The Journal of Open Source Software

# pylhe: A Lightweight Python interface to Les Houches Event files

**Alexander Puck Neuwirth** [1,2¶], **Matthew Feickert** [3], **Lukas Heinrich** [4], **and Eduardo Rodrigues** [5]

**1** University of Milan Bicocca **2** INFN Milan Bicocca **3** University of Wisconsin-Madison **4** Technical University of Munich **5** University of Liverpool **¶** Corresponding author

**DOI:** 10.xxxxxx/draft

**Software**

- Review
- Repository
- Archive

**Editor:**

**Submitted:** 15 December 2025
**Published:** unpublished





## Summary

pylhe is a lightweight Python library that provides a simple and efficient interface for reading and writing Les Houches Event (LHE) files, a standard format used by Monte Carlo event generators in high-energy physics (Alwall & others, 2007). The library enables memory-efficient streaming of events from `.lhe` and compressed `.lhe.gz` files through a pythonic iterator interface, allowing researchers to process arbitrarily large event files without loading all events into memory simultaneously.

Historically, the first standards for event representation in high-energy physics were the HEPEVT and HEPRUP common blocks (Boos & others, 2001), which provided a Fortran-based structure for storing event information. As the complexity of Monte Carlo event generators increased, the need for a more flexible and extensible format led to the development of the LHE file format. LHE introduced an XML structure that allowed for better organization of event data and facilitated interoperability between different tools. Typically, LHE files are used to describe parton-level events generated by matrix element generators that are passed to parton shower and hadronization programs producing HepMC files (Dembinski et al., 2022; Dobbs & Hansen, 2001; Verbytskyi et al., 2020), or directly to analysis frameworks such as Rivet (Bierlich et al., 2024; Bierlich & others, 2020; Buckley et al., 2013). In contrast to LHE files, HepMC stores fewer details about the particles, but also contains many more particles per event. Both files are plain text-based formats, making them human-readable but potentially large. Consequently, compression using gzip is common practice for both formats. In the future a binary format such as HDF5 (The HDF Group, 1998) could be considered for event storage to improve read/write performance and reduce file sizes further.

The LHE format stores information within `<init>` and multiple `<event>` blocks consisting of whitespace-separated values designed historically for straightforward parsing in Fortran. Further details can be found in the original definition of the Les Houches Event file standard (Alwall & others, 2007). Following the original publication there were two extensions to the LHE format, version 2.0 in 2009 (Butterworth & others, 2010) and version 3.0 in 2013 (Andersen & others, 2014). However, pylhe currently only implements the widely adopted extension on top of version 1.0, that is the addition of multiple weights via `<initrwgt>`, `<rwgt>`, `<weight>`, `<weights>`, `<wgt>`, and `<weightgroup>`. In the future, if there is a demand for `<scales>`, `<generator>`, `<pdfinfo>`, or `<clustering>`, support for these can be added as well.

## Statement of need

The LHE format is used by all major Monte Carlo event generators such as MadGraph (Alwall et al., 2014), POWHEG-BOX (Alioli et al., 2010; Frixione et al., 2007; Nason, 2004), Sherpa (Bothmann & others, 2019; Gleisberg et al., 2009), HERWIG (Bahr & others, 2008; Bellm &

others, 2016, 2020; Bewick & others, 2024; Corcella et al., 2001), Pythia (Bierlich & others, 2022; Sjostrand et al., 2006, 2008; Sjöstrand et al., 2015), Whizard (Kilian et al., 2011; Moretti et al., 2001). While interfaces for C/C++/Fortran exist in the respective generators, a lightweight and easy-to-use Python interface was missing until the inception of `pylhe` in 2015. Additionally, `pylhe` can serve as a crucial interface for emerging machine learning applications in particle physics, allowing researchers to efficiently extract event data for training neural networks and other machine learning models used in event classification, anomaly detection, and physics analysis.

# State of the field

Unlike the existing C/C++/Fortran interfaces provided by the Monte Carlo event generators, `pylhe` offers a pure Python interface that is more accessible and easier to use. When `pylhe` was first developed, there were no other Python libraries available for reading and writing LHE files. Nowadays, there are a few other smaller Python libraries with less adoption than `pylhe`, which provide only read functionality and are no longer actively maintained, such as `lhereader` (Albert, 2021; Biswas, 2017). For completeness, it should be mentioned that several LHE libraries exist in other programming languages, such as Go (`go-hep` (Binet et al., 2017)), Rust (`lhe` (Fukuda, 2018), `lhef` (Maier, 2024b; Weber, 2018), `event_file_reader` (Maier, 2024a)), Julia (`LHEF.jl` (Fischer & Ling, 2023)) and Haskell (`lhe.hs` (Lawrence, 2019)). These provide varying degrees of completeness.

# Software design

`pylhe` allows for easy reading and writing of `.lhe` and `.lhe.gz` files in Python, enabling seamless integration into modern data analysis workflows in high-energy physics. The pythonic event yielding approach allows for memory-efficient processing of arbitrarily large LHE files by streaming events one at a time rather than loading all of them at once into memory. Internally, `pylhe` uses `xml.etree.ElementTree` to parse the XML structure, since using the `lxml` library did not provide a significant speed up.

The library facilitates quick validation of event files through programmatic access to event structure and particle properties, making it straightforward to perform sanity checks on generated events. This can be done for example via the integration with Awkward Array (Pivarski et al., 2018) through the `to_awkward()` function, which converts LHE events into columnar data structures optimized for vectorized operations and efficient analysis of large datasets.

# Research impact statement

`pylhe` is regularly used in various research projects and publications within high-energy physics. Notably, it has been cited in Higgs studies (Brehmer et al., 2019; Feuerstake et al., 2025; Stylianou & Weiglein, 2024), in Supersymmetry (SUSY), Beyond the Standard Model (BSM) and dark matter searches (Anisha et al., 2023; Beresford et al., 2024; Beresford & Liu, 2019; Cheung et al., 2024; Kling, 2020; Zhou & Liu, 2022, 2025), and in forward physics studies (Kelly et al., 2022; Kling et al., 2023; Kling & Trojanowski, 2020). It is also employed in methodological studies involving machine learning techniques for event generation and analysis (Brehmer et al., 2020; Kofler et al., 2025).

# AI usage disclosure

Generative AI tools have been used in the development of this software and writing of the manuscript:

- Github's copilot has been used in reviewing pull requests.
- VScode's copilot has been used as an advanced autocomplete.
- ChatGPT has been used to identify the most pythonic solutions in case of ambiguity.

All the results generated by these tools have been reviewed by the authors and are correct to the best of our knowledge.


# Acknowledgements

We would additionally like to thank the contributors of pylhe and the Scikit-HEP community for their support.


# References


Albert, A. (2021). *Lhereader*. GitHub. https://github.com/AndreasAlbert/lhereader

Alioli, S., Nason, P., Oleari, C., & Re, E. (2010). A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, *06*, 043. https://doi.org/10.1007/JHEP06(2010)043

Alwall, J., Frederix, R., Frixione, S., Hirschi, V., Maltoni, F., Mattelaer, O., Shao, H.-S., Stelzer, T., Torrielli, P., & Zaro, M. (2014). The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, *07*, 079. https://doi.org/10.1007/JHEP07(2014)079

Alwall, J., & others. (2007). A Standard format for Les Houches event files. *Comput. Phys. Commun.*, *176*, 300–304. https://doi.org/10.1016/j.cpc.2006.11.010

Andersen, J. R., & others. (2014). *Les Houches 2013: Physics at TeV Colliders: Standard Model Working Group Report*. https://doi.org/https://doi.org/10.48550/arXiv.1405.1067

Anisha, Atkinson, O., Bhardwaj, A., Englert, C., Naskar, W., & Stylianou, P. (2023). BSM reach of four-top production at the LHC. *Phys. Rev. D*, *108*(3), 035001. https://doi.org/10.1103/PhysRevD.108.035001

Bahr, M., & others. (2008). Herwig++ Physics and Manual. *Eur. Phys. J. C*, *58*, 639–707. https://doi.org/10.1140/epjc/s10052-008-0798-9

Bellm, J., & others. (2016). Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, *76*(4), 196. https://doi.org/10.1140/epjc/s10052-016-4018-8

Bellm, J., & others. (2020). Herwig 7.2 release note. *Eur. Phys. J. C*, *80*(5), 452. https://doi.org/10.1140/epjc/s10052-020-8011-x

Beresford, L., Clawson, S., & Liu, J. (2024). Strategy to measure tau g-2 via photon fusion in LHC proton collisions. *Phys. Rev. D*, *110*(9), 092016. https://doi.org/10.1103/PhysRevD.110.092016

Beresford, L., & Liu, J. (2019). Search Strategy for Sleptons and Dark Matter Using the LHC as a Photon Collider. *Phys. Rev. Lett.*, *123*(14), 141801. https://doi.org/10.1103/PhysRevLett.123.141801

Bewick, G., & others. (2024). Herwig 7.3 release note. *Eur. Phys. J. C*, *84*(10), 1053. https://doi.org/10.1140/epjc/s10052-024-13211-9

Bierlich, C., Buckley, A., Butterworth, J. M., Gutschow, C., Lonnblad, L., Procter, T., Richardson, P., & Yeh, Y. (2024). Robust independent validation of experiment and theory: Rivet version 4 release note. *SciPost Phys. Codeb.*, *36*, 1. https://doi.org/10.21468/SciPostPhysCodeb.36

Bierlich, C., & others. (2020). Robust Independent Validation of Experiment and Theory: Rivet version 3. *SciPost Phys.*, *8*, 026. https://doi.org/10.21468/SciPostPhys.8.2.026

Bierlich, C., & others. (2022). A comprehensive guide to the physics and usage of PYTHIA 8.3. *SciPost Phys. Codeb.*, *2022*, 8. https://doi.org/10.21468/SciPostPhysCodeb.8

Binet, S., Wieck, B., Blyth, D., Busato, E., Ughetto, M., & Waller, P. (2017). Go-HEP: Libraries for high energy physics analyses in go. *Journal of Open Source Software*, *2*(17), 372. https://doi.org/10.21105/joss.00372

Biswas, D. (2017). *Lhereader*. GitHub. https://github.com/diptaparna/lhereader

Boos, E., & others. (2001, September). Generic User Process Interface for Event Generators. *2nd Les Houches Workshop on Physics at TeV Colliders*. https://arxiv.org/abs/hep-ph/0109068

Bothmann, E., & others. (2019). Event Generation with Sherpa 2.2. *SciPost Phys.*, *7*(3), 034. https://doi.org/10.21468/SciPostPhys.7.3.034

Brehmer, J., Dawson, S., Homiller, S., Kling, F., & Plehn, T. (2019). Benchmarking simplified template cross sections in $WH$ production. *JHEP*, *11*, 034. https://doi.org/10.1007/JHEP11(2019)034

Brehmer, J., Kling, F., Espejo, I., & Cranmer, K. (2020). MadMiner: Machine learning-based inference for particle physics. *Comput. Softw. Big Sci.*, *4*(1), 3. https://doi.org/10.1007/s41781-020-0035-2

Buckley, A., Butterworth, J., Grellscheid, D., Hoeth, H., Lonnblad, L., Monk, J., Schulz, H., & Siegert, F. (2013). Rivet user manual. *Comput. Phys. Commun.*, *184*, 2803–2819. https://doi.org/10.1016/j.cpc.2013.05.021

Butterworth, J. M., & others. (2010, March). THE TOOLS AND MONTE CARLO WORKING GROUP Summary Report from the Les Houches 2009 Workshop on TeV Colliders. *6th Les Houches Workshop on Physics at TeV Colliders*. https://doi.org/https://doi.org/10.48550/arXiv.1003.1643

Cheung, K., Kim, Y., Kwon, Y., Ouseph, C. J., Soffer, A., & Wang, Z. S. (2024). Probing dark photons from a light scalar at Belle II. *JHEP*, *05*, 094. https://doi.org/10.1007/JHEP05(2024)094

Corcella, G., Knowles, I. G., Marchesini, G., Moretti, S., Odagiri, K., Richardson, P., Seymour, M. H., & Webber, B. R. (2001). HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, *01*, 010. https://doi.org/10.1088/1126-6708/2001/01/010

Dembinski, H., Rodrigues, E., & Fedynitch, A. (2022). *Scikit-hep/pyhepmc: v2.1.1* (Version v2.1.1). Zenodo. https://doi.org/10.5281/zenodo.7013498

Dobbs, M., & Hansen, J. B. (2001). The HepMC C++ Monte Carlo event record for High Energy Physics. *Comput. Phys. Commun.*, *134*, 41–46. https://doi.org/10.1016/S0010-4655(00)00189-2

Feuerstake, F., Fuchs, E., Robens, T., & Winterbottom, D. (2025). Interference effects in resonant di-Higgs production at the LHC in the Higgs singlet extension. *JHEP*, *04*, 094. https://doi.org/10.1007/JHEP04(2025)094

Fischer, K., & Ling, J. (2023). *LHEF.jl*. GitHub. https://github.com/JuliaHEP/LHEF.jl

Frixione, S., Nason, P., & Oleari, C. (2007). Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, *11*, 070. https://doi.org/10.1088/1126-6708/2007/11/070

Fukuda, H. (2018). *Lhe-rs*. GitHub. https://github.com/hajifkd/lhe-rs

Gleisberg, T., Hoeche, Stefan., Krauss, F., Schonherr, M., Schumann, S., Siegert, F., & Winter, J. (2009). Event generation with SHERPA 1.1. *JHEP*, *02*, 007. https://doi.org/10.1088/1126-6708/2009/02/007

Kelly, K. J., Kling, F., Tuckler, D., & Zhang, Y. (2022). Probing neutrino-portal dark matter at the Forward Physics Facility. *Phys. Rev. D*, *105*(7), 075026. https://doi.org/10.1103/PhysRevD.105.075026

Kilian, W., Ohl, T., & Reuter, J. (2011). WHIZARD: Simulating Multi-Particle Processes at LHC and ILC. *Eur. Phys. J. C*, *71*, 1742. https://doi.org/10.1140/epjc/s10052-011-1742-y

Kling, F. (2020). Probing light gauge bosons in tau neutrino experiments. *Phys. Rev. D*, *102*(1), 015007. https://doi.org/10.1103/PhysRevD.102.015007

Kling, F., Kuo, J.-L., Trojanowski, S., & Tsai, Y.-D. (2023). FLArE up dark sectors with EM form factors at the LHC forward physics facility. *Nucl. Phys. B*, *987*, 116103. https://doi.org/10.1016/j.nuclphysb.2023.116103

Kling, F., & Trojanowski, S. (2020). Looking forward to test the KOTO anomaly with FASER. *Phys. Rev. D*, *102*(1), 015032. https://doi.org/10.1103/PhysRevD.102.015032

Kofler, A., Stimper, V., Mikhasenko, M., Kagan, M., & Heinrich, L. (2025). Flow annealed importance sampling bootstrap meets differentiable particle physics. *Mach. Learn. Sci. Tech.*, *6*(2), 025061. https://doi.org/10.1088/2632-2153/addbc1

Lawrence, S. (2019). *Lhe.hs*. GitHub. https://github.com/bytbox/lhe.hs

Maier, A. (2024a). *Event-file-reader*. GitHub. https://github.com/a-maier/event-file-reader

Maier, A. (2024b). *Lhef*. GitHub. https://github.com/a-maier/lhef

Moretti, M., Ohl, T., & Reuter, J. (2001). *O'Mega: An Optimizing matrix element generator*. 1981–2009. https://doi.org/https://doi.org/10.48550/arXiv.hep-ph/0102195

Nason, P. (2004). A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, *11*, 040. https://doi.org/10.1088/1126-6708/2004/11/040

Pivarski, J., Osborne, I., Ifrim, I., Schreiner, H., Hollands, A., Biswas, A., Das, P., Roy Choudhury, S., Smith, N., Goyal, M., Fackeldey, P., & Krommydas, I. (2018). *Awkward Array*. https://doi.org/10.5281/zenodo.4341376

Sjostrand, T., Mrenna, S., & Skands, P. Z. (2006). PYTHIA 6.4 Physics and Manual. *JHEP*, *05*, 026. https://doi.org/10.1088/1126-6708/2006/05/026

Sjostrand, T., Mrenna, S., & Skands, P. Z. (2008). A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, *178*, 852–867. https://doi.org/10.1016/j.cpc.2008.01.036

Sjöstrand, T., Ask, S., Christiansen, J. R., Corke, R., Desai, N., Ilten, P., Mrenna, S., Prestel, S., Rasmussen, C. O., & Skands, P. Z. (2015). An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, *191*, 159–177. https://doi.org/10.1016/j.cpc.2015.01.024

Stylianou, P., & Weiglein, G. (2024). Constraints on the trilinear and quartic Higgs couplings from triple Higgs production at the LHC and beyond. *Eur. Phys. J. C*, *84*(4), 366. https://doi.org/10.1140/epjc/s10052-024-12722-9

The HDF Group. (1998). *HDF5 – hierarchical data format version 5*. https://www.hdfgroup.org/HDF5/

Verbytskyi, A., Buckley, A., Grellscheid, D., Konstantinov, D., William Monk, J., Lönnblad, L., Przedzinski, T., & Pokorski, W. (2020). HepMC3 Event Record Library for Monte Carlo Event Generators. *J. Phys. Conf. Ser.*, *1525*(1), 012017. https://doi.org/10.1088/1742-6596/1525/1/012017

Weber, T. (2018). *Lhef-rs*. GitHub. https://github.com/tweber12/lhef-rs

Zhou, H., & Liu, N. (2022). Probing compressed higgsinos with forward protons at the LHC. *JHEP*, *10*, 092. https://doi.org/10.1007/JHEP10(2022)092

Zhou, H., & Liu, N. (2025). Search for nearly degenerate higgsinos via photon fusion with the semileptonic channel at the LHC. *Nucl. Phys. B*, *1010*, 116752. https://doi.org/10.1016/j.nuclphysb.2024.116752