Accelerating LHC event generation with simplified pilot runs and fast PDFs

Enrico Bothmann[1], **Andy Buckley**[2], **Ilektra A. Christidi**[3], **Christian Gütschow**[4], **Stefan Höche**[5], **Max Knobbe**[1,2], **Tim Martin**[6], **Marek Schönherr**[7]

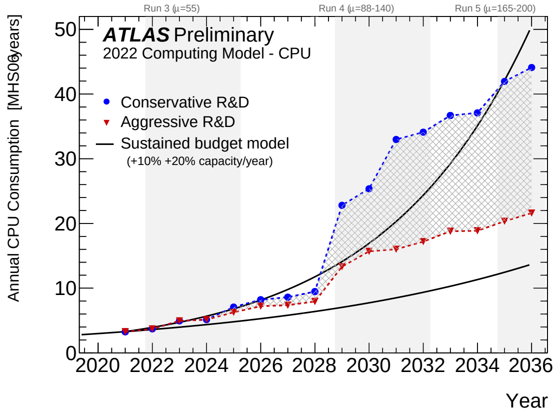
[1]Institut für Theoretische Physik, Georg-August-Universität Göttingen, 37077 Göttingen, Germany
[2]School of Physics & Astronomy, University of Glasgow, Glasgow, G12 8QQ, UK
[3]Centre for Advanced Research Computing, University College London, Gower Street, London, WC1E 6BT, UK
[4]Department of Physics and Astronomy, University College London, Gower Street, London, WC1E 6BT, UK
[5]Department of Physics, University of Warwick, Coventry, CV4 7AL, UK
[6]Theory Division, Fermi National Accelerator Laboratory, Batavia, IL, 60510, USA
[7]Institute for Particle Physics Phenomenology, Department of Physics, Durham University, Durham, DH1 3LE, UK

E-mail: `chris.g@cern.ch`

Abstract. High-precision calculations are an indispensable ingredient for the success of the LHC physics programme, yet their poor computing efficiency has been a growing cause for concern, threatening to become a paralysing bottleneck in the coming years. We present solutions to eliminate the apprehension by focussing on two major components of general-purpose Monte Carlo event generators: the evaluation of parton distribution functions, and the generation of perturbative matrix elements. We show that for the cost-driving event samples employed by the ATLAS experiment to model omnipresent, irreducible Standard Model backgrounds, such as weak boson or top-quark pair production in association with jets, these computational components dominate the overall run time by up to 80 %. We demonstrate that a reduction of the computing footprint of LHAPDF and SHERPA by factors of around 40 can be achieved for multi-leg NLO event generation.

1. Introduction

The main objective of the Large Hadron Collider (LHC) and its high-luminosity phase (HL-LHC) in the coming years will be to establish whether or not the Standard Model of particle physics describes Nature above the electroweak symmetry-breaking scale. To this end, Monte Carlo event generators are crucial, serving the dual purpose of providing realistic detector simulations as well as the final-state predictions to compare to the data. However, the poor computing efficiency of high-precision calculations is a growing concern for the LHC general-purpose experiments: All reasonable projections for the next decade exceed the expected budget (*see* Fig. 1), leaving the world's largest and most expensive experiment at risk of being statistically limited by the available Monte Carlo event samples.



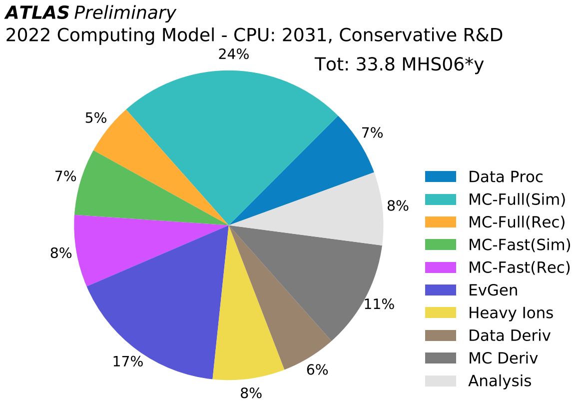


Figure 1. Estimated CPU requirements by the ATLAS experiment for the next decade, taken from [1]. The solid lines indicate a possible range for the available budget. Even the projection that assumes "aggressive R&D" to improve the computational footprint barely manages to stay within the optimistic budget forecast.

Figure 2. Breakdown of projected computing usage using the "Conservative R&D" model for Run-4, taken from [1]. Monte Carlo event generation ("EvgGen") is expected to take up 17% of the budget, closely following GEANT4-based detector simulation ("MC-Full(Sim)") with 24% of the total share.

In the case of the ATLAS experiment, Monte Carlo event generation is expected to take the second largest share of the projected CPU requirements for Run-4 (*see* Fig. 2), closely following GEANT4-based detector simulation which is expected to contribute a quarter to the projected CPU usage. The bulk of the event generation component is taken up by high-precision Standard Model calculations of standard candle processes, such as final states with weak boson or top-quark pair production in association with jets [2]. A systematic study of the CPU breakdown for these high-precision particle-level calculations, along with strategies aiming to reduce the computational footprint of these types of setups, has recently been discussed in [3], focussing on the concrete implementations of the LHAPDF [4] library for parton distribution function (PDF) evaluation as well as the SHERPA [5] event generator, which provides next-to-leading order (NLO) accurate matrix elements in the QCD and electroweak (EW) couplings.

Generated events are typically supplemented with multiple event weights reflecting the effects of electroweak corrections, and variations of scale, α_s and PDF-fit choices. These are computed using the on-the-fly reweighting techniques outlined in Refs. [6–8], including the re-evaluation of couplings (when varying the renormalisation scale) and PDFs (when varying the factorisation scale), of which the latter are particularly costly. While the impact of the performance improvements detailed in Section 2 of Ref. [3] has been assessed for a range of on-the-fly variation weights, showing that they lead to overall speed-ups of up to factors of 15 for representative setups, the exact configuration employed by the ATLAS experiment for their large scale event generation was not included in the original benchmarking exercise. In the following, we present a dedicated analysis of these configurations.

2. Technical setup

As test cases, we use setups for $pp \to e^+e^- + 0,1,2j$@NLO$+3,4,5j$@LO Drell-Yan production as well as $pp \to t\bar{t} + 0,1j$@NLO $+ 2,3j$@LO at 13 TeV at the LHC where the different multiplicities are merged using the MEPS@NLO algorithm [9–11]. Next-to-leading order precision in QCD is provided for up to two jets in the e^+e^- + jets setup and up to one jet in the $t\bar{t}$ + jets setup with the help of the OpenLoops library [12,13] for virtual corrections and an implementation of Catani-Seymour dipole subtraction in Amegic [14] and Comix [15]. The matching to a Catani-Seymour based parton shower [16] is performed using the S–Mc@Nlo technique [17, 18], an extension of the Mc@Nlo matching method [19] that implements colour and spin correlations in the first parton-shower emission, in order to reproduce the exact singularity structure of the hard matrix element.

A statistical over-representation of multijet events is achieved through biasing the unweighted event distribution. In the e^+e^-+jets case, the biasing is done in the maximum of the scale sum of all partonic jet transverse momenta (H_T) and the transverse momentum of the lepton pair (p_T^V). In the $t\bar{t}$ + jets case, the biasing is done in the maximum of the H_T and the average top-quark transverse momentum ($(p_\mathrm{T}^t + p_\mathrm{T}^{\bar{t}})/2$). Additionally, approximative EW corrections ($\mathrm{EW}_\mathrm{virt}$) are calculated which requires evaluation of the relatively expensive EW virtual corrections, and also the subleading Born corrections. These corrections are supplied as variation weights [6, 7] and are not applied to the nominal weight used for the unweighting. Furthermore, 7-point variations of the factorisation and renormalisation scales are included, which are calculated for both the matrix-element and the parton-shower parts of the event generation [8]. The events are also supplied with variation weights corresponding to the 100 Hessian eigenvectors of the nominal PDF set (`NNPDF30_nnlo_as_0118_hessian` [20]), along with the central values for the α_s variations `NNPDF30_nnlo_as_0117` and `NNPDF30_nnlo_as_0119`, the error set for the PDF4LHC15 set (`PDF4LHC15_nnlo_30_pdfas` [21]) as well as the central values for `MSHT20nnlo_as118` [22], `CT18NNLO_as_0118` [23] and `NNPDF31_nnlo_as_0118_hessian` [24]. As for the scale variations, these variation weights require a re-evaluation of the PDFs in both matrix elements and the parton shower.

3. Performance analysis

The impact of the performance improvements discussed in Ref. [3] is investigated in eight steps, with each step adding a new improvement as follows:

MePs@Nlo baseline
: The baseline setup, using the pre-improvement versions of Sherpa 2.2.11 and Lhapdf 6.2.3, *i.e.* using the CKKW scale setting procedure throughout as well as the standard S–Mc@Nlo matching technique. All one-loop corrections are provided by OpenLoops 2.1.2.

↳ Lhapdf 6.4.0
: The version of Lhapdf is increased to Lhapdf 6.4.0, implementing the improvements of Sec. 2 in Ref. [3].

↳ ⟨LC⟩–Mc@Nlo
: The full-colour spin-correlated S–Mc@Nlo algorithm is reduced to its leading-colour spin-averaged cousin, ⟨LC⟩–Mc@Nlo, which however is still applied before the unweighting. Note that this is the only step where a physics simplification occurs. For details *see* Sec. 3.1 of Ref. [3].

↳ pilot run
: The pilot run strategy of Sec. 3.2 in Ref. [3] is enabled, minimising the number of coefficients and variations needlessly computed for events that are going to be rejected in the unweighting step.

↳ ⟨LC⟩–Mc@Nlo–Css
: The ⟨LC⟩–Mc@Nlo matching is moved into the standard Css parton shower, *i.e.* it is now applied after the unweighting.

↳ Mcfm
: During the pilot run, the automatically generated one-loop QCD matrix elements provided

by OpenLoops are replaced by the manually highly optimised analytic expressions encoded in the Monte Carlo for FeMtobarn processes (Mcfm) [25–29]. Once the event is accepted, OpenLoops continues to provide all one-loop QCD and EW corrections, *see* Sec. 3.3 of Ref. [3].

↳ **pilot scale** Events are unweighted using a simple scale that depends solely on the kinematics of the final state and, thus, does not require a clustering procedure. The correct dependence on the actual factorisation and renormalisation scales determined through the CKKW algorithm is then restored through a residual event weight. For details *see* Sec. 3.4 of Ref. [3].

↳ **PPRS scale** A clustering-independent scale definition is used for the hard events in the $t\bar{t}$ + jets setup, as was already the case for the e^+e^- + jets setup. The impact of avoiding the clustering is larger for $t\bar{t}$+jets due to the different structure of the core process, which comprises four partons instead of two, along with a gluon-dominated initial state.

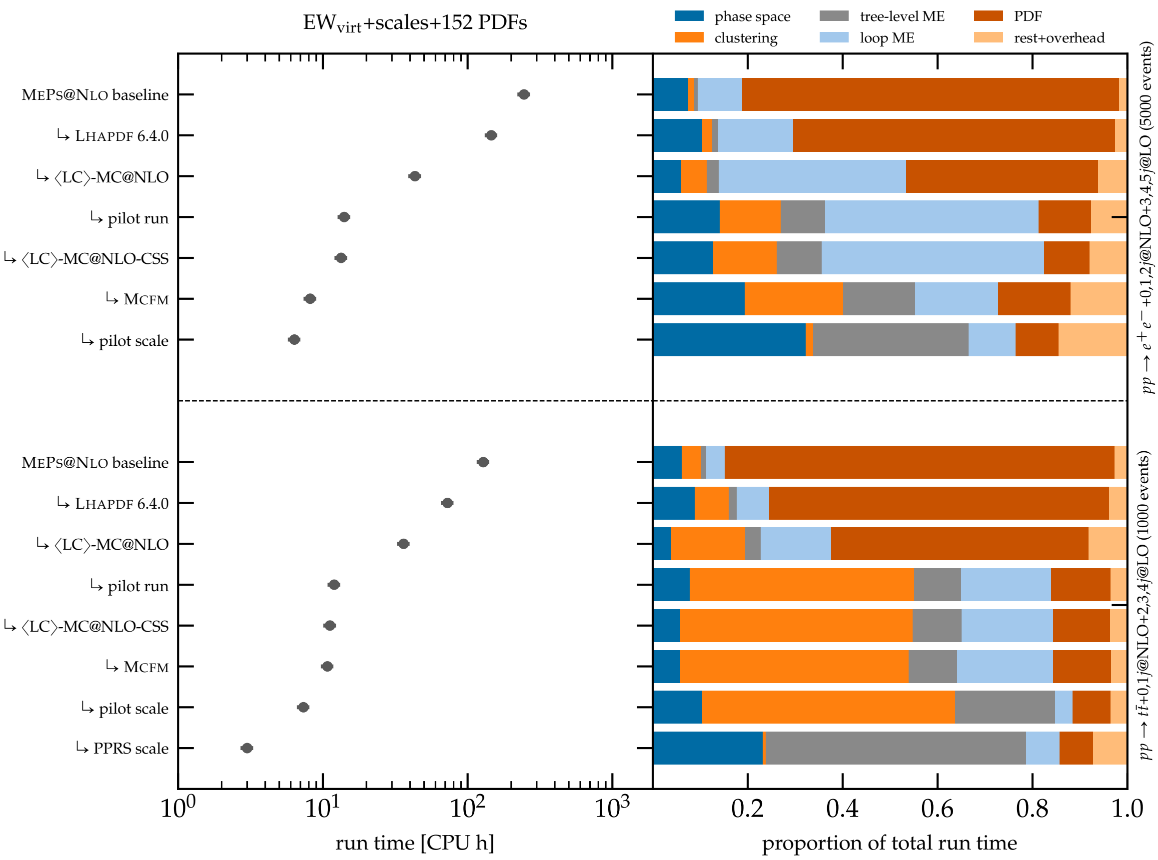


Figure 3. Overall run time after different performance improvements (left half), combined with the breakdown of the overall run time into a high-level calculation composition (right half). The timing is assessed by producing 5 000 particle-level events for $pp \to e^+e^- + 0, 1, 2j$@NLO + 3, 4, 5j@LO (upper half) and 1 000 particle-level events for $pp \to t\bar{t} + 0, 1j$@NLO + 2, 3, 4j@LO (lower half), both using MEPs@NLO.

Figure 3 shows the impact of each improvement on the total run time to generate 5 000 and 1 000 unweighted events for e^+e^-+jets and $t\bar{t}$+jets, respectively, on the left side. The composition of these run times for each of the steps is shown on the right side. For the total run times, horizontal error bars indicate a 10 % uncertainty estimate. Using LHAPDF 6.4 reduces the overall run time by about 40–50 %. Enabling the ⟨LC⟩–MC@NLO yields a speedup factor of 2–3. Moving the matched first shower emission into the normal CSS shower simulation, ⟨LC⟩–MC@NLO–CSS, result in an additional speed-up of 5–10 %.

When switching to use MCFM for the pre-unweighting loop calculations, another sizeable reduction in runtime by about 80 % is observed for the $e^+e^- +$ jets setup, whereas a negligible impact is seen for the $t\bar{t}+$jets setup, since only the $t\bar{t}$ process is currently implemented in MCFM but not the more costly $t\bar{t}j$ process, which is then provided by OPENLOOPS throughout.

Moreover, it can be seen that the large clustering component in the $t\bar{t}+$jets case is removed when calculating the hard events using a dedicated clustering-independent scale definition, similar to the $e^+e^- +$ jets setup. This improvement results in an additional speedup of a factor 2 for the ATLAS $t\bar{t}+$ jets setup.

The overall runtime of the ATLAS $e^+e^- +$ jets setup and the $t\bar{t}+$ jets setup is reduced by factors of 39 and 43, respectively. None of the components in the breakdown of the CPU composition for the best setups take up more than 40 % of the runtime a this point.

4. Summary

We presented a dedicated performance analysis of two major software packages used by the ATLAS experiment for large-scale event generation of state-of-the-art $pp \rightarrow e^+e^- + 0,1,2j$@NLO + $3,4,5j$@LO and $pp \rightarrow t\bar{t} + 0,1j$@NLO + $3,4j$@LO simulations at the LHC. We show that using the simple strategies to reduce the computational footprint discussed in [3], the overall runtime of these setups can be reduced by a factor of about 40, thereby achieving a major milestone set by the event generator working group of the HEP Software Foundation whilst paving the way towards affordable state-of-the-art event simulation in the HL-LHC era.

Acknowledgements

The authors would like to thank the Durham Performance Analysis Workshop series [30] which provided an important forum to gain the necessary insights during key phases of this work. This research was supported by the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE–AC02–07CH11359. The work of S.H. was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program. A.B. and M.S. are supported by the UK Science and Technology Facilities Council (STFC) Consolidated Grant programmes ST/S000887/1 and ST/T001011/1, respectively. A.B., I.C., C.G., and M.S. are further supported by the STFC SWIFT-HEP project (grant ST/V002627/1). E.B. and M.K. acknowledge support from BMBF (contract 05H21MGCAB) and funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 456104544. This work has received funding from the European Union's Horizon 2020 research and innovation programme as part of the Marie Skłodowska-Curie Innovative Training Network MCnetITN3 (grant agreement no. 722104). M.S. is supported by the Royal Society through a University Research Fellowship (URF\R1\180549) and an Enhancement Award (RGF\EA\181033, CEC19\100349, and RF\ERE\210397).

References

[1] ATLAS collaboration 2022 CERN-LHCC-2022-005
[2] ATLAS collaboration 2022 *JHEP* **08** 089 (*Preprint* 2112.09588)
[3] Bothmann E *et al.* 2022 *Eur. Phys. J. C* **82** 1128 (*Preprint* 2209.00843)
[4] Buckley A *et al.* 2015 *Eur. Phys. J.* **C75** 132 (*Preprint* 1412.7420)

[5] Bothmann E *et al.* 2019 *SciPost Phys.* **7** 034 (*Preprint* 1905.09127)
[6] Kallweit S *et al.* 2016 *JHEP* **04** 021 (*Preprint* 1511.08692)
[7] Bräuer S *et al.* 2020 *JHEP* **10** 159 (*Preprint* 2005.12128)
[8] Bothmann E *et al.* 2016 *Eur. Phys. J.* **C76** 590 (*Preprint* 1606.08753)
[9] Höche S *et al.* 2013 *JHEP* **04** 027 (*Preprint* 1207.5030)
[10] Gehrmann T *et al.* 2013 *JHEP* **01** 144 (*Preprint* 1207.5031)
[11] Höche S *et al.* 2014 *Phys.Rev.* **D89** 093015 (*Preprint* 1403.7516)
[12] Cascioli F *et al.* 2012 *Phys.Rev.Lett.* **108** 111601 (*Preprint* 1111.5206)
[13] Buccioni F *et al.* 2019 *Eur. Phys. J. C* **79** 866 (*Preprint* 1907.13071)
[14] Gleisberg T *et al.* 2008 *Eur. Phys. J.* **C53** 501–523 (*Preprint* 0709.2881)
[15] Gleisberg T and Hoeche S 2008 *JHEP* **12** 039
[16] Schumann S *et al.* 2008 *JHEP* **03** 038 (*Preprint* 0709.1027)
[17] Höche S *et al.* 2012 *JHEP* **09** 049 (*Preprint* 1111.1220)
[18] Höche S *et al.* 2012 *Phys. Rev.* **D86** 094042 (*Preprint* 1208.2815)
[19] Frixione S *et al.* 2002 *JHEP* **06** 029 (*Preprint* hep-ph/0204244)
[20] NNPDF collaboration 2015 *JHEP* **04** 040 (*Preprint* 1410.8849)
[21] Butterworth J *et al.* 2016 *J. Phys. G* **43** 023001 (*Preprint* 1510.03865)
[22] Bailey S *et al.* 2021 *Eur. Phys. J. C* **81** 341 (*Preprint* 2012.04684)
[23] Hou T J *et al.* 2021 *Phys. Rev. D* **103** 014013 (*Preprint* 1912.10053)
[24] NNPDF collaboration 2017 *Eur. Phys. J. C* **77** 663 (*Preprint* 1706.00428)
[25] Campbell J M *et al.* 1999 *Phys. Rev.* **D60** 113006 (*Preprint* hep-ph/9905386)
[26] Campbell J M *et al.* 2011 *JHEP* **07** 018 (*Preprint* 1105.0020)
[27] Campbell J M *et al.* 2015 (*Preprint* 1503.06182)
[28] Campbell J *et al.* 2019 *JHEP* **12** 034 (*Preprint* 1909.09117)
[29] Campbell J M *et al.* 2021 *Eur. Phys. J. C* **81** 1117 (*Preprint* 2107.04472)
[30] Basden A *et al.* 2021 URL https://doi.org/10.5281/zenodo.5155503