SWIFTGalaxy: a Python package to work with particle groups from SWIFT simulations

Kyle A. Oman ©[1,2]¶

1 Institute for Computational Cosmology, Physics Department, Durham University, United Kingdom **2** Centre for Extragalactic Astronomy, Physics Department, Durham University, United Kingdom ¶ Corresponding author

Summary

SWIFTGalaxy is an open-source astrophysics module that extends SWIFTSimIO (Borrow & Borrisov, 2020) to analyses of particles belonging to individual galaxies simulated with SWIFT (Schaller et al., 2024). It inherits from and extends the functionality of SWIFTSimIO's SWIFTDataset class. It understands the content of halo catalogues and therefore which particles belong to a galaxy or other group of particles, and its integrated properties. The particles occupy a coordinate frame that is enforced to be consistent, such that particles loaded on-the-fly will match e.g. rotations and translations of particles already in memory. Intuitive masking of particle datasets is also enabled. Utilities to make working in cylindrical and spherical coordinate systems more convenient are also provided. Finally, tools to iterate efficiently over multiple galaxies are provided.

Background

Cosmological hydrodynamical galaxy formation simulations begin with a representation of the early universe where its two key constituents—dark matter and gas—are discretized. The dark matter is usually represented with collisionless particles that respond only to the gravitational force, while gas may be represented by particles (Lagrangian tracers) or mesh cells (Eulerian tracers) and obeys the laws of hydrodynamics in addition to gravity. The SWIFT code (Schaller et al., 2024) takes the former approach, implementing the smoothed particle hydrodynamics (SPH, Gingold & Monaghan, 1977; Lucy, 1977) formalism. Observations of the cosmic microwave background are used to constrain a multiscale Gaussian random field to formulate the initial conditions for a simulation of a representative region of the early Universe, before the first galaxies formed (e.g. Bertschinger, 2001). In addition to hydrodynamics and gravity, galaxy formation models include additional physical processes such as radiative gas cooling, star formation, energetic feedback from supernovae and supermassive black hole accretion, and more. Many of these are formally unresolved, leading these to be collectively referred to as "sub-grid" elements of the models.

As time integration of a cosmological hydrodynamical galaxy formation simulation proceeds, outputs containing a "snapshot" of the simulation state are typically written at regular intervals. These contain tables of the properties of particles including at least positions, velocities, masses and unique identifiers. More physically complex particle types, such as stars and gas, carry additional properties such as temperature, metallicity, internal energy, and many more. SWIFT snapshot files include rich metadata recording the full configuration of the code and galaxy formation model, plus physical units and related information for all data.

For many types of analysis the snapshot files alone are insufficient: knowledge of which groups of particles "belong" to gravitationally bound structures—galaxies—is needed. In SWIFT

this is most commonly determined in a two-stage process. In the first stage particles are linked to neighbouring particles separated by less than a suitably chosen linking length to form connected "friends-of-friends" (FOF, Davis et al., 1985) groups. The second step consists of finding overdense, self-bound particle groups or "halos" within each FOF group. There are numerous algorithms that accomplish this task. The current implementation recommended for the SWIFT community is the HBT-Herons (Forouhar Moreno et al., 2025) implementation of the hierarchical bound-tracing plus (HBT+, Han et al., 2018) algorithm. The properties of halos can be tabulated in "halo catalogues", where again many algorithms exist. This is often done as part of the halo-finding process, but the recommendation for the SWIFT community is to move this to a separate step and use the spherical overdensity and aperture processor (SOAP, McGibbon et al., 2025) tool.

Statement of need

The collection of standard cosmological hydrodynamical simulation data products—snapshots of particles, halo membership information and halo catalogues—leads to a generic workflow to begin analysis of individual galaxies in such simulations. First, an object of interest is identified in the halo catalogue. Then, the halo membership information is queried to locate its member particles. Finally, the particles are loaded from the snapshot to proceed with analysis. The SWIFT community already maintains the SWIFTSimIO tool (Borrow & Borrisov, 2020) that supports reading in SWIFT snapshot files and SOAP catalogue files. It uses metadata in the files to annotate data arrays with physical units and relevant cosmological information. It is also able to efficiently select spatial sub-regions of a simulation by taking advantage of the fact the particles in cells of a "top-level cell" grid covering the simulation domain are stored contiguously. Finally, it includes some data visualisation tools.

SWIFTGalaxy extends SWIFTSimIO by implementing the workflow outlined above supplemented with additional features including coordinate transformations, data array masking and efficient iteration over galaxies. All of the SWIFTSimIO features, including use of the visualisation tools, are inherited by SWIFTGalaxy. The package prevents duplication of effort by users who would otherwise each need to implement the same steps independently. It also helps to avoid common errors, such as applying a coordinate transformation to one particle type (e.g. rotation of the dark matter) but forgetting to apply it to others (e.g. the gas remains unrotated) by enforcing a consistent data state where applicable.

A core design principle of SWIFTGalaxy is that only operations with an unambiguous implementation are included. In other words, SWIFTGalaxy tries to avoid making decisions for its users. This is a key difference when compared to other packages serving this purpose, such as pynbody (Pontzen et al., 2013) and yt (Turk et al., 2011) that can also be used with SWIFT data sets. Other packages are also less tailored to SWIFT and therefore less able to take advantage of the detailed structure of the data layout on disk and information available as metadata.

Despite being tailored for use with SWIFT simulations, SWIFTGalaxy is modular: it has interchangeable interfaces to different halo catalogue formats. In addition to the preferred SOAP format, the Velociraptor (Elahi et al., 2019) and Caesar formats are also supported, and adding support for other formats is straightforward by design. A "standalone" mode is also available so that the other features of SWIFTGalaxy can be used on arbitrary collections of particles even if no halo catalogue is available or desired.

SWIFTGalaxy is hosted on GitHub and has documentation available through ReadTheDocs. The first research article using the package has recently appeared (Trayford et al., 2025). Many more projects using it are currently ongoing, especially in the context of the COLIBRE simulations project (Chaikin et al., 2025; Schaye et al., 2025).

References

Bertschinger, E. (2001). Multiscale Gaussian random fields and their application to cosmological simulations. *The Astrophysical Journal Supplement Series*, *137*(1), 1–20. https://doi.org/10.1086/322526

Borrow, J., & Borrisov, A. (2020). swiftsimio: A Python library for reading SWIFT data. *The Journal of Open Source Software*, *5*(52), 2430. https://doi.org/10.21105/joss.02430

Chaikin, E., Schaye, J., Schaller, M., Ploeckinger, S., Bahé, Y. M., Benítez-Llambay, A., Correa, C., Forouhar Moreno, V. J., Frenk, C. S., Huško, F., Kugel, R., McGibbon, R., Richings, A. J., Trayford, J. W., Borrow, J., Crain, R. A., Helly, J. C., Lacey, C. G., Ludlow, A., & Nobels, F. S. J. (2025). COLIBRE: Calibrating subgrid feedback in cosmological simulations that include a cold gas phase. *arXiv e-Prints*, arXiv:2509.04067. https://doi.org/10.48550/arXiv.2509.04067

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, *292*, 371–394. https://doi.org/10.1086/163168

Elahi, P. J., Cañas, R., Poulton, R. J. J., Tobar, R. J., Willis, J. S., Lagos, C. del P., Power, C., & Robotham, A. S. G. (2019). Hunting for galaxies and halos in simulations with VELOCIraptor. *Publications of the Astronomical Society of Australia*, *36*, e021. https://doi.org/10.1017/pasa.2019.12

Forouhar Moreno, V. J., Helly, J., McGibbon, R., Schaye, J., Schaller, M., Han, J., Kugel, R., & Bahé, Y. M. (2025). Assessing subhalo finders in cosmological hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*. https://doi.org/10.1093/mnras/staf1478

Gingold, R. A., & Monaghan, J. J. (1977). Smoothed particle hydrodynamics: Theory and application to non-spherical stars. *Monthly Notices of the Royal Astronomical Society*, *181*, 375–389. https://doi.org/10.1093/mnras/181.3.375

Han, J., Cole, S., Frenk, C. S., Benitez-Llambay, A., & Helly, J. (2018). HBT+: An improved code for finding subhaloes and building merger trees in cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, *474*(1), 604–617. https://doi.org/10.1093/mnras/stx2792

Lucy, L. B. (1977). A numerical approach to the testing of the fission hypothesis. *The Astronomical Journal*, *82*, 1013–1024. https://doi.org/10.1086/112164

McGibbon, R., Helly, J., Schaye, J., Schaller, M., & Vandenbroucke, B. (2025). SOAP: A Python package for calculating the properties of galaxies and halos formed in cosmological simulations. *The Journal of Open Source Software*, *10*(111), 8252. https://doi.org/10.21105/joss.08252

Pontzen, A., Roškar, R., Stinson, G., & Woods, R. (2013). *pynbody: N-Body/SPH analysis for python*. Astrophysics Source Code Library, record ascl:1305.002.

Schaller, M., Borrow, J., Draper, P. W., Ivkovic, M., McAlpine, S., Vandenbroucke, B., Bahé, Y., Chaikin, E., Chalk, A. B. G., Chan, T. K., Correa, C., van Daalen, M., Elbers, W.,

Gonnet, P., Hausammann, L., Helly, J., Huško, F., Kegerreis, J. A., Nobels, F. S. J., … Xiang, Z. (2024). SWIFT: A modern highly-parallel gravity and smoothed particle hydrodynamics solver for astrophysical and cosmological applications. *Monthly Notices of the Royal Astronomical Society*, *530*(2), 2378–2419. https://doi.org/10.1093/mnras/stae922

Schaye, J., Chaikin, E., Schaller, M., Ploeckinger, S., Huško, F., McGibbon, R., Trayford, J. W., Benítez-Llambay, A., Correa, C., Frenk, C. S., Richings, A. J., Forouhar Moreno, V. J., Bahé, Y. M., Borrow, J., Durrant, A., Gebek, A., Helly, J. C., Jenkins, A., Lacey, C. G., … Nobels, F. S. J. (2025). The COLIBRE project: Cosmological hydrodynamical simulations of galaxy formation and evolution. *arXiv e-Prints*, arXiv:2508.21126. https://doi.org/10.48550/arXiv.2508.21126

Trayford, J. W., Schaye, J., Correa, C., Ploeckinger, S., Richings, A. J., Chaikin, E., Schaller, M., Benitez-Llambay, A., Frenk, C., & Husko, F. (2025). Modelling the evolution and influence of dust in cosmological simulations that include the cold phase of the interstellar medium. *arXiv e-Prints*, arXiv:2505.13056. https://doi.org/10.48550/arXiv.2505.13056

Turk, M. J., Smith, B. D., Oishi, J. S., Skory, S., Skillman, S. W., Abel, T., & Norman, M. L. (2011). yt: A multi-code analysis toolkit for astrophysical simulation data. *The Astrophysical Journal Supplement Series*, *192*(1), 9. https://doi.org/10.1088/0067-0049/192/1/9