. We find that the required computing time indeed scales much better with increased sample size. However, the cell size does not seem to decrease significantly with larger statistics. We conclude that further improvements are needed to render the LSH-based approach competitive compared to the linear nearest-neighbour search.

2.3 Relation to positive resampling

Let us comment on the relation between the current approach and the positive resampler suggested in [21]. There, events are collected in histogram bins and positive resampling, i.e.

² In principle, one could consider all events sharing buckets with e without violating the complexity constraints. However, in practice we only need a small fraction of these events to create the cell.

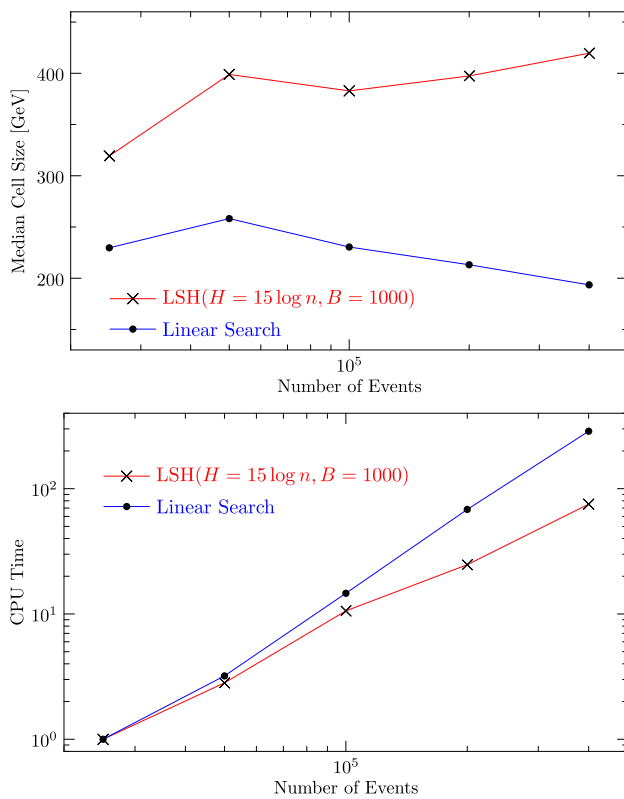


Fig. 3 Comparison between linear nearest-neighbour search and the LSH-based approach with $H = 15 \log n$ hash functions and buckets containing $B = 1000$ events. The upper panel shows the median cell size as a function of the number of input events. The lower panel illustrates the scaling of the computing time with the sample size. The respective timings are normalised to the first data point in order to facilitate comparison and eliminate constant factors resulting from differences in the implementation

replacing event weights by rescaled absolute values, is performed on each bin separately.

There is a clear parallel to the LSH-based method presented in Sect. 2.2. Where the LSH includes random projections of events onto coordinates, positive resampling projects events onto one or more observables defined by the chosen histogram, such as particle transverse momenta. The histogram bins then correspond to the buckets in the locality-sensitive hash tables. Following the same philosophy behind the random projection method, we expect events in the same histogram bins to be also close in other phase-space directions with high probability.

Positive resampling as presented in [21] is therefore similar to cell resampling with LSH-based nearest-neighbour search with a small number of $H \leq 3$ hash functions. An important difference is the definition of a cell. The equivalent for the positive resampler would be an entire bin, i.e. a complete hash bucket. For this reason, it is difficult to perform a rigorous direct comparison between the two approaches. However, we can get an idea by comparing cell resampling

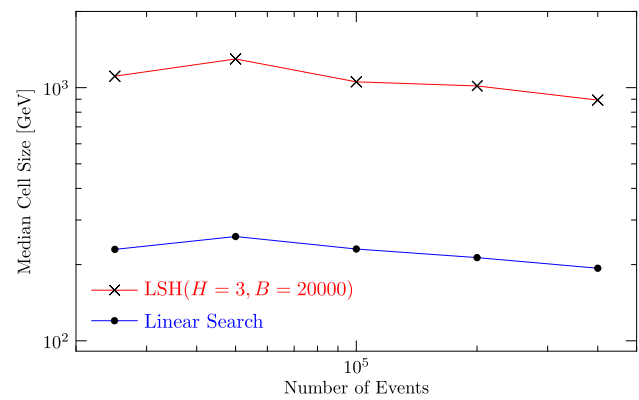


Fig. 4 Comparison of the median cell size between linear nearest-neighbour search and the LSH-based approach with three hash functions

Table 1 Parameters used for event generation

# events	6.1×10^7 (low-statistics input sample) 1.5×10^{10} (high-statistics reference sample)
μ_r, μ_f	$\frac{H_T}{2}$
\sqrt{s}	7 TeV
PDF	NNPDF 3.1 [33,34]
Jet definition	anti- k_r [35] $R = 0.4$ $p_{\perp} > 30$ GeV $ \eta < 4.4$
m_W	80.385 GeV
Γ_W	2.085 GeV
$1/\alpha_{\text{QED}}(0)$	132.232

with linear nearest-neighbour search and LSH-based search with small H . In Fig. 4 we compare the median cell sizes obtained in both approaches for different sample sizes. We have increased the bucket size to $B = 20000$ to ensure that sufficiently many nearest-neighbour candidates are found.

3 Cell resampling for the production of a W boson with two jets

We now demonstrate the proposed method through a highly non-trivial application. We consider the production of a leptonically decaying W^- boson together with at least two jets in proton-proton collisions at 7 TeV centre-of-mass energy calculated at next-to-leading order. Weighted events are generated using Sherpa [15] with OpenLoops [32]. Generation parameters are shown in Table 1.

The 2-parton samples receive contributions from Born, virtual, and subtraction terms, and the 3-parton samples from subtraction and real emission. The sum should be positive, provided reasonable choices for the renormalisation and fac-

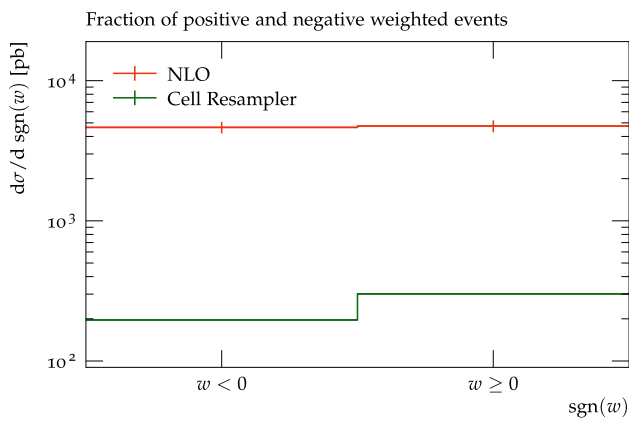


Fig. 5 Contributions of negative- and positive-weight events to the cross section according to the MC_XS RIVET analysis

torisation scales have been applied. We choose $\mu_f = \mu_r = H_T/2$. The subtraction terms can extend far in phase space, which can cause non-local bin-to-bin migration in an analysis of distributions, and issues for the current resampling based on phase space buckets. This issue can be reduced by applying the modified subtractions [36,37] to restrict their size. We have not done so in the current study, partly to illustrate the performance of the resampler in a “worst case” scenario.

To assess the performance of our method, we use 6.1×10^7 events as input and compare both this original NLO sample and the output of the cell resampler to a reference prediction obtained from a high-statistics sample with 1.5×10^{10} events. For the cell resampling, we choose the distance measure introduced in Sect. 2.1.1, and determine the contribution from transverse momentum differences by setting the parameter τ to 10, c.f. Eq. (13). Since the computing time required for cell resampling grows quadratically with the number of events, we split the input sample into nine samples of equal size and apply the resampling separately to each of them. At the same time, we impose an upper limit for the cell radius of 100 GeV in order to achieve cell sizes commensurate with the combined sample size. While this limit prevents us from eliminating *all* negative event weights, it still reduces their overall contribution by more than an order of magnitude. This is illustrated in Fig. 5.

The NLO calculation achieves a $\sim 10^2$ pb cross section as cancellation between terms of size $\sim 5 \times 10^3$ pb, whereas the cell resampler obtains the same cross section as a difference between terms of $\sim 3 \times 10^2$ pb and $\sim 2 \times 10^2$ pb. If we were to reweight the events to weight ± 1 , the original NLO sample would result in almost 20 times as many events compared to unweighting the resampled events. This corresponds to a Monte Carlo estimate of the uncertainty which is larger by a factor of approximately $\sim \sqrt{20}$. Conversely, to achieve the same estimated uncertainty as reached after resampling, a

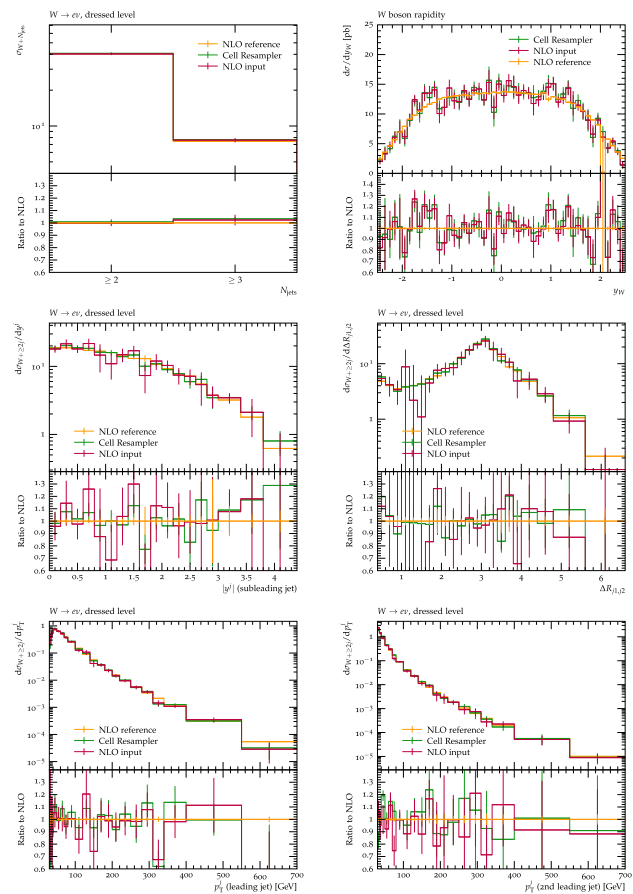


Fig. 6 The results from the pure NLO calculation and of resampler for a subset of the analyses performed. See text for further details

pure NLO sample would require the generation of almost 20 times as many events.

The events are parsed through the standard RIVET [38] analyses MC_XS and MC_WJETS, plus a relevant ATLAS analysis [39] in order to demonstrate the performance for calculations relevant for experimental measurements. As the point of this study is not to compare with experimental data, but to improve on the stability of the predictions, we will for simplicity use the calculation of just $pp \rightarrow e^- \bar{\nu}_e jj$ and not the positron channel. In addition, we also extract the rapidity distribution of the W boson using a custom analysis based on the RIVET framework.

Figure 6 compares the vanilla prediction for the NLO calculation and the result of passing the low-statistics NLO event sample through the resampler for several observables. The results for the inclusive two and three jet cross sections are shown in the top left plot. Both the low-statistics NLO result and that of the cell resampler are stable, as expected. The vertical lines indicate the statistical uncertainty estimated from the weight distribution interpreted as a Monte Carlo sample. The estimate of the uncertainty is larger from the pure NLO

sample than that of the cell resampler, simply because of the reduction in the variance of the weights in the event sample.

Having demonstrated that the predictions for inclusive two and three jet production are preserved by the cell resampling, we now consider differential distributions. The top right plot of Fig. 6 shows the rapidity distribution of the W boson. The level of agreement with the reference prediction within the Monte Carlo error estimate is similar between the original low-statistics NLO sample and the cell resampler. Concretely, the χ^2 per degree of freedom is 1.18 for the cell resampler and 1.25 for the low-statistics NLO sample when comparing to the reference distribution. We conclude that the nominal reduction of the statistical uncertainty through cell resampling is indeed accompanied by a stabilisation of the predicted distributions. Furthermore, the Monte Carlo estimate of the uncertainty in each bin correlates well with the statistical jitter between bins.

It is important to note here that contrary to the approach in [21], there is no input to the cell resampler about any analyses intended for the sample. This implies that the observed good agreement with the high-statistics reference predictions would not be limited to the rapidity distribution of the W boson. Indeed, the plots on the middle row of Fig. 6 show that the reference prediction is also matched well in the distribution in the absolute rapidity of the hardest jet, $|y_{j_1}|$ and $\Delta R_{j_1, j_2}$, both calculated according to the analysis in [39].

Finally, the bottom row in Fig. 6 shows the distributions on the transverse momentum of the hardest and second hardest jet. These distributions are significantly smoother straight from the NLO calculation. They are smoother still from the cell resampler. At the same time, the cells appear to be sufficiently small, such that neither these steeply falling distributions nor the peak in the $\Delta R_{j_1, j_2}$ distribution are smeared out to any visible degree.

4 Conclusions

We have presented *cell resampling* as a method to eliminate negative event weights. Negative weights are redistributed *locally* in phase space and any potential bias introduced by this redistribution becomes arbitrarily small given sufficient statistics.

We have demonstrated the real-world performance on the highly non-trivial example of the production of a W boson with two jets at NLO. It is straightforward to apply our method to arbitrary processes, and we provide an easy-to-use implementation available from <https://cres.hepforge.org/> to this end.

A central ingredient of cell resampling is the definition of a *distance function* between events in phase space. This function should mirror experimental sensitivity without referring to any specific analysis. We have proposed a simple met-

ric, which is shown to perform well in practice. The exploration of more sophisticated distance functions is a promising avenue towards future improvements of the cell resampling method.

Since the quality of the reweighted event samples increases systematically with their size, it is important that the computational cost of cell resampling scales well with the number of input events. To achieve this, we have explored an algorithm for nearest-neighbour search in phase space based on random projections and locality-sensitive hashing. While we find a significantly improved scaling behaviour compared to naive linear search, further improvements will be necessary to obtain the same increase in quality with growing statistics.

Acknowledgements We thank M. Schönherr for detailed communication about Sherpa and its interface to RIVET, and D. Maitre and T. Kuhl for discussions and comments on the manuscript. This work has received funding from the European Union's Horizon 2020 research and innovation programme as part of the Marie Skłodowska-Curie Innovative Training Network MCnetITN3 (grant agreement no. 722104), and the EU TMR network SAGEX agreement no. 764850 (Marie Skłodowska-Curie).

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: The present work does not use any experimental data. The simulated data can be reproduced readily from the information given in Table 1 using publicly available programs.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP³.

Appendix A: Seed selection strategies

When introducing cell resampling in Sect. 2, the only criterion for picking a cell seed was that the event weight has to be negative. We may wonder whether constructing cells in a specific order changes the final result. To answer this question, we assess three different strategies for selecting the seed for the next cell.

1. Choose the event with the most negative weight ("most negative").
2. Choose the next negative-weight event according to the order in which events were generated ("in order").

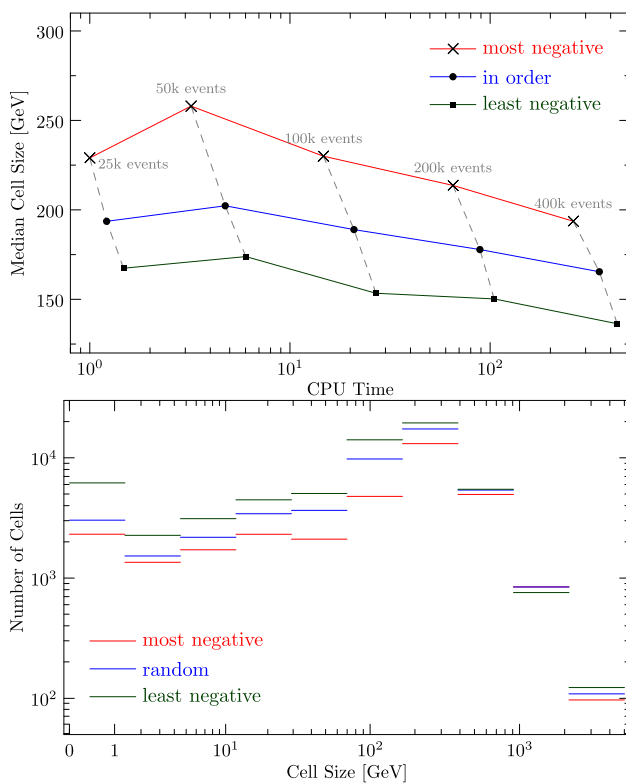


Fig. 7 Comparison of different strategies for selecting cell seeds. On the top, we show the median cell radius as a function of computing time. Timings are normalised to the fastest run with 25,000 events and the “most negative” strategy. On the bottom, we show the distribution of cell radii for the sample with 400,000 events

3. Choose the negative-weight event with weight closest to zero (“least negative”).

We are primarily interested in the quality of the outcome, i.e. the resampling should affect observables as little as possible. As a proxy for this very general requirement, we first consider the cell size distributions. After quality, a secondary criterion is speed. In Fig. 7, we compare the median cell sizes as a function of computing time for samples with between 25 and 400 thousand events. We also show the distribution of cell radii for the largest event sample.

We might be tempted to conclude that the “least negative” strategy is best: despite being the slowest for any given sample, it consistently produces smaller cells given either a fixed number of input events or even a fixed computing time budget. However, we also notice that this strategy not only produces more small-sized cells, but also a larger number of big cells. By choosing seeds with weight close to zero first, we tend to construct cells with increasing radii. Cells that are constructed early on may easily be subsumed by later cells, which implies that the time spend on constructing the earlier cell is wasted and the median cell size is no longer a good indicator of quality.

We can limit the impact of big cells by imposing a maximum cell radius. In this case, we can measure the quality of the resampling through the negative-weight factor

$$\hat{\sigma}_- = \frac{1}{\sigma} \frac{d\sigma}{d \operatorname{sgn} w} \Big|_{w=-1} = - \frac{\sum_{w_i < 0} w_i}{\sum_i w_i}, \quad (29)$$

where w_i is the weight of event i . For the initial event sample with 400,000 events we find $\hat{\sigma}_- \approx 26.1$. Resampling with a maximum cell radius of 100 GeV yields $\hat{\sigma}_- \approx 1.1$, with fluctuations at the level of one per cent between the different strategies. We do not observe any significant differences in run time.

In conclusion, we propose the following modus operandi. First, we estimate the typical cell size through test runs without imposing any limit on the radius. To save computing time we can employ the “most negative” strategy and extrapolate from smaller event samples. We then perform the actual resampling with a cell radius limit of the order of the estimated median. Following the above discussion, we do not expect any significant dependence of the final results on the cell selection strategy.

References

1. S. Catani, M.H. Seymour, A general algorithm for calculating jet cross-sections in NLO QCD. Nucl. Phys. B **485**, 291–419 (1997). [arXiv:hep-ph/9605323](#)
2. S. Catani, S. Dittmaier, M.H. Seymour, Z. Trocsanyi, The dipole formalism for next-to-leading order QCD calculations with massive partons. Nucl. Phys. B **627**, 189–265 (2002). [arXiv:hep-ph/0201036](#)
3. S. Frixione, B.R. Webber, Matching NLO QCD computations and parton shower simulations. JHEP **06**, 029 (2002). [arXiv:hep-ph/0204244](#)
4. P. Nason, G. Ridolfi, A positive-weight next-to-leading-order Monte Carlo for Z pair hadroproduction. JHEP **08**, 077 (2006). [arXiv:hep-ph/0606275](#)
5. J.C. Collins, D.E. Soper, G.F. Sterman, Transverse momentum distribution in Drell–Yan pair and W and Z boson production. Nucl. Phys. B **250**, 199–224 (1985)
6. V.S. Fadin, E. Kuraev, L. Lipatov, On the Pomernanchuk singularity in asymptotically free theories. Phys. Lett. B **60**, 50–52 (1975)
7. I.I. Balitsky, L.N. Lipatov, The Pomernanchuk singularity in quantum chromodynamics. Sov. J. Nucl. Phys. **28**, 822–829 (1978)
8. S. Höche, F. Krauss, M. Schönherr, F. Siegert, QCD matrix elements + parton showers: the NLO case. JHEP **04**, 027 (2013). [arXiv:1207.5030](#)
9. L. Lönnblad, S. Prestel, Merging multi-leg NLO matrix elements with parton showers. JHEP **03**, 166 (2013). [arXiv:1211.7278](#)
10. L. Lönnblad, S. Prestel, Unitarising matrix element + parton shower merging. JHEP **02**, 094 (2013). [arXiv:1211.4827](#)
11. GEANT4 Collaboration, S. Agostinelli et al., GEANT4—a simulation toolkit. Nucl. Instrum. Methods A **506**, 250–303 (2003)
12. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer 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**, 079 (2014). [arXiv:1405.0301](#)
13. J. Bellm et al., Herwig 7.0/Herwig++ 3.0 release note. Eur. Phys. J. C **76**, 196 (2016). [arXiv:1512.01178](#)

14. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., An introduction to PYTHIA 82. *Comput. Phys. Commun.* **191**, 159–177 (2015)
15. Sherpa Collaboration, E. Bothmann et al., Event generation with Sherpa 2.2. *SciPost Phys.* **7**, 034 (2019). [arXiv:1905.09127](https://arxiv.org/abs/1905.09127)
16. K. Danziger, S. Höche, F. Siegert, Reducing negative weights in Monte Carlo event generation with Sherpa. [arXiv:2110.15211](https://arxiv.org/abs/2110.15211)
17. R. Frederix, S. Frixione, S. Prestel, P. Torrielli, On the reduction of negative weights in MC@NLO-type matching procedures. [arXiv:2002.12716](https://arxiv.org/abs/2002.12716)
18. C. Gao, J. Isaacson, C. Krause, i-Flow: high-dimensional integration and sampling with normalizing flows. *Mach. Learn. Sci. Technol.* **1**, 045023 (2020). [arXiv:2001.05486](https://arxiv.org/abs/2001.05486)
19. E. Bothmann, T. Janßen, M. Knobbe, T. Schmale, S. Schumann, Exploring phase space with neural importance sampling. *SciPost Phys.* **8**, 069 (2020). [arXiv:2001.05478](https://arxiv.org/abs/2001.05478)
20. C. Gao, S. Höche, J. Isaacson, C. Krause, H. Schulz, Event generation with normalizing flows. *Phys. Rev. D* **101**, 076002 (2020). [arXiv:2001.10028](https://arxiv.org/abs/2001.10028)
21. J.R. Andersen, C. Gütschow, A. Maier, S. Prestel, A positive resampler for Monte Carlo events with negative weights. *Eur. Phys. J. C* **80**, 1007 (2020). [arXiv:2005.09375](https://arxiv.org/abs/2005.09375)
22. B. Nachman, J. Thaler, Neural resampler for Monte Carlo reweighting with preserved uncertainties. *Phys. Rev. D* **102**, 076004 (2020). [arXiv:2007.11586](https://arxiv.org/abs/2007.11586)
23. B. Stienen, R. Verheyen, Phase space sampling and inference from weighted events with autoregressive flows. [arXiv:2011.13445](https://arxiv.org/abs/2011.13445)
24. P.T. Komiske, E.M. Metodiev, J. Thaler, Metric space of collider events. *Phys. Rev. Lett.* **123**, 041801 (2019). [arXiv:1902.02346](https://arxiv.org/abs/1902.02346)
25. M. Crispim Romão, N.F. Castro, J.G. Milhano, R. Pedro, T. Vale, Use of a generalized energy Mover's distance in the search for rare phenomena at colliders. *Eur. Phys. J. C* **81**, 192 (2021). [arXiv:2004.09360](https://arxiv.org/abs/2004.09360)
26. M. Cacciari, G.P. Salam, Dispelling the N^3 myth for the k_t jet-finder. *Phys. Lett. B* **641**, 57–61 (2006). [arXiv:hep-ph/0512210](https://arxiv.org/abs/hep-ph/0512210)
27. P. Indyk, R. Motwani, Approximate nearest neighbors: Towards removing the curse of dimensionality, in *Proceedings of the 30th ACM Symposium on Theory of Computing* (1998), pp. 604–613
28. J. Leskovec, A. Rajaraman, J. Ullman, *Mining of Massive Datasets* (Cambridge University Press, Cambridge, 2020)
29. W.B. Johnson, J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space. *Contemp. Math.* **26**, 189–206 (1984)
30. M.E. Muller, A note on a method for generating points uniformly on N-dimensional spheres. *Commun. Assoc. Comput. Mach.* **2**, 19–20 (1959)
31. G. Marsaglia, Choosing a point from the surface of a sphere. *Ann. Math. Stat.* **43**, 645–646 (1972)
32. F. Buccioni, J.-N. Lang, J.M. Lindert, P. Maierhöfer, S. Pozzorini, H. Zhang et al., OpenLoops 2. *Eur. Phys. J. C* **79**, 866 (2019). [arXiv:1907.13071](https://arxiv.org/abs/1907.13071)
33. NNPDF Collaboration, R.D. Ball et al., Parton distributions from high-precision collider data. *Eur. Phys. J. C* **77**, 663 (2017). [arXiv:1706.00428](https://arxiv.org/abs/1706.00428)
34. A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht et al., LHAPDF6: parton density access in the LHC precision era. *Eur. Phys. J. C* **75**, 132 (2015). [arXiv:1412.7420](https://arxiv.org/abs/1412.7420)
35. M. Cacciari, G.P. Salam, G. Soyez, The anti- k_t jet clustering algorithm. *JHEP* **04**, 063 (2008). [arXiv:0802.1189](https://arxiv.org/abs/0802.1189)
36. Z. Nagy, Three jet cross-sections in hadron hadron collisions at next-to-leading order. *Phys. Rev. Lett.* **88**, 122003 (2002). [arXiv:hep-ph/0110315](https://arxiv.org/abs/hep-ph/0110315)
37. Z. Nagy, Next-to-leading order calculation of three jet observables in hadron hadron collision. *Phys. Rev. D* **68**, 094002 (2003). [arXiv:hep-ph/0307268](https://arxiv.org/abs/hep-ph/0307268)
38. C. Bierlich et al., Robust independent validation of experiment and theory: Rivet version 3. *SciPost Phys.* **8**, 026 (2020). [arXiv:1912.05451](https://arxiv.org/abs/1912.05451)
39. ATLAS Collaboration, G. Aad et al., Measurements of the W production cross sections in association with jets with the ATLAS detector. *Eur. Phys. J. C* **75**, 82 (2015). [arXiv:1409.8639](https://arxiv.org/abs/1409.8639)