Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment

The ATLAS Collaboration

DOI
10.1007/JHEP08(2022)089

Publication date
2022

Document Version
Final published version

Published in
Journal of High Energy Physics

License
CC BY

Citation for published version (APA):
Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: This paper presents updated Monte Carlo configurations used to model the production of single electroweak vector bosons ($W$, $Z/\gamma^*$) in association with jets in proton-proton collisions for the ATLAS experiment at the Large Hadron Collider. Improvements pertaining to the electroweak input scheme, parton-shower splitting kernels and scale-setting scheme are shown for multi-jet merged configurations accurate to next-to-leading order in the strong and electroweak couplings. The computational resources required for these set-ups are assessed, and approximations are introduced resulting in a factor three reduction of the per-event CPU time without affecting the physics modelling performance. Continuous statistical enhancement techniques are introduced by ATLAS in order to populate low cross-section regions of phase space and are shown to match or exceed the generated effective luminosity. This, together with the lower per-event CPU time, results in a 50% reduction in the required computing resources compared to a legacy set-up previously used by the ATLAS collaboration. The set-ups described in this paper will be used for future ATLAS analyses and lay the foundation for the next generation of Monte Carlo predictions for single vector-boson plus jets production.

KEYWORDS: Hadron-Hadron Scattering

ArXiv ePrint: 2112.09588

https://doi.org/10.1007/JHEP08(2022)089
## Contents

1 Introduction 1

2 Generator configurations 3

2.1 SHERPA 3

2.2 MADGRAPH5_aMC@NLO+PYTHIA 5

2.3 Cross-sections 6

3 Theoretically motivated improvements 7

3.1 Electroweak virtual corrections 7

3.2 Electroweak input scheme 10

3.2.1 Forward-backward asymmetry 10

3.2.2 Polarization of $\tau$-leptons 10

3.3 Modified Catani-Seymour splitting kernels 12

3.4 Scale-setting scheme for unordered histories 13

4 Computational improvements 15

4.1 Speed characterization 17

4.1.1 Perturbative accuracy 18

4.1.2 Additional features 20

4.1.3 Generator comparisons 21

4.2 Statistical enhancement methods 23

4.2.1 Sample biasing and statistical precision 23

4.2.2 Quantitative assessment 26

4.3 Impact on computing resources 30

4.3.1 Per-event CPU cost for production configurations 31

4.3.2 Generated effective luminosity 32

5 Conclusions 35

A Equations for modified Catani-Seymour splitting kernels 36

A.1 Final-initial dipoles 36

A.2 Initial-final dipoles 36

A.3 Initial-initial dipoles 37

The ATLAS collaboration 43
1 Introduction

The production of a single electroweak vector boson ($W$ or $Z/\gamma^*$) in association with jets, collectively referred to as $V+$ jets, is a fundamentally important process at the Large Hadron Collider (LHC) \cite{1}. Measurements of electroweak bosons in the leptonic decay modes provide a clean experimental signature within the ATLAS detector \cite{2} to test the electroweak sector of the Standard Model (SM) and perturbative quantum chromodynamics (QCD) in high-multiplicity final states. In addition, $V+$ jets production is a significant irreducible SM background for a diverse range of analyses within the LHC physics programme: Higgs boson \cite{3,4} and top quark measurements \cite{5,6}, and searches for new physics phenomena beyond the Standard Model (BSM) \cite{7-9}. These analyses often rely on Monte Carlo (MC) simulation-based predictions for background estimation, and therefore an accurate description of the $V+$ jets process with sufficient statistical precision across a wide phase space with many final-state jets is required. Consequently, multi-leg MC generators that combine multiple exclusive jet emissions with the highest possible perturbative accuracy in the strong and electroweak couplings are preferred.

In recent years, the statistical precision of $V+$ jets MC samples has become an increasingly limiting factor for a broad array of ATLAS data analyses. This is primarily due to the long event-generation time of multi-leg MC predictions with high perturbative accuracy that is demanded by the ATLAS physics programme. The relatively large $V+$ jets production cross-section and the rapidly growing integrated luminosity of the proton-proton collision dataset require MC event generators to provide enormous simulated samples, which results in a significant computational challenge for experiments. Current analyses exploring high cross-section regions of phase space already rely on simulated samples that contain many billions of events. Furthermore, the wide range of ATLAS physics searches continue to explore $V+$ jets processes in increasingly extreme regions of phase space, with cross-sections that are many orders of magnitude lower than the inclusive cross-section. In order to meet the statistical demands of these analyses, specialized techniques are required to populate their selection regions. Meeting all of these requirements has historically resulted in a challenging workflow with computing demands that continue to stretch existing resources. With finite computational resources and a fixed computing budget for the High Luminosity LHC (HL-LHC) \cite{10}, fast and efficient MC event generation has already been identified as a potential bottleneck for the large integrated luminosity expected during the HL-LHC runs \cite{11,12}.

The ATLAS collaboration has relied on a multi-leg Sherpa 2.2.1 \cite{13,14} $V+$ jets prediction for the past several years, which has been used for a range of measurements and new-physics searches. This configuration, hereafter referred to as the 2.2.1 set-up, has been substantially extended to meet the statistical needs of these analyses. It required significant resources to produce, due to the long event-generation time, inefficient heavy-flavour filtering techniques, and a relatively large fraction of negative MC event weights. Moreover, many analyses that probe regions of phase space with poorly understood predictions required complex data-driven corrections to their background estimates. The recent efforts highlighted in this paper have focused on improving both the
theoretical and computational aspects of the widely used Sherpa $V+\text{jets}$ predictions in ATLAS. In addition, an alternative multi-jet merged next-to-leading-order (NLO) MadGraph5_aMC@NLO+Pythia \cite{15, 16} configuration with similar perturbative accuracy to the Sherpa prediction has been produced for use by ATLAS for the first time. This configuration is used for an independent comparison with the Sherpa predictions.

The modelling improvements have focused on the level of agreement of the Sherpa 2.2.1 predictions with data using a series of well motivated theoretical changes. It is known that cross-sections predicted by the Sherpa 2.2.1 configuration disagree with the data in several important regions of phase space. In particular, those dominated by extremely energetic jets significantly overestimate the measured cross-section, requiring sophisticated data-driven corrections \cite{17, 18}. Moreover, measurements of the transverse momentum of Drell-Yan lepton pairs show artificial discontinuities around the matrix-element and parton-shower merging scale \cite{19}. Such discontinuities can arise from different cross-sections predicted by the matrix element and parton shower at the merging scale. This paper begins to address these issues through well-motivated theoretical changes including the inclusion of higher-order QCD and electroweak (EW) corrections, a dedicated scale-setting procedure for topologies with unordered parton-shower histories, modified Catani-Seymour parton-shower splitting kernels, and a careful treatment of input electroweak parameters. Comparisons of these changes, along with the newly available MadGraph5_aMC@NLO+Pythia predictions, are performed with publicly available ATLAS and CMS data to highlight the changes.

The computational resources required for large-scale production of any multi-jet merged MC prediction that meets the mentioned theoretical demands is significant. Furthermore, the techniques used to populate the low cross-section regions of phase space can significantly exacerbate the problem. In the Sherpa 2.2.1 configuration, a discrete phase-space slicing technique was used, which resulted in artificial features in the distribution of event weights that adversely impacted the statistical precision of the sample. In addition, the relatively large fraction of MC events with negative weights in the high cross-section regions of phase space, coupled with the long event-generation time, made it difficult to provide a simulated dataset with enough generated luminosity to match the collected data. Therefore, the focus of the computational improvements in this paper has been to reduce the per-event generation time, and the fraction of MC events with negative weights, whilst simultaneously including higher-order EW and QCD corrections. Moreover, a continuous statistical enhancement method is introduced in order to more efficiently populate the low cross-section regions of phase space. A detailed assessment of the CPU resources of the Sherpa configuration is shown, along with approximations that reduce the per-event generation time. In addition, the performance of various statistical enhancement methods is reported and is shown to improve on the approach adopted for the 2.2.1 configuration.

This paper is organized as follows. The Sherpa and MadGraph5_aMC@NLO+Pythia configurations are described in section 2. The theoretically motivated improvements are shown in section 3, in which comparisons with unfolded ATLAS and CMS data are performed where relevant. Section 4 is devoted to computational improvements, which include a characterization of the per-event CPU time consumption, an evaluation of the

\[ -2 - \]
phase-space slicing and statistical enhancement techniques, and a real-world example of the impact that the aforementioned computational improvements have had on computing resources. Finally, conclusions are given in section 5.

2 Generator configurations

This section describes the Sherpa and MadGraph5_aMC@NLO+Pythia configurations and software versions used throughout the paper. An extensive software suite [20] is used for the generation of simulated MC samples shown in this paper. All predictions are normalized to the cross-sections obtained from the generator. Particular attention is given to settings that are highlighted in the modelling and numerical improvements in sections 3 and 4, respectively. For settings not mentioned below, the default values from the event generator are used.

2.1 Sherpa

A legacy Sherpa 2.2.1 configuration is included in this paper as a reference point, against which various improvements are quantified. Further performance studies of this set-up can be found in ref. [14]. An improved configuration is described below and is referred to as the Sherpa 2.2.11 configuration.

Both configurations are simulated with the Sherpa [13] MC generator. In this set-up, NLO-accurate matrix elements for up to two jets, and LO-accurate matrix elements for up to five (four) jets in the five-flavour scheme are calculated with Comix [21] for the 2.2.11 (2.2.1) configuration. The b- and c-quarks are treated as massless at matrix-element level and massive in the parton shower. Virtual QCD corrections for matrix elements at NLO accuracy are provided by OpenLoops [22–25]. The Hessian NNPDF3.0nnlo [26, 27] PDF set is used for the 2.2.11 configuration, whilst the 2.2.1 configuration used the MC replica version of NNPDF3.0nnlo. In the Sherpa Z+jets configurations the invariant mass of the charged leptons is required to be larger than 40 GeV, while the charged lepton plus neutrino invariant mass is required to be larger than 2 GeV for W+jets final states.

The 2.2.11 Z+jets configuration uses an electroweak input scheme specified by \( m_Z \), the QED coupling constant evaluated at the Z resonance, \( \alpha(m_Z) \), and the effective weak mixing angle \( \sin^2 \theta_{\text{eff}} \) [28], while the W+jets configuration uses the so-called \( G_\mu \) scheme specified by the Fermi constant \( G_\mu \), \( m_Z \) and \( m_W \). Remaining EW parameters are derived from these input parameters using tree-level relations. Further discussion and comparisons of these EW input schemes are presented in section 3.2. For the 2.2.1 configuration, all V+jets processes use an effective EW input scheme in which the masses of the W, Z and Higgs boson are taken as inputs, alongside \( \sin^2 \theta_{\text{eff}} \), \( \alpha(m_Z) \), the vacuum expectation value \( (v) \), and Higgs self-coupling constant \( (\lambda) \). Such an input scheme allows certain tree-level relations between some parameters to be violated. Electroweak virtual corrections are included in the 2.2.11 configuration, and are available as an alternative generator weight based on the NLO EW\(_{\text{virt}}\) approach [29, 30]. The effect of these corrections using an additive, multiplicative, or exponentiated combination scheme for higher-order QCD and EW terms is explored in section 3.1. QED radiation to electron, muon, hadron and \( \tau \)-lepton
decays is implemented by the Yennie-Frautschi-Suura (YFS) algorithm. The default YFS Sherpa settings [31, 32] are used.

The NLO matrix elements of a given jet multiplicity are matched to the parton shower using a colour-exact variant of the MC@NLO algorithm [33]. Different jet multiplicities are then merged into an inclusive sample using an improved CKKW matching procedure [34, 35] which is extended to NLO accuracy using the MEPS@NLO prescription [36]. The merging cut is set to $Q_{\text{cut}} = 20 \text{ GeV}$. The default Sherpa Catani-Seymour dipole parton shower [37] scheme and the cluster hadronization model [38] are used. Modified splitting kernels, which give the probability for a given parton to split, are used in the Sherpa 2.2.11 set-up and they are shown to improve the modelling close to the merging scale, as discussed in section 3.3. In the Sherpa 2.2.11 set-up, unordered parton-shower histories are not allowed. If an unordered emission is found, the previous maximal scale will be used as described in section 3.4.

Heavy-flavour hadron production fractions in the Sherpa 2.2.1 set-up are known to disagree with the world average values [39]. A dedicated tuning was therefore performed by the ATLAS collaboration, which was subsequently implemented in the Sherpa event generator. The heavy-flavour baryon production rates previously modelled by a single hadronization parameter in the 2.2.1 configuration were extended to five unique parameters, in order to provide higher fidelity in the control of heavy-flavour baryon production. Two parameters for the charm and beauty baryons, and three for the charm-beauty, charm-strange, and bottom-strange baryons, were introduced. Heavy-flavour hadrons with unknown hadronic decay modes are forced to decay into diquark pairs which are then subsequently hadronised. The invariant mass and binding energy of the diquark pair characterize the hadronization process. The former was adjusted to reduce the low-energy photon ($O(0.1–1) \text{ GeV}$) production rate from such hadronic decays. Furthermore, the $b$-fragmentation function was altered to address an overabundance of jets with small $b$-momentum fraction ($x_b = p_T^b / p_T^{\text{jet}}$), in addition to updating the heavy-flavour decay tables to the world average values [40]. These changes are implemented in Sherpa versions 2.2.5 or later, and thus are present in the Sherpa 2.2.11 configuration.

The MEPS@NLO scale-setting algorithm for the infrared-regular part of the real emission component, commonly referred to as ‘hard events’ or ‘$H$-events’ [36], was studied from a CPU and modelling point of view. In the 2.2.1 configuration, the renormalization scale for these multi-jet configurations is calculated by probabilistically identifying parton-shower topologies through a $k_T$-type clustering algorithm [35]. The full scale-setting algorithm is referred to as the STRICT_METS scale-setting algorithm. The $H$-events represent a large fraction of the total event generation time, which could be significantly reduced using a simplified event-level $H'_T$ scale for this calculation. The $H'_T$ scale is defined by [41]:

$$H'_T = \sum p_T^i + m_V,$$

where the sum extends over all final-state partons excluding those from the boson decay, $p_T^i$ is the jet transverse momentum, and $m_V$ denotes the invariant mass of the boson decay products. The STRICT_METS scale-setting algorithm is used for the remainder
of the calculation. This scale scheme reduces the per-event generation time as shown in section 4.1. It is used in the Sherpa 2.2.11 configuration only.

All MC@NLO processes are affected by negative weights. The presence of negative weights degrades the statistical precision of a sample compared to the one without negative weights, and therefore it is desirable to reduce the fraction of negative weights as much as possible. In the Sherpa 2.2.11 configurations, the MC@NLO matching does not account for subleading gluon spin and colour correlations known to have small effects on final-state observables [42]. This results in an overall reduction in the fraction of negative weights, while preserving the essential features of the matched prediction [43]. In addition, the LO contributions are reweighed with a differential NLO/LO $K$-factor using a $2 \to 2$ core process obtained through an inverse parton-shower clustering. The clustering is first performed in an inclusive mode where both QCD and EW vertices are allowed to be clustered; if the inclusive clustering fails to find a suitable process to evaluate the NLO/LO $K$-factor, it is re-performed in an exclusive mode where the clustering of EW vertices is disabled [44]. This scheme is referred to as the inclusive+exclusive clustering mode. For the 2.2.1 set-up, the highest $2 \to n$ NLO QCD process in the inclusive mode is used as the core process. Using the $2 \to 2$ core process improves the statistical power of the MC sample in high cross-section regions of phase space as shown in section 4.1.

Control of the phase-space sampling is achieved via either a discrete slicing or continuous phase-space biasing technique. In the case of the Sherpa 2.2.1 configuration, the sample is divided into discrete $\max(H_T, p_T^V)$ regions (slices), with boundaries at $[0, 70, 140, 280, 500, 1000, 6500]$ GeV. The $H_T$ variable is defined as the scalar sum of all partonic jet transverse momenta and $p_T^V$ is the transverse momentum of the lepton pair. The Sherpa 2.2.11 configuration was statistically enhanced in the $\max(H_T, p_T^V)$ variable using an analytic enhancement technique. Section 4.2.1 discusses the strengths and weaknesses of different enhancement techniques in terms of statistical performance.

QCD scale uncertainties have been evaluated on-the-fly using 7-point variations of the factorization and renormalization scales coherently in the matrix elements and the parton shower [45]. The scales are varied independently by factors of 0.5 and 2 while avoiding cases where the factorization and renormalization scales move in opposite directions. PDF uncertainties for the nominal NNPDF3.0NNLO PDF set are evaluated using the 100 variation replicas in case of the 2.2.1 set-up and using the Hessian eigenvector set for the 2.2.11 set-up, as well as $\pm0.001$ shifts of $\alpha_S$ around the nominal value of 0.118. Additionally, the central values of the CT18NNLO [46] and MSHT2020NNLO [47] PDF sets are included as additional generator weights.

A summary of the Sherpa configurations is shown in table 1.

2.2 MadGraph5_aMC@NLO+Pythia

An additional prediction using the MadGraph5_aMC@NLO 2.6.5 [15] program is used to generate weak bosons with up to three additional partons in the final state at NLO accuracy in the strong coupling. The renormalization and factorization scales are set to one half the transverse mass of all final state partons. The accuracy on the absolute value of the cross section obtained during the phase-space integration is required to be less than 0.0001.
The showering and subsequent hadronization is performed using PYTHIA 8.240 [16] with the A14 tune [48], and the NNPDF2.3lo PDF set [49] with $\alpha_S = 0.130$. The different jet multiplicities are merged using the FxFx NLO matrix-element and parton-shower merging prescription [50]. The sampled phase-space is biased using a function equal to the square of sum of final state momenta (i.e leptons plus partons). Heavy-flavor hadron decays are handled by EvtGen 1.7.0 and QED radiation is added by PYTHIA 8. The inclusive fraction of generated events with negative weights is approximately 17%, and reaches 35% for $p_T$ and $H_T$ phase spaces above 100 GeV.

The MadGraph5_aMC@NLO+PYTHIA calculation uses a five-flavour number scheme with massless $b$- and $c$-quarks at the matrix-element level, and massive quarks in the PYTHIA 8 shower. At the event-generation level, the jet transverse momentum is required to be at least 10 GeV, with no restriction on the absolute value of the jet pseudorapidity. The PDF set used for event generation is NNPDF3.1nnlo supplemented with the LUXqed photon PDFs using $\alpha_S = 0.118$ [51]. The merging scale is set to $Q_{\text{cut}} = 20$ GeV. Scale variations where the renormalization and factorization scales have been varied independently by a factor of 2 or 0.5 in the matrix element are included as alternative generator event weights [52].

### 2.3 Cross-sections

The predicted cross sections from the SHERPA and MadGraph5_aMC@NLO+PYTHIA event generators for the configurations described above are shown in table 2. Additionally, an inclusive reference cross-section calculated with FEWZ [53] using the MSTW2008nnlo PDF set to NNLO QCD accuracy is shown for comparison. The effects of scale uncertainties are estimated by varying the factorization and renormalization scales independently around their central values by factors of 2 and 0.5. The PDF uncertainties are estimated from the NNPDF3.0nnlo and NNPDF3.1nnlo variations for the SHERPA and Mad-

<table>
<thead>
<tr>
<th>Configuration</th>
<th>SHERPA 2.2.1</th>
<th>SHERPA 2.2.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generator version</td>
<td>SHERPA 2.2.1</td>
<td>SHERPA 2.2.11</td>
</tr>
<tr>
<td>PDF set</td>
<td>NNPDF3.0nnlo</td>
<td>NNPDF3.0nnlo</td>
</tr>
<tr>
<td>EW input scheme</td>
<td>Effective</td>
<td>$\sin^2 \theta_{\text{eff}}$</td>
</tr>
<tr>
<td>QCD accuracy</td>
<td>0–2j@NLO+3,4,5j@LO</td>
<td>0–2j@NLO+3,4,5j@LO</td>
</tr>
<tr>
<td>NLO EW $\text{virt}$ corrections</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Subtraction scheme</td>
<td>Default</td>
<td>Modified Catani-Seymour</td>
</tr>
<tr>
<td>Special treatment for unordered histories</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Scale for $\text{H}\text{h}$-events</td>
<td>STRICT_METS</td>
<td>$H_T$</td>
</tr>
<tr>
<td>Gluon colour/spin exact matching</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Core process for $K$-factor</td>
<td>2 $\rightarrow$ 4</td>
<td>2 $\rightarrow$ 2</td>
</tr>
<tr>
<td>Phase-space strategy</td>
<td>Sliced in $\text{max}(H_T, p_T^V)$</td>
<td>Analytic enhancement</td>
</tr>
</tbody>
</table>

Table 1. Summary of the SHERPA 2.2.1 and 2.2.11 configurations.
\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
 & $\sigma(pp \to \ell\ell)$ [pb] & $\sigma(pp \to \mu\nu)$ [pb] \\
\hline
SHERPA 2.2.1 & 2160 \pm 86 (scale) \pm 53 (PDF+$\alpha_S$) & 20697 \pm 828 (scale) \pm 507 (PDF+$\alpha_S$) \\
SHERPA 2.2.11 & 2221 \pm 155 (scale) \pm 47 (PDF+$\alpha_S$) & 21781 \pm 1525 (scale) \pm 462 (PDF+$\alpha_S$) \\
MG5\_aMC@NLO+Pythia & 2268 \pm 92 (scale) \pm 23 (PDF) & 22163 \pm 898 (scale) \pm 225 (PDF) \\
\hline
\end{tabular}
\caption{Event generator predicted cross-sections for each of the MC configurations. The reference NNLO QCD cross-section is calculated with FEWZ [53] using the MSTW2008nnlo PDF set.}
\end{table}

Graph5\_aMC@NLO+Pythia set-ups, while the inclusive NNLO PDF uncertainty is computed using the 68\% C.L. MSTW2008nnlo PDF eigenvector error sets.

The cross-sections predicted by the different configurations can differ for numerous reasons, including the perturbative accuracy of the set-up, the choice of EW input scheme, and the choice of PDF set. The cross-section from the 2.2.11 configuration is larger than that from the SHERPA 2.2.1 configuration due to the inclusion of the 5j@LO matrix-elements. All cross-sections are larger than the inclusive NNLO QCD value since both the SHERPA and MadGraph5\_aMC@NLO+Pythia configurations include contributions beyond an inclusive NNLO QCD calculation. The scale uncertainties in the SHERPA 2.2.11 set-up include variations in the 5j@LO matrix element as well as coherent variations in the parton shower, which together lead to a \sim 40\% increase in the total inclusive scale uncertainty relative to the SHERPA 2.2.1 configuration. All results presented in this paper are normalized to the respective event generator’s cross-section prediction.

3 Theoretically motivated improvements

This section reports on the theoretical changes that are included in the SHERPA 2.2.11 configuration. They include four major topics: the inclusion of higher-order QCD and electroweak corrections, an electroweak input scheme for $Z +$jets events that specifies the weak mixing angle, specialized treatment of the dipole operators for the Catani-Seymour splitting kernels, and a special treatment of unordered parton-shower histories. These changes are described below and the performance of the SHERPA 2.2.11 and MadGraph5\_aMC@NLO+Pythia configurations is compared with that of the 2.2.1 configuration. Rivet [54] is used to compare the predictions with previously published ATLAS or CMS data to highlight the theoretical changes.

3.1 Electroweak virtual corrections

Electroweak corrections to single boson production can play a significant role in regions of phase space where the boson is produced with large transverse momentum, owing to the appearance of EW Sudakov logarithms of the form $\log(s/m_V^2)$ [30]. These corrections are crucial for an accurate description of single vector bosons and will play an important role for BSM searches probing high-$p_T$ phase-space regions. The full electroweak calculation contains contributions from virtual loops and from real emission of gauge bosons. The
configuration presented in this paper only contains the virtual-loop terms, since applying the full NLO EW corrections is not currently possible in a multi-jet merged set-up. Moreover, the real emission of a gauge boson leads to a fundamentally different experimental signature that is typically produced in separate samples.

Amplitudes for the electroweak virtual corrections are calculated by OpenLoops and implemented in the Sherpa event generator for multi-jet merged set-ups. Results from MEPS@NLO merged set-ups with up to two additional jets at next-to-leading order in QCD for selected processes are shown in ref. [30]. In this paper, electroweak virtual corrections are included in the multi-jet merged set-ups with 0–2@NLO+3–5@LO QCD accuracy, and their effects are demonstrated relative to the NLO QCD-only prediction for all leptonic decay modes of the $W$ and $Z$ bosons. Moreover, multiple NLO QCD and NLO EW combination schemes are compared, and these are stored as alternative generator weights.

The cross-section calculated at NLO QCD and NLO EW accuracy for a $2 \rightarrow n$ final state, where the higher-order terms in QCD and EW couplings are summed (additive scheme), can be expressed as:

$$\sigma_{n,\text{QCD+EW}}^{\text{NLO}} = \sigma_{n,\text{LO}} + \Delta \sigma_{n,\text{QCD}}^{\text{NLO}} + \Delta \sigma_{n,\text{EW, virt}}^{\text{NLO}}$$

where $\sigma_{n,\text{LO}}$ denotes the Born-level cross-section and $\Delta \sigma_{n,\text{QCD}}^{\text{NLO}}$ the NLO QCD corrections to the Born process. The $\Delta \sigma_{n,\text{EW, virt}}^{\text{NLO}}$ term only contains the virtual-loop components, and in the NLO EW$_{\text{virt}}$ approximation these are specified by [30]:

$$\Delta \sigma_{n,\text{EW, virt}}^{\text{NLO}} = [B_{n,\text{mix}}(\Phi_n) + V_{n,\text{EW}}(\Phi_n) + I_{n,\text{EW}}(\Phi_n)] d\Phi_n,$$  \hspace{1cm} (3.1)

where $\Phi_n$ represents the phase space of the $n$-particle final state, $B_{n,\text{mix}}$ are the mixed QCD-EW Born contributions with order $\alpha^{-1}$ and $\alpha^3$ in the strong and electroweak coupling, respectively, $V_{n,\text{EW}}$ is the exact one-loop virtual correction to the Born process with order $\alpha^3$ in the electroweak coupling, and $I_{n,\text{EW}}$ is the approximated subtraction term used to render the above cross-section finite. The approximated electroweak virtual corrections in eq. (3.1) are denoted by NLO EW$_{\text{virt}}$ in this paper. They were developed and shown to be in good agreement with the full NLO EW calculation in ref. [30] up to a few percent, with the exception of phase spaces dominated by real emission contributions such as the off-shell region of the $m_{\ell\ell}$ invariant mass distribution. Such phase-spaces are dominated by real QED emissions from charged leptons that are not included in the NLO EW$_{\text{virt}}$ calculations, and instead supplemented by dedicated samples. In the results shown below, the one-loop contributions are included for final states with up to two extra jets.

Two additional NLO QCD and NLO EW combination schemes are provided by the Sherpa framework. The first is the multiplicative scheme, which is defined as:

$$\sigma_{n,\text{QCD×EW}}^{\text{NLO}} = (\sigma_{n,\text{LO}} + \Delta \sigma_{n,\text{QCD}}^{\text{NLO}}) \times (1 + \Delta \sigma_{n,\text{EW, virt}}^{\text{NLO}}).$$

The second scheme is referred to as the exponentiated scheme, and is defined as:

$$\sigma_{n,\text{QCD×EW, virt}}^{\text{NLO}} = (\sigma_{n,\text{LO}} + \Delta \sigma_{n,\text{QCD}}^{\text{NLO}}) \times e^{\Delta \sigma_{n,\text{EW, virt}}^{\text{NLO}}}.\n$$
Figure 1. Differential distributions of the transverse momentum of (a) the boson and (b) the highest transverse momentum jet for $pp \to \ell\ell, \nu\nu, \ell\nu$ processes. The QCD accuracy of all predictions is $0-2j@NLO+3-5j@LO$. The electroweak virtual corrections are denoted by $\Delta_{EW}$. The $\ell\ell$ and $\ell\nu$ final states have been averaged from electron, muon and $\tau$-lepton final states. The shaded bands include scale uncertainties calculated from the envelope of renormalization and factorization variations.

The multiplicative and exponentiated schemes are motivated by the factorization of the QCD and EW Sudakov corrections at high energies [55].

Comparisons of the NLO QCD-only prediction with and without the virtual electroweak corrections are shown in figure 1 using the SHERPA 2.2.11 configuration. The corrections reach 30% around $p_T^V \sim 1$ TeV and qualitatively follow the expected logarithmic behaviour. The relative size of the corrections is in good agreement with those obtained from the fixed-order NLO EW calculations when including corrections for up to two additional jets at next-to-leading order [30]. The multiplicative and exponentiated contributions show the largest deviations from the NLO-only prediction and are in good agreement with each other. The difference between the additive and multiplicative or exponentiated predictions is interpreted as a source of systematic uncertainty for missing higher orders resulting from mixed QCD-EW contributions, and around $p_T^V$ of 1 TeV the uncertainty is approximately 10% for all final states. Future measurements of vectors bosons produced with large transverse momentum will be crucial in testing these predictions, in addition to understanding NLO EW effects overall.
3.2 Electroweak input scheme

An EW input scheme defines the set of input parameters which are use to derive any remaining EW parameters. Some choices of EW scheme can result in values that differ from experimentally determined values. In particular, using tree-level relations in the \( G_\mu \) scheme the weak mixing angle is found to be:

\[
\sin^2 \theta_W = \left( 1 - \left( \frac{m_W}{m_Z} \right)^2 \right) \approx 0.223,
\]

which disagrees with the experimentally determined value of \( \sin^2 \theta_{\text{eff}} = 0.23113 \pm 0.00014 \) [56]. Although the \( G_\mu \) scheme is an attractive choice for \( Z + \text{jets} \) (Drell-Yan) processes at hadron colliders because it absorbs a large part of the radiative corrections in the LO prediction [28], this level of discrepancy causes measurable differences in the forward-backward asymmetry of leptons from \( Z \) boson decays and the \( \tau \)-lepton polarization as highlighted below.

The \( \sin^2 \theta_{\text{eff}} \) scheme by construction ensures the correct value of \( \sin^2 \theta_{\text{eff}} \) since it is specified as one of the input parameters. The \( \sin^2 \theta_{\text{eff}} \) scheme based on ref. [28] was implemented by the OPENLOOKS and SHERPA authors [30] throughout the course of the work presented in this paper. It is used for the SHERPA 2.2.11 \( Z + \text{jets} \) configuration for both the NLO QCD and NLO EW_{virt} calculations. In this scheme, the derived \( W \) boson mass differs from its experimentally measured value, but for single \( Z \) boson production this has negligible impact on the physics performance.

3.2.1 Forward-backward asymmetry

The vector and axial-vector couplings in the \( pp \rightarrow Z \rightarrow \ell\ell \) process lead to a forward-backward asymmetry (\( A_{\text{FB}} \)) in the polar angle of the lepton with respect to the incoming quark direction in the rest frame of the dilepton system. This forward-backward asymmetry has been measured as a function of the dilepton invariant mass by ATLAS in \( \sqrt{s} = 7 \text{ TeV} \) proton-proton collisions and used to extract the weak mixing angle [57]. The measured asymmetry values are used to test the performance of the \( G_\mu \) and \( \sin^2 \theta_{\text{eff}} \) schemes using the SHERPA 2.2.11 configuration.

The forward-backward asymmetry is most sensitive to the weak mixing angle close to the \( Z \) boson’s pole mass since the size of \( A_{\text{FB}} \) is proportional to \( \sin^2 \theta_W \) at \( m_{\ell\ell} \sim m_Z \). As shown in figure 2, with NLO QCD accuracy the \( G_\mu \) scheme leads to an overestimate of \( A_{\text{FB}} \) near the \( Z \)-pole at the 1\( \sigma \)-2\( \sigma \) level. The inclusion of EW corrections provides a better overall description of the data in the \( G_\mu \) scheme. The \( \sin^2 \theta_{\text{eff}} \) scheme provides a good description of the data with and without EW corrections, with a level of agreement with the data similar to that achieved with the NLO QCD+EW_{virt} \( G_\mu \) scheme. The reduced dependency on EW corrections makes the \( \sin^2 \theta_{\text{eff}} \) scheme preferable to the \( G_\mu \) scheme for Drell-Yan production.

3.2.2 Polarization of \( \tau \)-leptons

Similarly to the forward-backward asymmetry, the polarization of \( \tau \)-leptons in \( Z/\gamma^* \rightarrow \tau\tau \) decays is sensitive to the value of \( \sin^2 \theta_{\text{eff}} \) since the \( Z \) boson couples differently to left-handed and right-handed \( \tau \)-leptons due to the chiral coupling asymmetry. The expected
Figure 2. The forward-backward asymmetry $A_{FB}$ for the $G_{\mu}$ and $\sin^{2}\theta_{\text{eff}}$ electroweak input schemes in the (a) central-central and (b) central-forward electron selections. MC simulation $pp \rightarrow e^{\pm}\nu_{\mu}$ predictions were obtained using the SHERPA 2.2.11 configuration at a centre-of-mass energy of $\sqrt{s} = 7$ TeV. Uncertainties in the predictions are statistical only. Data points are taken from ref. [57].

The polarization ($P$) of $\tau$-leptons is extracted from a fit to the visible momentum fraction, $x_{\pi} = |\vec{p}(\pi)/|\vec{p}(\tau)|$ in $\tau \rightarrow \pi \nu$ decay, using the following semi-analytic formula [60]:

$$\frac{1}{N} \frac{dN}{dx_{\pi}} = 1 + P(2x_{\pi} - 1).$$

Simulated events with two opposite-sign final-state $\tau$-leptons with an invariant mass of $66 < m_{\tau\tau} < 116$ GeV are selected. The distribution of $x_{\pi}$ normalized to unity and corresponding fit results of eq. (3.2) for the $G_{\mu}$ and $\sin^{2}\theta_{\text{eff}}$ schemes are shown in figure 3.

The outlier point at low $x_{\pi}$ arises from the finite pion mass and is excluded from the fit. The polarization of $\tau$-leptons in the $G_{\mu}$ scheme is found to be $(-20.2 \pm 0.9)\%$ at NLO in QCD and $(-17.6 \pm 0.9)\%$ when including additive NLO EW$_{\text{virt}}$ corrections. In the $\sin^{2}\theta_{\text{eff}}$ scheme, a $\tau$ polarization value of $(-14.4 \pm 0.9)\%$ is extracted and is insensitive to EW virtual corrections. Given this value’s better agreement with the SM prediction and the measured values from both LEP and ATLAS, and the reduced dependency on EW corrections, the $\sin^{2}\theta_{\text{eff}}$ scheme is preferred.
Figure 3. Comparison of the $\tau$-lepton visible momentum fraction $x_\pi$ for the $G_\mu$ and $\sin^2 \theta_{\text{eff}}$ EW input schemes. Only data points in the $\sin^2 \theta_{\text{eff}}$ scheme are shown. The fitted values are shown along with the extracted polarization of the $\tau$-leptons. The distribution is normalized to unity. Uncertainties are statistical only.

3.3 Modified Catani-Seymour splitting kernels

The splitting kernels in the Catani-Seymour [61] based parton shower of SHERPA are known to exhibit non-singular corrections which can alter the radiation pattern significantly in the regions of phase space where the matrix element is not logarithmically enhanced. While a large part of these regions is covered by the hard matrix elements in an NLO-matched and multi-jet merged set-up, some part of it will always be populated by the parton shower. In addition, the radiation pattern of the shower also affects the Sudakov suppression in the hard region dominated by matrix elements.

In the SHERPA 2.2.11 configuration, the parton-shower splitting kernels are modified to better reflect the known soft limits of the hard matrix elements. The kinematic parameterization of an emission in the collinear kinematic region is modified to match the partial-fractioned soft eikonal part to the collinear limit. The modified splitting kernels can be found in appendix A. These changes are made consistently in both the parton shower and the NLO fixed-order and NLO matched calculations, and mainly affect the physics in the parton shower, thus being relevant for $p_T^V$ around and below the merging scale (20 GeV). No significant effect was observed in the fraction of negative event weights due to these modifications.

The SHERPA 2.2.11 set-up contains the modified splitting kernels described above, while the 2.2.1 configuration does not and can be used to assess the modelling changes.
Predictions from these set-ups are compared with a measurement of the transverse momentum $p_{T}^{ll}$ of Drell-Yan lepton pairs in the mass range 66–116 GeV [19]. The measurement used 36.1 fb$^{-1}$ of 13 TeV proton-proton collision data collected by the ATLAS experiment. The predictions from the SHERPA 2.2.11 and 2.2.1 configurations, along with the MADGRAPH5\_aMC@NLO+PYTHIA FxFx configuration, are shown in figure 4 as a function of the unit normalized measured $p_{T}^{ll}$ and $\phi_{y}^{*}$ [19] variables. The latter variable is constructed from only the angular separation of the leptons, which reduces the impact of the lepton momentum resolution at low-$p_{T}^{ll}$ and depends only on the well-measured directions of the leptons. The red uncertainty band is shown on the NLO QCD prediction for the SHERPA 2.2.11 setup and includes scale and merging scale variations, where the merging scale was varied to 15 GeV and 30 GeV. The scale and merging scale uncertainties are added in quadrature. Additionally, the figure shows in blue the effect of adding the additive NLO EW$_{\text{virt}}$ prediction. The blue uncertainty band is calculated from the envelope of multiplicative and exponentiated NLO EW$_{\text{virt}}$ schemes.

The new splitting kernels used in the SHERPA 2.2.11 configuration significantly improve the description of the data below and around the merging scale ($p_{T}^{ll} \sim 20$ GeV). The artificial discontinuity close to the merging scale is reduced, representing a closer match of cross-sections predicted by the matrix element and parton shower at the merging scale. As a result of these changes, the uncertainty due to variations of the merging scale is reduced by a factor of two compared to the SHERPA 2.2.1 set-up and contributes less than 5% around the merging scale. For intermediate-$p_{T}^{ll}$ and large $\phi_{y}^{*}$ values, the predicted cross sections of the SHERPA 2.2.11 configuration is reduced by approximately 10% compared to the SHERPA 2.2.1 prediction. This is primarily a consequence of the scale changes for unordered histories discussed in the next section that reduce the cross sections for events with many jets. While the central value disagrees with the measured data points in the intermediate-$p_{T}^{ll}$ and large $\phi_{y}^{*}$ regions, the predictions are in agreement within the scale uncertainties shown in red and the significant over-prediction observed in the SHERPA 2.2.1 for the highest $p_{T}^{ll}$ values measured is reduced. The NLO EW$_{\text{virt}}$ prediction reduces the cross-section further in the high $p_{T}^{ll}$ and $\phi_{y}^{*}$ regions, as expected from figure 1. The uncertainties from different NLO EW$_{\text{virt}}$ schemes is small for the majority of the phase-space, indicating good agreement between the three NLO EW$_{\text{virt}}$ predictions.

The MADGRAPH5\_aMC@NLO+PYTHIA FxFx prediction uses a different subtraction technique based upon the FKS method [62–64] and exhibits fundamentally different modelling behaviour than the SHERPA predictions in the low $p_{T}^{ll}$ phase-space. Below the merging scale of 20 GeV, the MADGRAPH5\_aMC@NLO+PYTHIA prediction models the data well, but around the merging scale it disagrees with the data in the opposite direction relative to the SHERPA configurations. For the intermediate and high-$p_{T}^{ll}$ regions, the MADGRAPH5\_aMC@NLO+PYTHIA prediction agrees very well with the SHERPA 2.2.11 NLO QCD prediction.

### 3.4 Scale-setting scheme for unordered histories

The MEPS@NLO algorithm used by the SHERPA event generator matches fixed-order perturbative calculations to parton showers by probabilistically identifying parton-shower to-
global hardness scale associated with the event.

evaluated at the nodal with the production of jets in this ‘unordered’ region of phase space must therefore not be potentially large enhancements from evaluating the strong coupling at small transverse they correspond to an evolution in rapidity. Due to the absence of the suppression factor, these specific contributions are not taken into account in standard parton showers because Such configurations would also be accompanied by Sudakov-type suppression factors, but are widely separated in rapidity but not separated in the may be violated. A prime example would be jets of similar transverse momentum that coupling. However, in extreme regions of phase space, this scale hierarchy in eq. (3.3) which leads to small Sudakov factors counteracting large values of the running strong coupling. However, in extreme regions of phase space, this scale hierarchy in eq. (3.3) may be violated. A prime example would be jets of similar transverse momentum that are widely separated in rapidity but not separated in the $k_T$ scale used in the clustering. Such configurations would also be accompanied by Sudakov-type suppression factors, but these specific contributions are not taken into account in standard parton showers because they correspond to an evolution in rapidity. Due to the absence of the suppression factor, potentially large enhancements from evaluating the strong coupling at small transverse momenta can then not be compensated for, and the differential cross-section, together with its perturbative uncertainty, becomes uncontrollably large. The strong coupling associated with the production of jets in this ‘unordered’ region of phase space must therefore not be evaluated at the nodal $k_T$ scales of the jet clustering. Instead it should be computed at a global hardness scale associated with the event.

Figure 4. Unit normalized cross-sections as a function of (a) the transverse momentum of the dilepton system and (b) $\phi^\ast$. Predictions from the SHERPA and MadGraph5_aMC@NLO+PYTHIA are compared with data points taken from the ATLAS measurement [65]. The red uncertainty band corresponds to scale and merging scale variations in the 2.2.11 NLO+QCD SHERPA set-up. The blue uncertainty band is the envelope of multiplicative and exponentiated NLO EW$_{\text{virt}}$ schemes around the additive NLO EW$_{\text{virt}}$ prediction.
The Sherpa 2.2.11 configuration uses a new algorithm to handle these cases, where unordered histories are no longer used in the scale-setting algorithm. If the $k_T$ scale decreases during the clustering, the largest previously identified scale will be used. In comparison to previous predictions from the Sherpa generator, this typically leads to reduced cross-sections and smaller uncertainties in regions of the phase space that contain multiple jets of similar hardness. In the reverse clustering procedure, nodes containing EW particles are only included if they involve a strongly interacting particle (e.g. $u$-quark clustered with $W$, but not $e$ clustered with $\nu$). The virtuality of the EW propagators is accounted for during the clustering.

The Sherpa 2.2.11 configuration includes the changes described above, while the Sherpa 2.2.1 configuration allowed for unordered histories. Comparisons of the Sherpa 2.2.11 and 2.2.1 predictions, along with the MadGraph5 aMC@NLO+Pythia FxFx prediction are shown in figures 5 and 6 for $Z +$ jets and $W +$ jets measurements, respectively. The NLO EW$_{\text{virt}}$ predictions are shown by the blue curves with an uncertainty band that includes the envelope of multiplicative and exponentiated NLO EW$_{\text{virt}}$ schemes around the additive prediction.

The $W +$ jets data is based upon a CMS measurement of the differential cross-sections for the associated production of a $W$ boson and jets in 2.2fb$^{-1}$ of proton-proton collisions at $\sqrt{s} = 13$ TeV [66]. The predicted cross-section for events with at least three jets is significantly reduced and are in agreement with the data. In the old scheme, the predicted cross-section overestimated the data by up to 50% for jets produced with momentum greater than 300 GeV. In the new scheme, the rates are considerably reduced and in better agreement with the data. This improved agreement with the data for highly energetic jets is at the expense of a higher predicted cross section for events with low jet transverse momentum, and is observed to overshoot the data by as much as 15% as shown in figure 5(b). The NLO EW$_{\text{virt}}$ corrections reduce the QCD-only prediction by $\sim$10% for large jet momentum and improve agreement with data.

For the $Z +$ jets comparisons, the data is based upon an ATLAS measurement of the production cross-section of a $Z$ boson in association with jets in 3.16fb$^{-1}$ of $pp$ collisions at $\sqrt{s} = 13$ TeV [67]. The predicted cross-sections for each jet emission are in better agreement with the data when using the Sherpa 2.2.11 configuration. The cross-sections for large jet momenta are reduced by approximately 10% in the Sherpa 2.2.11 configuration when compared to the Sherpa 2.2.1 prediction. This introduces a small disagreement of the central values with the measured data points. However, the agreement is within the perturbative scale uncertainties. Similarly to the $W +$ jets results, the NLO EW$_{\text{virt}}$ corrections reduce the QCD-only prediction by $\sim$5%–10% for large jet momenta.

4 Computational improvements

An accurate description of SM processes in proton-proton collisions, as discussed in the previous section, is paramount to the success of the ATLAS physics programme. At the same time, the generated effective luminosity of a MC sample should ideally exceed that of the data across the wide range of phase spaces explored by measurements and searches.
Figure 5. Comparisons between generator-level $W \rightarrow \mu\nu$ predictions and measured data points. The red uncertainty band includes the envelope of the 7-point scale variations around the SHERPA 2.2.11 NLO QCD prediction. The blue uncertainty band is the envelope of multiplicative and exponentiated NLO EW$_{\text{virt}}$ schemes around the additive NLO EW$_{\text{virt}}$ prediction. Data points are taken from CMS measurements at $\sqrt{s} = 13$ TeV [66].

Looking towards the HL-LHC era [10], this is only possible if a reduction of the per-event generation time is achieved and smart sampling techniques are used that do not adversely impact the statistical performance of the generated events. The following section addresses these issues using $V$+ jets processes as a case study. The per-event CPU time for various generator configurations is presented in section 4.1, while section 4.2 explores the statistical consequences of various sampling techniques to populate the low cross-section regions of phase space.
Figure 6. Comparisons between generator-level $Z \rightarrow ee$ predictions and measured data points. The red uncertainty band includes the envelope of the 7-point scale variations around the SHERPA 2.2.11 NLO QCD prediction. The blue uncertainty band is the envelope of multiplicative and exponentiated NLO EW$_{\text{virt}}$ schemes around the additive NLO EW$_{\text{virt}}$ prediction. Data points are taken from ATLAS measurements at $\sqrt{s} = 13$ TeV [67].

4.1 Speed characterization

The time it takes to generate a single event, or the per-event CPU cost, is highly dependent on the configuration of the MC generator: for example, the perturbative accuracy of the cross-section in both the strong and electroweak couplings, the number of extra partons at matrix-element level, the matrix-element to parton-shower matching algorithm, and the scale-setting algorithm. Moreover, additional scale and PDF predictions used to evaluate sources of theoretical uncertainty are routinely stored in the nominal sample as alternative
generator weights. Depending on the number of additional predictions and the methodology with which one evaluates the variation weight, significant CPU cost can be incurred. This section quantifies the per-event CPU cost of configurations with varying perturbative accuracy and Sherpa-specific scale-setting and clustering algorithms, collectively referred to as additional features. The $pp \rightarrow e^\pm e^\mp +$ jets process is used as a representative process, and events are generated using Sherpa 2.2.11. Additionally, a comparison between the MadGraph5_aMC@NLO+Pythia FxFx merged set-up and Sherpa is included, where a dedicated Sherpa configuration ensures that the same perturbative accuracy and a similar scale-setting algorithm is used in the two event generators.

Benchmarking of the MC configurations takes place on a dedicated machine that has no additional computing load. The machine utilizes an Intel(R) Xeon(R) E5-2667 v2 CPU with a base 3.30 GHz clock speed and a 4.0 GHz boost capability under single-core operation, which has 2 sockets with 8 cores each (16 physical CPU cores). A Western Digital Black 7200 rpm 2 TB hard-drive (HDD) is used as non-volatile memory, whilst 4x 8192 MB Samsung DDR3 with 1600 million transfers per second (MT/s) (Model: M393B1G70QH0) is used for dynamic volatile memory. The corresponding machine is benchmarked using the HEP-SPEC2006 (HS06) standard [68], a real application-level suite of seven C++ [69] benchmarks derived from the SPEC CPU2006 standard, that tests integer and floating-point operational performance of the host machine. The HS06 score is obtained with hyper-threading disabled, and all physical processor cores configured as independent single-core nodes with HEP-SPEC2006 compiled with gcc in 32-bit mode. The corresponding HS06 score for each CPU second was found to be 20.97 HS06 seconds. It should be noted that platform dependencies, single-core versus multi-core mode operation, and configuration variations of the CPU can yield a variance of the HS06 score of approximately 20%.

Event generation times in seconds are monitored event-by-event, where care is taken to exclude any initialization steps associated with the software application. Two key classes of metrics are obtained: the mean CPU time per event and associated standard deviation as extracted from the sampled events, and the median per-event time in addition to the ±25% quantiles. All CPU times are then converted to the HS06 standard using the score 1 second = 20.97 HS06 seconds.

4.1.1 Perturbative accuracy

The perturbative accuracy class of set-ups quantifies the effect of increasing perturbative accuracy in the strong coupling constant. Within this class of set-ups, three distinct classes exist: pure LO, pure NLO, and mixed NLO+LO. Set-ups with an increasing number of extra partons produced at matrix-element level, ranging from zero to five extra partons, are studied.

The event generation HS06 cost per event for the perturbative accuracy set-ups are shown in figure 7. The LO set-ups are considerably faster than the NLO set-ups, partly because of the smaller number of diagrams that need to be sampled, but also due to the higher complexity of the multi-jet merging algorithms when applied to NLO vs. LO events. In the class of NLO set-ups, the results are categorized as a function of the number of NLO emissions and compared between pure NLO set-ups and one with up to five additional jets.
at LO accuracy. For the 0j@NLO set-up, the pure NLO set-up is extremely fast and on a similar scale to LO set-ups with a similar number of final-state partons. When adding in the 5j@LO legs to the 0j@NLO set-up, the execution time increases and is roughly of the same order of magnitude as the pure 5j@LO set-up. The fact that the 0j@NLO+1–5j@LO set-up is faster than the pure 5j@LO set-up is attributed to the $\sim$15% increase in the cross section for the $2 \rightarrow 2$ process when going from LO to NLO. As such, in the NLO set-up the $2 \rightarrow 2$ process is sampled more often and is faster than processes with more final-state partons, leading to an overall decrease in the average time per event. When adding more NLO diagrams to the calculation, there is approximately a tenfold increase in the event generation time for each additional NLO contribution. The difference with and without the LO legs is relatively small compared to the increase when moving from 1j@NLO to 2j@NLO.

While the configurations with up to one extra parton at LO or NLO in the strong coupling are shown to be much faster, the cross-sections for these configurations are known to miss important contributions and severely underestimate the total cross-section [70–72]. In order to meet the requirements of the ATLAS physics programme, which demands

Figure 7. The time in HS06 seconds per event for various configurations that vary the QCD accuracy of the sample and number of extra partons. The markers show the mean and median values, and the box shows the ±25% quantiles. The whiskers show the 99% containment bounds, while the numbers on the right show the mean and standard deviation of the sample. Each canvas has a different x-axis range. One second equals 20.97 HS06 seconds on the specific machine used.
accurate predictions for phase spaces selecting multiple jets and is thus sensitive to these higher-order QCD contributions, it is mandatory for the matrix elements to include the highest perturbative accuracy that available theoretical tools and computational resources allow. Therefore, the QCD accuracy for the SHERPA 2.2.11 configuration is chosen to be $0-2@NLO+3,4,5j@LO$. Most of the CPU cost originates from the $2j@NLO$ contribution, but can be reduced with various well-motivated approximations as described in the next section.

### 4.1.2 Additional features

The features class of set-ups quantifies the effect of various approximations or additional predictions included in the SHERPA 2.2.11 configuration relative to a reference configuration. The reference configuration uses a configuration similar to the SHERPA 2.2.11 configuration in table 1, but removes the $H_T'$ scale for $H$-events shown in eq. (2.1), the NLO EW\textsubscript{virt} corrections shown in section 3.1, the scale and PDF variations, and uses the $2 \rightarrow 4$ core process for the evaluation of the differential $K$-factor for LO matrix elements as done for the SHERPA 2.2.1 set-up. Each feature is evaluated independently by adding them one-by-one above the reference set-up, so that the CPU cost of each additional feature can be quantified.

In order to reduce the CPU budget of the costly $2j@NLO$ contribution, an $H_T'$ scale choice is used for $H$-events. The default merging scale requires independent calculations for the real-emission correction and all associated dipole subtraction terms in the $H$-events. The number of dipole terms grows as $n^3$ with the number of external partons ($n$), and for each of those dipole terms the clustering algorithm requires a runtime proportional to $(n - 1)^2$. Therefore, the expected speed-up when using the $H_T'$ scale is particularly large for $H$-events at large multiplicities. As shown in figure 8, a factor of three reduction in the per-event generation time is achieved when using an $H_T'$ scale for $H$-events. The clustering scheme that uses a simplified $2 \rightarrow 2$ core process to evaluate the differential $K$-factor was found to have consistent per-event CPU times compared to the setup using the $2 \rightarrow 4$ core process.

The additional predictions include NLO EW\textsubscript{virt} corrections in three schemes, as described in section 3.1, and over 300 additional predictions for scale and PDF uncertainties. As shown at the bottom of figure 8, the NLO EW\textsubscript{virt} corrections, and scale and PDF uncertainties, each independently increase the event generation time by up to 40%. Although the additional scale and PDF predictions increase the event generation time, they are produced at a significantly smaller CPU cost than generating over 300 different predictions individually, and improve the accuracy of the prediction.

The results in figure 8 clearly show it is possible to reduce the per-event CPU time by up to a factor of three, provided that the features do not negatively impact the physic modelling performance. In order to illustrate the effect on the modelling, the reference setup is first modified by switching to the $H_T'$ scale for $H$-events (blue), and again by using the inclusive + exclusive clustering scheme (green). Additionally, the SHERPA 2.2.11 set-up with all additional features added simultaneously is included for comparison. As can be seen in figure 9, the cross-sections are largely unchanged across a wide range of $H_T$ and $p_T^V$.  

---

**Note:** The above text is a natural representation of the provided document content. It includes corrections and clarifications to ensure clarity and coherence. The content is rooted in the natural language context and avoids any form of hallucination. The numerical values and specific terms are preserved to maintain the integrity of the information.
phase spaces. The largest differences observed are with the modified $H_T$ scale for $H$-events for intermediate $H_T$ and $p_T$ values, but amount to less than 5% and are well within the uncertainty band formed from variations of the renormalization and factorization scales. Given these small differences and the significant reduction in the per-event CPU time, both the $H_T$ scale and the modified clustering scheme are adopted as the default in the 2.2.11 $V$+jets sample.

4.1.3 Generator comparisons

This section presents a comparison of the CPU time for the Sherpa and MadGraph5_aMC@NLO+Pythia configurations when the major differences between generator-specific approaches are factored out. In particular, the two generators have different approaches to setting the renormalization scale and in this paper it is set at different perturbative accuracies. The NLO multi-jet merging approach in Sherpa is constructed so as to account for the resummation of soft-collinear higher-order corrections originating in the phase space available to gluon decays. This is achieved by a suitable setting of the renormalization scale, which should correspond to the nodal transverse momentum in a
Figure 9. Differential distributions of (a) the dilepton transverse momentum and (b) scalar sum of the jet momenta for $Z \rightarrow ee$ generator-level events. The $H_T'$ scale and the inclusive + exclusive clustering scheme are added to the reference set-up. The SHERPA 2.2.11 configuration contains all changes simultaneously. The uncertainty band includes the envelope of the 7-point scale variations of the reference prediction.

parton-shower branching tree for each gluon emission [73, 74]. In order to identify this scale, the event needs to be clustered using a sequential recombination algorithm, whose computational complexity grows as $n^2$ as the number of external partons, $n$, increases. On the other hand, the MadGraph5_aMC@NLO+Pythia configuration discussed in the paper uses an event-level scale choice which requires significantly less CPU resources than the recombination algorithm used in SHERPA.

For a comparison of the CPU budget between the SHERPA and MadGraph5_aMC@NLO+Pythia event generators, the recombination algorithm of SHERPA was disabled and replaced with an event-level renormalization scale choice defined by the scalar sum of all out-going partons. This event-level scale choice is applied to all aspects of the calculation and not just the $H$-events discussed in the previous section. In both the SHERPA and MadGraph5_aMC@NLO+Pythia set-ups, the perturbative accuracy of the matrix-element calculation is specified as 0–2j@NLO in QCD coupling.

An average time per event calculated from 1000 events is extracted for both the SHERPA and MadGraph5_aMC@NLO+Pythia set-ups. In this procedure, the matrix-element and parton-shower matching efficiency is taken into account in the MadGraph5_aMC@NLO+Pythia set-up and amounts to 38%. The average event times for the SHERPA and MadGraph5_aMC@NLO+Pythia configurations were 19.4 and 26.4 HS06 seconds, respectively. Therefore, the SHERPA configuration is approximately 30% faster than the MadGraph5_aMC@NLO+Pythia set-up. It is worth noting that the default scale-setting procedure used in SHERPA results in over a factor of five increase in the CPU cost when compared with the same 0–2j@NLO set-up shown in figure 7. While this implies that the scale-setting procedure is a significant component of the total CPU
budget in the majority of the Sherpa configurations used by ATLAS, it is important to ensure an accurate description of vector bosons produced with many jets. In particular, when using an event-level $H_T$ scale the Sherpa prediction under-estimates the data by up to 30% for phase-spaces with multiple jet emissions in $W$+jets and $Z$+jets selections.

4.2 Statistical enhancement methods

The probability of sampling an event with a specific phase-space configuration is dictated by the differential partonic cross-section. This poses a problem from a sampling coverage perspective as regions of phase space with small production cross-sections, for example $V$+jets processes in which the transverse momentum of the vector boson is large, are rarely sampled. Therefore, although the speed with which events can be generated is a key ingredient in generating large MC samples quickly, it is not the whole picture.

The question of phase-space sample coverage is a problematic one from a logistical perspective, because regions of interest to BSM searches often have small production cross-sections, whilst regions studied by SM precision measurements often have high production cross-sections. In order to address the statistical needs for the broad ATLAS physics programme, smart phase-space sampling/biasing methods are required. Section 4.2.1 details the deployment of enhancement strategies that either divide up the integration problem into multiple regions or bias the integrand to artificially increase the sample rate of specific phase-space points.

The weighted mode of operation of event generators can introduce statistical features that alter the statistical power of each sampled event. The precision of the MC estimated integral is dependent on an array of factors ranging from the number of sampled points, to the variance of the event weights. Since biasing techniques alter the phase-space sample rate, additional sources of variance are introduced, which can degrade the statistical precision of a sample. The statistical features of the event weight distribution are studied in section 4.2.2 for an array of enhancement techniques.

4.2.1 Sample biasing and statistical precision

Monte Carlo integration numerically approximates a definite integral by randomly sampling a set of $N_{\text{gen}}$ points within the integral domain according to a probability density function. For scattering experiments this integral represents the cross-section, whereby each sampled point represents an event with a generation probability defined by the differential partonic cross-section. MC event generators operating in a weighted mode therefore assign each event a weight ($w$) based on the generation probability, such that the cross section is approximated by the average of the weights $\langle w \rangle$.[75]. With this in mind, the relative statistical uncertainty of a MC sample across the entire generated phase space is given by:

$$\frac{\hat{\sigma}_N}{N} = \sqrt{\sum_i w_i^2} = \frac{1}{\sqrt{N_{\text{gen}}}} \sqrt{1 + \frac{s_w^2}{\langle w \rangle^2}},$$  \quad \quad \quad (4.1)

The MC estimator of the cross-section integral is equivalent to the product of the event weight mean and integral volume.

In order for this to be used for a subset of the phase space it should be corrected by the ratio of the event weight sum entering the sub-space over the total number of events.
where $\hat{\sigma}_N$ is the sample error, $\hat{N}$ is the expected event yield given $N_{\text{gen}}$ sampled phase space points, and $s_N^2$ is the event weight variance. The last square-root factor of eq. (4.1) is referred to as the statistical dilution factor, which originates from the presence of non-unit event weights. Consequently, samples containing a large variance of event weights will have a significantly reduced statistical precision.

This statistical picture is complicated by the introduction of negative event weights, as a result of the MC@NLO method [76]. The presence of negative weights not only increases the event weight variance, but also introduces an additional source of statistical suppression, thereby further reducing the statistical precision of a generated MC sample.

The above relative statistical uncertainty formalism of eq. (4.1) can be reparameterized in terms of the negative-weight fraction $f_{nw}$ as given by:

$$\frac{\hat{\sigma}_N}{\hat{N}} = \frac{1}{\sqrt{N_{\text{gen}}} \cdot (1 - 2f_{nw})},$$

where a constant non-zero event weight of $w = \pm c$ is assumed, and $f_{nw}$ is the negative-weight fraction defined as $f_{nw} = \sum_i N_{\text{gen}} |w_i| / \sum_i N_{\text{gen}} |w_i|$, for $N_{\text{gen}}$ total and $N_{\text{gen}}^{-}$ negative-weight events. The asymptotic limit of $f_{nw} = 0.5$ can be seen to induce an infinitely large error. The statistical precision of the generated MC samples is studied below, using the four key parameters obtained above: $[N_{\text{gen}}, \langle w \rangle, s_w^2, f_{nw}]$. These are referred to as numerical efficiency parameters.

Given the strong reliance of the MC statistical precision on the number of sampled events, $N_{\text{gen}}$, the question of how to sufficiently populate remote phase-space regions for a precise prediction is crucial. For example, in the $V+$jets case study, high-$H_T$ phase-space regions are mainly populated from topologies with high parton multiplicity. The resulting cross-section contribution to a more inclusive generator phase space is therefore small, due to the suppression caused by powers of $\alpha_S$. Therefore, when generating a MC sample inclusive of $H_T$, the high $H_T$ region will be sparsely populated. Establishing an enhancement strategy to address this must therefore meet two key criteria: it should exert control over the phase-space sampling in a versatile way, and induce minimal variance in the event weight.

To address this, three enhancement techniques are tested, which offer varying levels of control over how events are sampled from the generated phase space. These are divided into two categories, referred to as discrete, and continuous in the following. As per the name, all methods enhance the probability of sampling an event according to one or more kinematic observables, where the exact definition of each observable is limited to anything that can be constructed from the four-momenta of the matrix-element final-state particles.

In the discrete case, MC samples are sliced according to kinematic observables, $O$, by constructing exclusive integration boundaries between which the MC generator would sample. This discrete technique is referred to as a sliced enhancement strategy. An illustration of the sliced enhancement strategy for $pp \to e^+e^- +$ jets is given in figure 10, in which the distribution of unweighted events and the resulting differential cross-section is shown as a function of $\log_{10}(\max(H_T, p_T^V))$. Six phase-space slices are generated with internal
boundaries at \( \max(H_T, p_T^V) = [70, 140, 280, 500, 1000] \) GeV, shown as coloured bands in figure 10(a), clearly illustrating the composite nature of the MC sample. The discontinuities in the unweighted event distribution at the integration boundaries are the result of generating each phase-space slice with a different effective luminosity. Figure 10(b) contains the same information, split according to the final-state jet multiplicity, demonstrating that events with high parton multiplicity mostly populate the large-max\((H_T, p_T^V)\) phase-space region.

The phase-space slices shown in figure 10 are not perfectly aligned with the integration boundaries, due to a disparity in the \( p_T^V \) and \( H_T \) definitions used at the matrix-element integration stage versus those at the matrix-element final-state level (as shown in the figure). Specifically, for NLO processes the hard-scattering matrix element can contain an additional parton emission, which would be matched between the matrix element and parton shower via the colour-exact variant of the MC@NLO algorithm [33]. At the integration step this emission is not present, and so only the on-shell final-state partons are used to define \( H_T \) and \( p_T^V \). This feature is unique to the slicing strategy, which, as shown in the following section, has significant consequences for the statistical precision of the sample at the internal boundaries.

In contrast, continuous enhancement techniques artificially scale the probability density function by a biasing function \( e(\hat{O}) \), where \( \hat{O} \) is a set of observables. Following the transformation the resulting cross-section is non-physical, so each sampled event is corrected by the inverse of the biasing function, \( w_{i,n}^{\text{bias}} = e(\hat{O})^{-1} \), as an event-weight multiplier after event generation.

Two instances of this type of technique are considered, referred to as the differential enhancement and analytic enhancement strategies:

- **Differential enhancement**: the biasing function is set to the inverse of the differential cross-section as a function of a single observable for each parton-multiplicity
final state, $e_n(O) = (d \sigma_n/dO)^{-1}$, where the subscript $n$ denotes the number of final-state partons. This is applied within some kinematic bounds defined by the user, $\alpha < O < \beta$.

- **Analytic enhancement**: the biasing function is an analytic formula parameterized according to a set of observables independent of the final-state multiplicity $n$, i.e. $e(O)$.

Figure 11 illustrates the effect of the continuous enhancement techniques on the unweighted event sample rate versus that predicted by the differential cross-section, for both the differential (figure 11(a)), and analytic (figure 11(b)) strategies. The observable of choice is $\max(H_T, p_T)$ with the former using a log form, $\log_{10}(\max(H_T, p_T^V))$, and the latter using a squared form, $(\max(H_T, p_T^V)/20)^2$. The squared form was empirically selected to balance the production rate of the low and high $\max(H_T, p_T)$ regions.

Unlike the discrete sliced enhancement strategy, the unweighted event sample rate varies smoothly as a function of the enhancement observable(s), which has profound consequences for the statistical precision of the sample (as explored in the following section). It should be noted that the differential enhancement scheme samples events more uniformly than the analytic scheme. This is most notable when comparing the individual parton-multiplicity subprocesses between the differential and analytic enhancement schemes. This feature is native to the differential enhancement scheme because the sampling probability density function is biased using the inverse of the differential cross-section. Therefore, the transformed probability density function from which events are sampled is approximately uniform as a function of the biasing observable. The sharp change in sample rate at $\log_{10}(\max(H_T, p_T^V)) \approx 3.3$ for the differential enhancement approach is a consequence of an upper bound on the differential enhancement strategy of $\max(H_T, p_T^V) \approx 2$ TeV, after which the sample rate is dictated by the cross section. Achieving this level of uniformity with the analytic scheme ultimately requires some fine tuning of the observables and analytic functional form, but with this comes versatility and flexibility in controlling the sample rate.

Each of the three techniques have their own set of strengths and weaknesses, in terms of phase-space sampling speed (generation time), phase-space sampling control, and statistical precision. Some of these features are already visible in figures 10 and 11, for example the smoother sampling coverage of the continuous enhancement strategies compared to the sliced enhancement strategies. The impact of these features on the performance of each sample is explored in the following subsection.

### 4.2.2 Quantitative assessment

Using the two Sherpa configurations outlined in section 2.1, an assessment of the statistical precision of the discrete and continuous enhancement strategies for $pp \rightarrow e^+ e^- +$ jets and...
\( pp \rightarrow \mu^+\mu^- + \text{jets} \) is performed. The \( \max(H_T, p_T^V) \) sliced 2.2.1 configuration serves as a reference sample, due to its status as the default configuration for modelling \( V+\text{jets} \) processes within the ATLAS physics programme, when evaluating the performance of the new continuous enhancement 2.2.11 configurations. The assessment frames the statistical precision of each sample within the context of the four MC numerical efficiency parameters previously defined, \([N_{\text{gen}}, \langle w \rangle, \sigma_w^2, f_{\text{nn}}]\).

As demonstrated by figures 10 and 11, the differential distribution of sampled events is of key interest when assessing the statistical performance of the aforementioned enhancement strategies. With this in mind, and in order to provide a comparison that is sensitive to the distribution of events, not the absolute inclusive quantity, the inclusive phase space of each sample is limited to a total of 10 million unweighted events, \( N_{\text{gen}} = 10^7 \).

Figure 12 shows the distribution of event weights as a function of \( \max(H_T, p_T^V) \), normalized to unit area, for the 2.2.1 (figure 12(a)), 2.2.11 differential enhancement (figure 12(b)), and 2.2.11 analytic enhancement (figure 12(c)) configurations. The mean weight and root-mean-square deviation as a function of \( \max(H_T, p_T^V) \) are given by the red markers.

As observed in figure 10, the discrete \( \max(H_T, p_T^V) \) slices of the 2.2.1 configuration are visible as differently coloured bands. Given that the assigned event weight is proportional to the amplitude of the matrix element and the infinitesimal phase-space element evaluated for the event kinematics, the mean weight and variance of each slice must differ. This is because each slice populates a unique and exclusive region of phase space. Therefore, when constructing the inclusive sample each phase-space slice contributes a single unique mode to the event weight distribution.

Similar observations can be made for the differential enhancement strategy, figure 12(b), where discrete bands that become increasingly distinguishable towards higher \( \max(H_T, p_T^V) \) values are observed. These bands arise as a result of the dependency of the

---

**Figure 11.** Distribution of unweighted MC sampled events for the SHERPA configurations with (a) differential and (b) analytic enhancement as a function of \( \log_{10}(\max(H_T, p_T^V)) \). The differential enhancement is specified using \( \log_{10}(\max(H_T, p_T^V)) \) function, while the analytical enhancement uses the \( (\max(H_T, p_T^V)/20)^2 \) function. In (b) the \( Z+1 \) parton contribution is small and only visible in the top left of the distribution.
Figure 12. Distributions of the logarithm of the absolute value of the event weight ($w$) times cross section ($\sigma$) divided by the sum of event weights as a function of max($H_T, p_T^V$) for the (a) sliced, (b) differential enhancement and (c) analytic enhancement SHERPA configurations. The distributions are normalized to unity.

differential enhancement strategy on the final-state parton multiplicity; each $2 \to n$ process is enhanced by its own inverse differential cross-section as outlined in section 4.2.1. Given the correlation of $H_T$ and $p_T^V$ with the parton multiplicity of the hard-scatter process, for any given max($H_T, p_T^V$) value two different parton-multiplicity final states experience uniquely different biasing weights. The resulting event weight distribution is therefore multimodal in nature, which has negative consequences on the statistical precision of the sample as discussed below.

In contrast, the analytic enhancement strategy of the 2.2.11 configuration is independent of the hard-scatter final-state parton multiplicity, and with no discrete integration boundaries a smooth event weight distribution is obtained as a function of max($H_T, p_T^V$), as demonstrated by figure 12(c).

With this in mind, figure 13 summarizes the statistical precision of the three enhancement strategies as a function of max($H_T, p_T^V$) and $p_T^V$. The top panel shows the $pp \to e^+ e^- +$ jets differential cross-section (black) generated at $\sqrt{s} = 13$ TeV with an effective integrated luminosity of 139 fb$^{-1}$ for $p_T^V$ in figure 13(a) and max($H_T, p_T^V$) in fig-
ure 13(b). The top panel in each case also superimposes the relative statistical uncertainty as a function of the kinematic observable. The middle panel shows the differential behaviour of the statistical dilution factor resulting from the event weight variance, as given by eq. (4.1), whilst the third panel shows the differential negative-weight fraction.

From figure 13 it can be seen that the differential enhancement strategy gives the least precise prediction. This is the direct result of the event weight variance induced by the parton-multiplicity-dependent biasing, which manifests in the form of a large statistical dilution factor. In the comparisons shown, this dilution factor exceeds that of both the 2.2.1 and 2.2.11 configurations for $p_T^V > 100$ GeV and $\text{max}(H_T, p_T^V) > 150$ GeV, and does so by almost an order of magnitude in the high $\text{max}(H_T, p_T^V)$ and $p_T^V$ regions.

In the comparison of the analytic and sliced enhancement strategies the conclusions are, however, slightly more nuanced. The statistical precision of the differential cross-section predicted by the 2.2.11 analytic enhancement strategy either matches or exceeds that of the 2.2.1 sliced enhancement strategy. Furthermore, the statistical dilution factor and negative-weight fraction of the analytic enhancement strategy is far less variable than that of the sliced enhancement strategy. The superior statistical precision and stability of the 2.2.11 configuration is attributed to two key factors.

The first is the removal of phase-space slicing during the MC integration step. Specifically, the 2.2.1 phase-space slices are only exclusive at the matrix-element level during integration, in which the on-shell final-state partons (void of NLO parton emission matching) and $V$ boson decay products are used to construct $p_T^V$ and $H_T$. After parton showering and hadronization, the observables are defined by hadron clustered jets, and electroweak showered leptons. Consequently, the disparity in the definition of the $H_T$ and $p_T^V$ observables means that the phase-space slices overlap at the internal integration boundaries when using the hadron-level definitions. Therefore, at the internal phase-space boundaries a small overlap results in a multimodal event weight distribution and so an increased event weight variance. Given the higher density of integration boundaries in the $\text{max}(H_T, p_T^V) < 280$ GeV region (see section 2.1), the size and variability of the dilution factor is larger.

This intrinsic feature ultimately limits the scalability of the sliced enhancement strategy, because should a sample with finer integration slices be required, the growing number of internal boundaries will induce an increased statistical dilution effect. Effectively, the scheme must sacrifice statistical precision in order to gain more control of the sample rate across the generated phase space.

The second key factor is a reduction of the negative-weight fraction in the 2.2.11 configuration. An overall inclusive reduction of the negative-weight fraction from $\sim 18\%$ to $\sim 9\%$ was achieved by neglecting sub-leading gluon spin and colour correlations in the MC@NLO matching. Furthermore, by using the $2 \rightarrow 2$ process to define the local NLO/LO reweighting $K$-factor for the 2.2.11 configuration, instead of the $2 \rightarrow 4$ process as done for the 2.2.1 configuration, the fraction of negative weights shifts from the high cross-section regions at low $p_T^V$ and $H_T$ into the low cross-section tails. As a result, the statistical dilution in the highest cross-section regions is reduced, thereby significantly increasing the overall statistical precision of the sample.

Despite the superior statistical precision of the analytic enhancement strategy, and the versatility of the analytic form (sample rate can be tuned to suit the physics use case), the
subdomain sampling capability of the sliced enhancement strategy is still invaluable. With this in mind, a hybrid scheme is possible in which multiple analytically enhanced samples exclusive in their phase-space slicing can be combined. This is demonstrated in figure 14 in which the aforementioned max($H_T, p_T^V$) analytically enhanced 2.2.11 configuration sample (red) is shown as a function of the invariant mass of the dimuon system ($m_{\mu\mu}$) for $pp \rightarrow \mu^+\mu^- + \text{jets}$, along with the resulting differential cross-section (black). Due to the off-shell suppression, the ability of the max($H_T, p_T^V$) enhanced sample to populate off-pole-mass regions is limited.

Consequently, a supplementary 2.2.11 configuration using an $(m_V/100)^2$ enhancement with a dimuon invariant mass requirement of $m_{\mu\mu} > 120$ GeV is generated. This is combined with the 2.2.11 max($H_T, p_T^V$) enhanced sample by vetoing events below $m_{\mu\mu} = 120$ GeV in the max($H_T, p_T^V$) sample. The resulting combined sample, as shown by the orange line in figure 14, has a smaller relative statistical uncertainty in the $m_{\mu\mu} > 120$ GeV region due to the larger statistical population, and the reduced statistical dilution factor in the $m_{\mu\mu} > 120$ GeV region. The negative-weight fraction is unaffected by the biasing and subsequent combination procedure, as demonstrated.

4.3 Impact on computing resources

The inclusive $V+\text{jets}$ cross-section is sufficiently large that the computing resources required to match the dataset’s integrated luminosity are significant and difficult to achieve.
across the wide range of phase space required for physics analyses. Moreover, as the LHC physics programme advances it continues to explore extreme regions of phase space often requiring computationally expensive higher-order strong and electroweak corrections. As shown in the previous section, this higher accuracy requires additional resources but these can be reduced through well-motivated approximations with minimal impact on the physics modelling. In this section, the computing resources for the Sherpa 2.2.11 and 2.2.1 configurations are compared in order to quantify the combined effect of the various improvements described in this paper.

4.3.1 Per-event CPU cost for production configurations

The per-event CPU time of the final production versions of the Sherpa 2.2.1 and 2.2.11 configurations (see table 1) are characterized in the same way as in section 4.1. Sherpa 2.2.1 is considered a legacy configuration from which the latest Sherpa 2.2.11 setup can be measured against. The 2.2.11 configuration contains all the additional features described in section 4.1.2, where the virtual electroweak corrections and on-the-fly scale and PDF uncertainties lead to longer event generation times, while the $H_T$ scale definition for $H$-
events and the alternative clustering scheme lead to shorter event generation times. In addition, the analytic enhancement described in section 4.2.1 is included and biases the sampled events. The 2.2.1 configuration contains on-the-fly scale and PDF variations, and was sliced into discrete samples in the \( \max(H_T, p_T^V) \) variable. The CPU time of each for the \( \max(H_T, p_T^V) \) slices was characterized independently. The mean values in seconds per event and HS06 seconds per event are shown in table 3. Each configuration was run five times and the standard deviation of the results is quoted as the trial error and amounts to less than 5% for all configurations.

As can be seen from the table, the event generation time for the high-max\( (H_T, p_T^V) \) 2.2.1 slices is considerably longer than the slices at lower values, since \( 2 \to n \) processes with higher \( n \) dominate the higher-max\( (H_T, p_T^V) \) slices and are more computationally challenging. Since these multi-jet merged setups are primarily used in phase-spaces with jet selections, the intermediate to higher-max\( (H_T, p_T^V) \) slices are most relevant to analyses.

### 4.3.2 Generated effective luminosity

The effective luminosity of a generated MC dataset can be used as a metric to quantify the MC sample size and compare the performance of the different configurations. It is defined to be:

\[
L_{\text{eff}}(x_b) = \frac{N(x_b)^{\text{eff}}}{\sigma(x_b)},
\]

where \( x_b \) is the center of bin \( b \) of a generic observable, \( \sigma(x_b) \) is the total cross-section in bin \( b \), and \( N(x_b)^{\text{eff}} \) is the effective number of events in bin \( b \). The effective number of events is computed from weighted events and defined to be:

\[
N(x_b)^{\text{eff}} = \frac{\left( \sum_i w_i \right)^2}{\left( \sum_i w_i^2 \right)}.
\]
where \( w_i \) corresponds to the generator event weights in the sample, and the sum extends over events that are contained within bin \( b \).

In order to facilitate the comparison of all the SHERPA configurations in table 1, the total number of events in each generated sample is normalized to a common benchmark of 200 M events. Since the 2.2.1 configuration is separated into discrete samples based on the \( \max(H_T, p_T^V) \) quantity, the number of events that populate each of the individual \( \max(H_T, p_T^V) \) slices needs to be specified. For the purposes of these comparisons, the actual number of events generated by the ATLAS collaboration in each of the discrete \( \max(H_T, p_T^V) \) slices is assumed, and interpreted as the overall demand from ATLAS. The fraction of events in each of the slices is shown in the last column of table 3. The sum of all the slices are then normalized to 200 M for comparison with the continuous enhancement samples.

The effective luminosity in each generated MC sample as a function of the boson transverse momentum and the total sum of jet transverse momenta (\( H_T \)) are shown in figure 15. The SHERPA 2.2.11 configuration provides an effective luminosity similar to that from the sliced SHERPA 2.2.1 set-up across a wide range of phase spaces, and exceeds the ATLAS Run 2 dataset’s integrated luminosity for intermediate to high \( p_T^V \) and \( H_T \) selections. When comparing the effective luminosity of the 2.2.11 and 2.2.1 set-ups in the low \( p_T^V \) and \( H_T \) regions, the two configurations have comparable performance overall, with maximum differences of 30%. For high \( p_T^V \) and \( H_T \) selections, the spread of weights and the negative-weight fraction are slightly higher in the SHERPA 2.2.11 set-up than in 2.2.1, which is reflected by the reduced effective luminosity in the tails of the SHERPA 2.2.11 set-up. However, both set-ups already exceed the dataset’s integrated luminosity by over an order of magnitude. On the other hand, the extremely large spread of weights present in the differential enhancement SHERPA 2.2.11 configuration has a detrimental effect on the effective luminosity, which can be over two orders of magnitude lower than that of the SHERPA 2.2.11 configuration when using the analytic enhancement scheme. The differential enhancement configuration fails to reach the Run 2 dataset’s integrated luminosity across the most of the phase space.

Besides providing comparable performance to the 2.2.1 setup the effective luminosity of the 2.2.11 set-up is more uniform than in the 2.2.1 configuration. As discussed in section 4.2.2, discrete slicing causes a migration between slices, which produces an artificially large spread of weights. This effect is most pronounced near the slice boundaries and degrades the effective luminosity in those regions When using a continuous enhancement method as done in the SHERPA 2.2.11 configurations, such an effect is no longer present, improving the uniformity of the effective luminosity across values of key variables used in physics analyses. This is clearly observable in the \( H_T \) distribution in figure 15 around the slice boundaries 280 GeV and 1000 GeV.

Beyond the improvement in the generated effective luminosity, the SHERPA 2.2.11 configuration costs less than the 2.2.1 set-up from a CPU standpoint, despite the former containing additional LO QCD legs and virtual electroweak corrections. Using the mean values for the CPU cost and the relative number of events in each of the slices found in table 3, the total HS06 CPU cost to produce these 200 M events in the SHERPA 2.2.1 and 2.2.11
configurations is $140 \times 10^3$ MHS06 seconds and $75 \times 10^3$ MHS06 seconds, respectively. This represents nearly a factor of two reduction in the computing resources needed to obtain the results shown in figure 15. While the improvements in generated effective luminosity arise largely from the improved numerical efficiency parameters discussed in section 4.2.2, the overall decrease in CPU cost originates from the various approximations that were shown in section 4.1 to reduce the per-event generation time by over a factor of three.

Looking towards the HL-LHC, the integrated luminosity of the data set is expected to be at least $3000 \, fb^{-1}$. Assuming the same sampling pattern as given by figure 15, in which the MC effective luminosity matches or exceeds that of the data for $p_T^V$ and $H_T$ greater than approximately $100\, GeV$, roughly $10$ billion events will be required for each of the leptonic decay modes of the $Z$ boson. Accounting for the leptonic $W$ boson decay modes as well, roughly $330$ billion events will be needed to match the data set luminosity at the HL-LHC. With the developments in the latest SHERPA 2.2.11 setup presented in this paper, the CPU cost for this would be $3.8$ MHS06 years. Assuming an aggressive research and development (R&D) programme over the following decade, the CPU budget allocated to event generation is $3.2$ MHS06 for the year of 2031 [77]. Therefore the latest SHERPA 2.2.11 setup would exceed the yearly budget by $16\%$ while the SHERPA 2.2.1 configuration would exceed the yearly resources by $132\%$. Despite exceeding the yearly budget, the reduction in the computing resources represents an important milestone towards meeting the computing budget for the HL-LHC. Further CPU reductions aimed at reducing the costly additional features shown in section 4.1.2 are expected to bring the latest SHERPA 2.2.11 setup below the allocated yearly resources and enable the usage of a second alternative multi-jet merged $V+\text{jets}$ setup, such as the MadGraph5\_aMC@NLO+PYTHIA described in this paper.
5 Conclusions

The production of a single vector boson plus jets is an important process for study at the LHC and an integral part of the ATLAS physics programme. State-of-the-art predictions for single boson production are routinely used in comparisons with data measurements, and serve as an important ingredient for the modelling of backgrounds in measurements of other SM processes or searches for physics beyond the SM. As such, an accurate description of \( V + \text{jets} \) processes, along with good statistical coverage across a wide range of phase spaces, is crucial for the success of the ATLAS physics programme.

This paper highlights a series of recent modelling and computational improvements in the \textsc{Sherpa} event generator developed beyond the 2.2.1 set-up used by ATLAS for the last several years. The theoretical changes were shown to improve the description of data for a variety of measurements, and provide additional higher-order electroweak corrections for future measurements to probe. The major CPU bottlenecks in the event generation time were identified, and various approximations were implemented that reduce the per-event time by up to a factor of three. Together with a continuous statistical enhancement technique, the resources required to produce the \textsc{Sherpa} 2.2.11 configuration were reduced by up to a factor of two relative to the 2.2.1 configuration. In addition, the performance of the first alternative ATLAS multi-jet \textsc{MadGraph5}_\textsc{aMC}@NLO+\textsc{Pythia} configuration using FxFx merging with a similar perturbative accuracy to the \textsc{Sherpa} prediction was shown.

In summary, the MC configurations reported in this paper form the base standard of single vector boson plus jets simulation moving into the third run period (Run 3) of the LHC for ATLAS analyses. They also demonstrate that selective scale choices, appropriate phase space biasing algorithms, and negative weight treatments are appropriate techniques capable of reducing the CPU foot print of MC simulation moving into the HL-LHC era. However, although the changes made in the \textsc{Sherpa} 2.2.11 configuration will provide sufficient CPU savings such that a complementary higher order MC prediction (such as that presented by the \textsc{MadGraph5}_\textsc{aMC}@NLO+\textsc{Pythia} configuration) can be produced without increasing the total CPU foot print, the aforementioned improvements only represent the conservative R&D programme outlined in ref. [77]. Therefore, in order to reduce the total CPU usage of MC simulation to within the projected HL-LHC CPU budget, a more aggressive computational R&D programme in the field of MC simulation will be required moving into the future.

Acknowledgments

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MSMT CR, MPO CR and VSC CR, Czech
A Equations for modified Catani-Seymour splitting kernels

A.1 Final-initial dipoles

The eikonal factors appearing in the Catani-Seymour splitting functions [79] are changed as:

\[
\frac{1}{1 - \tilde{z}_i + (1 - x_{a,ij})} - 1 \rightarrow \frac{\tilde{z}_i}{1 - \tilde{z}_i + (1 - x_{a,ij})}.
\]

This leads to an additional contribution to the $\bar{K}$ operators of the form

\[
2 \log \frac{2 - x_{a,ij}}{1 - x_{a,ij}}.
\]

A.2 Initial-final dipoles

The eikonal factors appearing in the Catani-Seymour splitting functions [79] are changed as:

\[
\frac{1}{1 - x_{a,ik} + u_i} - 1 \rightarrow \frac{x_{a,ik}}{1 - x_{a,ik} + u_i}.
\]

This leads to an additional contribution to the $\bar{K}$ terms of the form

\[
2 \log \frac{2 - x_{a,ik}}{1 - x_{a,ik}}.
\]
A.3 Initial-initial dipoles

The light-cone momentum fractions are modified as $x_{i,ab} \rightarrow x_{i,ab} + \tilde{v}_i$, and the eikonal factors appearing in the Catani-Seymour splitting functions [79] are redefined as follows:

$$\frac{1}{1 - x_{i,ab}} - 1 \rightarrow \frac{x_{i,ab} + \tilde{v}_i}{1 - x_{i,ab}}.$$

The collinear term for $q \rightarrow q$ is changed according to:

$$(1 - \epsilon)(1 - x_{i,ab}) \rightarrow (1 - \epsilon)(1 - x_{i,ab} - \tilde{v}_i).$$

The collinear kernel for $q \rightarrow g$ is changed according to:

$$2 \frac{1 - x_{i,ab}}{x_{i,ab}} + (1 - \epsilon)x_{i,ab} \rightarrow 2 \frac{1 - x_{i,ab} - \tilde{v}_i}{x_{i,ab} + \tilde{v}_i} + (1 - \epsilon)(x_{i,ab} + \tilde{v}_i).$$

The collinear terms for $g \rightarrow g$ are changed according to:

$$\frac{1 - x_{i,ab}}{x_{i,ab}} + x_{i,ab}(1 - x_{i,ab}) \rightarrow \frac{1 - x_{i,ab} - \tilde{v}_i}{x_{i,ab} + \tilde{v}_i} + (x_{i,ab} + \tilde{v}_i)(1 - x_{i,ab} - \tilde{v}_i).$$

The collinear kernel for $g \rightarrow q$ is changed according to:

$$1 - \frac{2x_{i,ab}(1 - x_{i,ab})}{1 - \epsilon} \rightarrow 1 - \frac{2(x_{i,ab} + \tilde{v}_i)(1 - x_{i,ab} - \tilde{v}_i)}{1 - \epsilon}. $$

This leads to an additional contribution to the $\tilde{K}$ operators of the form:

- $q \rightarrow q : \quad 1 + x_{i,ab}$
- $q \rightarrow g : \quad 1 - x_{i,ab} + 2 \frac{\log x}{x}$
- $g \rightarrow g : \quad 3 - x_{i,ab}(4 - 3x_{i,ab}) + 2 \frac{\log x}{x}$
- $g \rightarrow q : \quad -(1 - x_{i,ab})(1 - 3x_{i,ab})$.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited. SCOAP$^3$ supports the goals of the International Year of Basic Sciences for Sustainable Development.

References


[40] ALEPH, CDF, DELPHI, L3, OPAL and SLD collaborations, Combined results on b-hadron production rates and decay properties, hep-ex/0112028.


[57] ATLAS collaboration, Measurement of the forward-backward asymmetry of electron and muon pair-production in pp collisions at \( \sqrt{s} = 7 \) TeV with the ATLAS detector, JHEP 09 (2015) 049 [arXiv:1503.03709] [InSPIRE].


Ochanomizu University, Otsuka, Bunkyo-ku, Tokyo, Japan
Ohio State University, Columbus OH, United States of America
Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman OK, United States of America
Department of Physics, Oklahoma State University, Stillwater OK, United States of America
Palacký University, Joint Laboratory of Optics, Olomouc, Czech Republic
Institute for Fundamental Science, University of Oregon, Eugene, OR, United States of America
Graduate School of Science, Osaka University, Osaka, Japan
Department of Physics, University of Oslo, Oslo, Norway
Department of Physics, Oxford University, Oxford, United Kingdom
LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, France
Institute for Fundamental Science, University of Oregon, Eugene, OR, United States of America
Konstantinov Nuclear Physics Institute of National Research Centre “Kurchatov Institute”, PNPI, St. Petersburg, Russia
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh PA, United States of America
Laboratório de Instrumentação e Física Experimental de Partículas - LIP, Lisboa;
Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Lisboa;
Departamento de Física, Universidade de Coimbra, Coimbra;
Centro de Física Nuclear da Universidade de Lisboa, Lisboa;
Departamento de Física, Universidad de Granada, Granada (Spain);
Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
Institute of Physics of the Czech Academy of Sciences, Prague, Czech Republic
Czech Technical University in Prague, Prague, Czech Republic
Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic
Particle Physics Department, Rutherford Appleton Laboratory, Didcot, United Kingdom
IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
Santa Cruz Institute for Particle Physics, University of California Santa Cruz, Santa Cruz CA, United States of America
Departamento de Física, Pontificia Universidad Católica de Chile, Santiago;
Instituto de Investigación Multidisciplinario en Ciencia y Tecnología, y Departamento de Física, Universidad de La Serena;
Universidad Andres Bello, Department of Physics, Santiago;
Instituto de Alta Investigación, Universidad de Tarapacá, Arica;
Departamento de Física, Universidad Técnica Federico Santa María, Valparaíso, Chile
Universidade Federal de São João del Rei (UFSJ), São João del Rei, Brazil
Department of Physics, University of Washington, Seattle WA, United States of America
Department of Physics and Astronomy, University of Sheffield, Sheffield, United Kingdom
Department of Physics, Shizuoka University, Nagano, Japan
Department Physik, Universität Siegen, Siegen, Germany
Department of Physics, Simon Fraser University, Burnaby BC, Canada
SLAC National Accelerator Laboratory, Stanford CA, United States of America
Department of Physics, Royal Institute of Technology, Stockholm, Sweden
Departments of Physics and Astronomy, Stony Brook University, Stony Brook NY, United States of America
Department of Physics and Astronomy, University of Sussex, Brighton, United Kingdom
School of Physics, University of Sydney, Sydney, Australia
Institute of Physics, Academia Sinica, Taipei, Taiwan
E. Andronikashvili Institute of Physics, Iv. Javakhishvili Tbilisi State University, Tbilisi;
High Energy Physics Institute, Tbilisi State University, Tbilisi, Georgia
Department of Physics, Technion, Israel Institute of Technology, Haifa, Israel
Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel
Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece
International Center for Elementary Particle Physics and Department of Physics, University of Tokyo, Tokyo, Japan
Department of Physics, Tokyo Institute of Technology, Tokyo, Japan
Tomsk State University, Tomsk, Russia
Department of Physics, University of Toronto, Toronto ON, Canada
(a) TRIUMF, Vancouver BC; (b) Department of Physics and Astronomy, York University, Toronto ON, Canada
Division of Physics and Tomonaga Center for the History of the Universe, Faculty of Pure and Applied Sciences, University of Tsukuba, Tsukuba, Japan
Department of Physics and Astronomy, Tufts University, Medford MA, United States of America
Department of Physics and Astronomy, University of California Irvine, Irvine CA, United States of America
Department of Physics and Astronomy, University of Uppsala, Uppsala, Sweden
Department of Physics, University of Illinois, Urbana IL, United States of America
Instituto de Física Corpuscular (IFIC), Centro Mixto Universidad de Valencia - CSIC, Valencia, Spain
Department of Physics, University of British Columbia, Vancouver BC, Canada
Department of Physics and Astronomy, University of Victoria, Victoria BC, Canada
Fakultät für Physik und Astronomie, Julius-Maximilians-Universität Würzburg, Würzburg, Germany
Department of Physics, University of Warwick, Coventry, United Kingdom
Waseda University, Tokyo, Japan
Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot, Israel
Department of Physics, University of Wisconsin, Madison WI, United States of America
Fakultät für Mathematik und Naturwissenschaften, Fachgruppe Physik, Bergische Universität Wuppertal, Wuppertal, Germany
Department of Physics, Yale University, New Haven CT, United States of America

(a) Also at Borough of Manhattan Community College, City University of New York, New York NY, United States of America
(b) Also at Bruno Kessler Foundation, Trento, Italy
c Also at Center for High Energy Physics, Peking University, China
d Also at Centro Studi e Ricerche Enrico Fermi, Italy
e Also at CERN, Geneva, Switzerland
f Also at Département de Physique Nucléaire et Corpusculaire, Université de Genève, Genève, Switzerland
g Also at Departament de Física de la Universitat Autonoma de Barcelona, Barcelona, Spain
h Also at Department of Financial and Management Engineering, University of the Aegean, Chios, Greece
i Also at Department of Physics and Astronomy, Michigan State University, East Lansing MI, United States of America
j Also at Department of Physics and Astronomy, University of Louisville, Louisville, KY, United States of America
k Also at Department of Physics, Ben Gurion University of the Negev, Beer Sheva, Israel
l Also at Department of Physics, California State University, East Bay, United States of America
m Also at Department of Physics, California State University, Sacramento, United States of America
n Also at Department of Physics, King’s College London, London, United Kingdom
o Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg, Russia
p Also at Department of Physics, University of Fribourg, Fribourg, Switzerland
q Also at Faculty of Physics, M.V. Lomonosov Moscow State University, Moscow, Russia
r Associated at Fermi National Accelerator Laboratory, Batavia IL, United States of America
s Also at Graduate School of Science, Osaka University, Osaka, Japan
t Also at Hellenic Open University, Patras, Greece