Astronomy and Computing 57 (2026) 101159



Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom



Full length paper



The FLAMINGO simulations data release

John C. Helly [a,*], Robert J. McGibbon [b], Joop Schaye [b], Matthieu Schaller [b,c], William McDonald [b], Joey Braspenning [b,d], Jeger C. Broxterman [b,c], Emily E. Costello [e], Willem Elbers [a], Victor J. Forouhar Moreno [b], Carlos S. Frenk [a], Adrian Jenkins [a], Roi Kugel [b], Ian G. McCarthy [e], Jaime Salcido [e], Marcel P. van Daalen [b], Bert Vandenbroucke [b], Tianyi Yang [e]

[a] Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham, DH1 3LE, UK
[b] Leiden Observatory, Leiden University, PO Box 9513, Leiden, 2300 RA, The Netherlands
[c] Lorentz Institute for Theoretical Physics, Leiden University, PO Box 9506, Leiden, 2300 RA, The Netherlands
[d] Max–Planck-Institut für Astronomie, Königstuhl 17, Heidelberg, D-69117, Germany
[e] Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool, L3 5RF, UK

ARTICLE INFO

Dataset link: The FLAMINGO Simulations (Original data)

Keywords:
Cosmology: theory
Large-scale structure of Universe
Galaxies: clusters: general
Galaxies: clusters: intracluster medium
Galaxies: formation

ABSTRACT

We describe the public release of > 2.3 petabytes of data from the FLAMINGO cosmological simulations. The suite consists of hydrodynamical simulations that include radiative cooling, star formation, stellar mass loss and the resulting chemical enrichment, supernova feedback, and two implementations of AGN feedback. Neutrinos are simulated explicitly using particles. Data products include snapshots, halo and galaxy catalogues, HEALPix all-sky lightcone maps, particle data for lightcone maps, and power spectra. The FLAMINGO set includes 22 hydrodynamical simulations. In addition, there are 16 gravity-only simulations, including the $10\,080^3$ particles FLAMINGO-10k run, with initial conditions that match those of the corresponding hydrodynamical runs. The fiducial hydrodynamical simulations span three numerical resolutions that have each been calibrated to reproduce the present-day galaxy stellar mass function and gas fractions in low-redshift clusters. Other simulations systematically vary the galaxy stellar mass function, cluster gas fractions, cosmology (including neutrino masses), and/or the nature of dark matter, in volumes of 1 $cGpc^3$. The release includes hitherto unpublished simulations that use extra dark matter particles. While we provide a facility for downloading complete simulation outputs, we recognize that for many users this will not be possible due to limited local storage or network bandwidth. We implement a web service that enables users to explore available outputs and selectively download datasets or parts of datasets.

1. Introduction

Cosmological simulations are widely used to study a large variety of topics in galaxy formation and cosmology. Simulations are needed to predict observational quantities, such as the matter power spectrum or the halo mass function, enabling the measurement of the cosmological parameters on which they depend. Simulations are also required to model factors that complicate the analysis of observations, such as selection and projection effects.

Dark-matter-only (DMO) simulations, i.e. simulations that account for the effects of baryons in the initial conditions but not for the subsequent evolution, have long been the workhorse for comparisons with observational surveys. However, while baryonic matter is expected to trace the dark matter on very large scales, on non-linear scales ($\lesssim$10 Mpc), cooling flows, star formation, and galactic winds driven by stellar and active galactic nucleus (AGN) feedback can alter the distribution of baryons. As observational methods become more precise and probe smaller scales, baryonic effects become increasingly important. Indeed, modelling baryonic effects is, for example, essential for ongoing weak lensing surveys such as *Euclid* (e.g. Semboloni et al., 2011) and *LSST* (e.g. Robertson et al., 2026).

While baryonic effects can be modelled by post-processing DMO simulations (e.g. Schneider and Teyssier, 2015; Aricò et al., 2021), or by using simplified models built on top of DMO frameworks (e.g. Mead et al., 2015; Semboloni et al., 2013; Debackere et al., 2020), such models rely on hydrodynamical simulations to motivate, optimize and validate their methods.

Furthermore, observations rely on the detection of baryonic matter, either diffuse gas or stars. Examples of the former are secondary cosmic

* Corresponding author.
E-mail address: j.c.helly@durham.ac.uk (J.C. Helly).

https://doi.org/10.1016/j.ascom.2026.101159
Received 13 May 2026; Accepted 22 June 2026
Available online 26 June 2026
2213-1337/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

microwave background (CMB) anisotropies, such as the thermal and kinetic Sunyaev–Zel'dovich (SZ) effects, X-ray emission, and fast radio burst (FRB) dispersion measures. Examples of the latter are galaxy clustering and redshift space distortions. In the case of gas, the signal depends on the physical conditions such as the temperature, density, and chemical composition, while in the case of stars it depends on the way in which galaxies populate dark matter haloes. Hydrodynamical simulations can directly predict such observables.

A drawback of using hydrodynamical simulations, as opposed to post-processed DMO simulations, is that hydrodynamical simulations are more expensive computationally, both in terms of computing time and memory/storage requirements. In small volumes (side length $L \lesssim 10^2$ Mpc), which can be simulated at relatively high resolution (particle mass $\lesssim 10^6$ M$_\odot$), it is possible to model the low-mass end of the galaxy population and the detailed properties of massive galaxies. Examples of such simulations are HorizonAGN (Dubois et al., 2014), EAGLE (Schaye et al., 2015; Crain et al., 2015), IllustrisTNG (Pillepich et al., 2018; Nelson et al., 2019a), Simba (Davé et al., 2019), ASTRID (Bird et al., 2022), FIREbox (Feldmann et al., 2023), CROCODILE (Oku and Nagamine, 2024), (X)FABLE (Bigwood et al., 2025a), and COLIBRE (Schaye et al., 2026; Chaikin et al., 2026). However, these simulations do not provide good statistics for rare objects, such as galaxy clusters, and they cannot model the distribution of matter on scales comparable to or larger than their box size. Hence, for cosmology and the study of clusters and massive galaxies, large volumes ($L \gtrsim 10^3$ Mpc), which can currently only be modelled with relatively low resolution ($\gg 10^6$ M$_\odot$), tend to be required. Examples of such simulations are Cosmo-OWLS (Le Brun et al., 2014), BAHAMAS (McCarthy et al., 2017), MillenniumTNG (Pakmor et al., 2023), Magneticum Pathfinder (Dolag et al., 2025), and Frontier-E (Frontiere et al., 2025).

Another drawback of hydrodynamical simulations is that they depend on uncertain subgrid processes. To maximize the utility of hydrodynamical simulations, it is therefore important to calibrate them to observations that are relevant to the goals of the project and to vary the uncertain ingredients. This was the approach taken for the BAHAMAS simulations, for which the feedback was varied relative to a model that reproduces the galaxy stellar mass function, which is important for the galaxy-halo connection, and cluster gas fractions, which is important for the matter power spectrum. In addition, to study the interplay between variations in cosmology and baryonic physics, it is necessary to vary the cosmology.

Here we present the public release of the data from the Virgo Consortium's FLAMINGO[1] project (Schaye et al., 2023; Kugel et al., 2023). FLAMINGO consists of a large suite of cosmological hydrodynamical simulations and their DMO counterparts. Compared with simulations focusing on galaxy formation, the volumes are large ($L \geq 1$ cGpc), but the resolution is modest (baryonic particle mass $\geq 1 \times 10^8$ M$_\odot$). The simulations include not only cold dark matter and baryons, but also neutrinos. The subgrid physics in the fiducial model is calibrated on the observed $z = 0$ galaxy stellar mass function, gas fractions of low-z clusters, and black hole masses in massive present-day galaxies. The calibration is repeated for three different numerical resolutions, spanning two orders of magnitude in particle mass. The suite currently includes ten variations in baryon physics, which systematically vary the galaxy stellar mass function and/or cluster gas fractions. It also includes two different implementations of AGN feedback, thermally-driven winds or kinetic jets. In addition, there are nine different cosmologies (ten for gravity-only), including variations in the parameters of the ΛCDM model, the neutrino mass, decaying dark matter, and a hitherto unpublished cosmology assuming evolving dark energy (the last model is currently only available as DMO).

Following the pioneering work of Lemson and the Virgo Consortium (2006), who presented a publicly-available relational database containing the halo and galaxy properties of the Millennium simulation (Springel et al., 2005b), partial or full releases of the large datasets generated by cosmological simulation campaigns have become a common practice. This has enabled a very large number of studies by scientists not directly affiliated with the teams producing the simulations and has been a major step for open science. The practice was later extended beyond semi-analytic models with examples including the Illustris full data release (Nelson et al., 2015), the EAGLE galaxy catalogues and merger trees (McAlpine et al., 2016) followed by the EAGLE snapshot data (The EAGLE team, 2017), and the release of the IllustrisTNG suite (TNG50, TNG100, and TNG300; Nelson et al. (2019b)). More recently, there have been public releases of zoom-in simulations of suites of individual haloes from FIRE-2 (Wetzel et al., 2023), AURIGA (Grand et al., 2024) and TNG-Cluster (Nelson et al., 2024), collections of small-volume simulations (e.g. the CAMELS set Villaescusa-Navarro et al., 2023, which comprises nearly 17 000 simulations and more than 2 PB of data), and radiation-magnetohydrodynamical simulations of the high redshift universe including THESAN (Garaldi et al., 2024) and Sphinx20 (Katz et al., 2023).

All of these simulation data releases, with the exception of Sphinx20, provide full downloads of raw simulation data. In some cases, facilities for selective extraction, online search and online analysis are also provided. For example, the Illustris, IllustrisTNG and TNG-Cluster releases provide a web-based API to fetch information about specific objects or cut out parts of snapshots, the EAGLE data release includes a database of halos and galaxies which can be queried with SQL, and CAMELS allows data access via Jupyter notebooks.

We follow these practices and provide both full downloads and a mechanism for selective access to outputs, releasing all the FLAMINGO data products that we can afford to keep on disc. This includes snapshots, halo and galaxy catalogues, HEALPIX lightcone maps, particle lightcones, and power spectra. The total volume of the data released exceeds 2.4 PB. The data are hosted at the DiRAC Memory Intensive Service[2] in Durham, UK, where the simulations were run. The large size of the simulations makes full downloads problematic because many users will not have access to sufficient network bandwidth or storage. We address this by providing a mechanism to request single datasets or parts of datasets from the output files. Users with more limited computing resources may still be able to exploit subsets of the data, such as particular galaxies extracted from the snapshots or halo catalogues containing just a few quantities of interest. For ease of use, this system is designed to resemble the tools researchers would use to work with local data.

Since the publication of Schaye et al. (2023), FLAMINGO has already resulted in ≈80 refereed papers[3] on topics such as high-redshift galaxies, clusters, and large-scale structure. Some of these papers have resulted in additional data products that are also part of this release. Thanks to the efforts of the authors of completed and upcoming publications, the strengths and weaknesses of the simulations are fairly well characterized and documented, but we have only scratched the surface of the potential scientific applications.

For (interactive) visualizations and videos of the FLAMINGO simulations, see https://flamingo.strw.leidenuniv.nl/. For access to the data and online documentation, see https://flamingo.strw.leidenuniv.nl/data.html.

This paper is organized as follows. The FLAMINGO simulations and data products are described in Sections 2 and 3, respectively. Section 4 discusses the design and implementation of the data service. In Section 5 we briefly summarize the results of published comparisons with data, the interpretation of comparisons between the different FLAMINGO resolutions, and known issues and bugs. Section 6 provides advice on how to acknowledge the usage of the released data. Finally, we conclude in Section 7.

[1] The acronym stands for **F**ull-hydro **L**arge-scale structure simulations with **A**ll-sky **M**apping for the **I**nterpretation of **N**ext **G**eneration **O**bservations.

[2] https://dirac.ac.uk/memory-intensive-durham/.

[3] See https://flamingo.strw.leidenuniv.nl/papers.html.

2. FLAMINGO simulations

This section contains a brief overview of the FLAMINGO simulations. Further details on the simulation methods and the galaxy formation model are provided in Schaye et al. (2023), while the calibration of the subgrid model is detailed in Kugel et al. (2023).

2.1. The FLAMINGO model

All simulations were performed using the open-source simulation code SWIFT (Schaller et al., 2024).[4] In particular, gravity is solved for using a 4th-order fast-multipole method (e.g. Cheng et al., 1999) coupled to a particle-mesh method for long-range forces (e.g. Bagla and Ray, 2003), the gas is evolved using the SPHENIX (Borrow et al., 2022) flavour of Smoothed Particle Hydrodynamics (SPH), and neutrinos are followed using the δf-method of Elbers et al. (2021).

The FLAMINGO simulations include subgrid prescriptions for radiative cooling and heating following Ploeckinger and Schaye (2020), star formation and an entropy floor for the dense gas based on the method of Schaye and Dalla Vecchia (2008), the stellar evolution and chemical enrichment model of Wiersma et al. (2009), and core collapse supernova feedback implemented kinetically following Chaikin et al. (2023). Supermassive black holes seeding, gas accretion, merging, and repositioning are modelled using ingredients from Springel et al. (2005a), Booth and Schaye (2009) and Bahé et al. (2022). The feedback from AGN is modelled either as thermally-driven winds (Booth and Schaye, 2009) or using the collimated jet model of Huško et al. (2022).

FLAMINGO includes three different resolutions, labelled by the rounded log base 10 of the baryonic particle mass in units of the solar mass: the high-resolution m8, intermediate-resolution m9, and low-resolution m10 (see Table 1). The variations in cosmology and/or galaxy formation physics all use m9 resolution in 1 cGpc boxes. The fiducial model is available at all resolutions and, at m9 resolution, also in a larger volume of 2.8 cGpc on a side.

As described by Kugel et al. (2023), four free parameters of the subgrid models associated with stellar and AGN feedback were calibrated using a Gaussian process emulator, trained on a Latin hypercube of simulations of small volumes, to predict the observables as a function of the input parameters. A Monte Carlo Markov Chain search was then used in combination with the emulator to find the best-fitting value reproducing the target datasets, while accounting for observational errors. In particular, the fiducial model was calibrated to match the $z = 0$ galaxy stellar mass function from the DR4 release of the GAMA survey (Driver et al., 2022) and simultaneously the gas fractions in low-redshift clusters inferred from (pre-eRosita) X-ray and weak-lensing data aggregated by Kugel et al. (2023). For the galaxy mass function the lower limit of the mass range used in the calibration depends on the simulation resolution: $M_* > 10^{8.67}$, $10^{9.92}$, and $10^{11.17}\,\mathrm{M}_\odot$ for m8, m9, and m10, respectively, while the upper limit is always $10^{11.5}\,\mathrm{M}_\odot$. For the cluster gas fractions, the lower limit of the halo mass considered during the calibration is always $M_{500c} = 10^{13.5}\,\mathrm{M}_\odot$, while the upper mass limits are $10^{13.73}$, $10^{14.36}$, and $10^{14.53}\,\mathrm{M}_\odot$ for m8, m9, and m10, respectively.

Besides generating sets of simulation parameters matching the data, additional simulations were generated where the target data was shifted by particular amounts with respect to observations. For the cluster gas fraction variations, the observed gas fractions were shifted up and down compared to the results by $\pm N\sigma$, where σ is the error in the data (see Kugel et al., 2023, for the exact definitions). Similar variations were generated for shifts in the stellar mass function. The free parameters were calibrated independently for simulations using the thermal and jet AGN models and for simulations using different resolutions.

Since the $z = 0$ galaxy stellar mass function and the gas fractions of clusters in the mass and redshift ranges above were directly used to calibrate the FLAMINGO model parameters, these quantities (and closely related quantities such as the stellar mass to halo mass relation and cluster stellar and baryon mass fractions in the corresponding mass ranges) should not be treated as predictions of the model. Additionally, FLAMINGO uses the same AGN feedback efficiency parameters as Booth and Schaye (2009), who chose values which reproduce the normalization of the relation between galaxy stellar mass and central black hole mass. This relation is therefore a product of a previous calibration. Other halo and galaxy properties, such as those presented in section 6 of Schaye et al. (2023), were not used in the calibration process and may be considered to be predictions of the model.

The initial conditions (ICs) were generated using the MONOFONIC code (Hahn et al., 2021; Elbers et al., 2022) using a 3-fluid formalism with a separate transfer function for each of the species. The baseline cosmological model used for the simulations is the maximum likelihood values of the flat ΛCDM + neutrinos model fitted to the DES year 3 data release (Abbott et al., 2022) combined with external probes, i.e. their "3 × 2pt + All Ext". model. The parameter values are given in the first row of Table 3.

2.2. Model variations

The main properties of the 22 hydrodynamical and 16 DMO simulations released here[5] are given in Tables 1 and 2, respectively. Most runs simulate a volume of $(1\ \mathrm{cGpc})^3$ with the intermediate resolution ('m9') model, which corresponds to a (initial) mean baryonic particle mass of $1.07 \times 10^9\ \mathrm{M}_\odot$ and a mean cold dark matter particle mass of $5.65 \times 10^9\ \mathrm{M}_\odot$. The cosmological variations (with parameters given in Table 3) are all performed at this resolution, and so are the feedback variations. The set of m9 runs also includes the 'L2p8_m9' simulation, which uses the fiducial feedback and cosmology but in a volume of $(2.8\ \mathrm{cGpc})^3$, i.e. in a volume 22× larger, and was the largest cosmological hydrodynamical simulation (by number of resolution elements) run to $z = 0$ when first published. The two simulations 'L1_m8' and 'L1_m10' are run in the same $(1\ \mathrm{cGpc})^3$ volume as the L1_m9 simulation, but use 8× higher and lower mass resolutions, respectively, with the subgrid parameter calibration performed at each resolution as described above. Apart from the simulation L1_m9_extraDM, all runs use equal numbers of baryonic and dark matter particles, while the number of neutrino particles is a factor 1.8^3 smaller. All simulations in a given volume use the same phases in the initial conditions, allowing objects to be matched between models.

The cosmology variations (see Table 3) include runs based on the best-fitting ΛCDM model to the Planck data with larger neutrino masses, the low-σ_8 'lensing cosmology' of Amon et al. (2023), and ΛCDM extensions allowing for decaying dark matter (See Elbers et al., 2025, for details).

In addition to the feedback variations described above, there are two hydrodynamical simulations at m9 resolution in the 1 cGpc box. In the 'NoCooling' run radiative cooling (but not photo-heating), star formation, and feedback have all been switched off. Besides the equations of gravity and hydrodynamics, only heating sources (by adiabatic compression, shocks, and photoionization by the UV background radiation) are included in the physics. The last variation is the 'L1_m9_extraDM' model, which uses the fiducial calibration but super-samples the dark matter by a factor of 4. This leads to baryon and dark matter particles

[4] Publicly available, including the exact version used for the FLAMINGO runs, at www.swiftsim.com. The parameter files for the various FLAMINGO simulations are provided on the data release website as well as in the SWIFT code repository.

[5] This set extends the original 16 hydrodynamical and 12 DMO runs presented by Schaye et al. (2023) with additional model variations, primarily varying the values of the cosmological parameters.

Table 1
Hydrodynamical simulations in the FLAMINGO suite. Table replicated from Schaye et al. (2023) and expanded with newer models. The first four lines list the simulations that use the fiducial galaxy formation model and assume the fiducial cosmology, but use different volumes and resolutions. The remaining lines list the model variations, which all use a 1 cGpc box and intermediate resolution. The columns list the simulation identifier (where m8, m9 and m10 indicate $\log_{10}$ of the mean baryonic particle mass and correspond to high, intermediate, and low resolution, respectively; absence of this part implies m9 resolution); the number of standard deviations by which the observed stellar masses are shifted before calibration, Δm_*; the number of standard deviations by which the observed cluster gas fractions are shifted before calibration, $\Delta f_{\rm gas}$; the AGN feedback implementation (thermal or jets); the comoving box side length, L; the number of baryonic particles, $N_{\rm b}$ (which equals the number of CDM particles, $N_{\rm CDM}$, except for the L1_m9_extraDM simulation where $N_{\rm CDM} = 4\times N_{\rm b}$); the number of neutrino particles, N_ν; the initial mean baryonic particle mass, $m_{\rm g}$; the mean CDM particle mass, $m_{\rm CDM}$; the Plummer-equivalent comoving gravitational softening length, $\epsilon_{\rm com}$; the maximum proper gravitational softening length, $\epsilon_{\rm prop}$; the assumed cosmology which is specified in Table 3; and the paper reference introducing the simulations: (a) Schaye et al. (2023), (b) McCarthy et al. (2025), (c) Elbers et al. (2025), (d) this paper.

Identifier	Δm_* (σ)	$\Delta f_{\rm gas}$ (σ)	AGN	L (cGpc)	$N_{\rm b}$	N_ν	$m_{\rm g}$ ($\rm M_\odot$)	$m_{\rm CDM}$ ($\rm M_\odot$)	$\epsilon_{\rm com}$ (ckpc)	$\epsilon_{\rm prop}$ (pkpc)	Cosmology	Ref.
L1_m8	0	0	Thermal	1.0	3600^3	2000^3	1.34×10^8	7.06×10^8	11.2	2.85	D3A	a
L1_m9	0	0	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	a
L1_m10	0	0	Thermal	1.0	900^3	500^3	8.56×10^9	4.52×10^{10}	44.6	11.40	D3A	a
L2p8_m9	0	0	Thermal	2.8	5040^3	2800^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	a
fgas+2σ	0	+2	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	a
fgas−2σ	0	−2	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	a
fgas−4σ	0	−4	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	a
fgas−8σ	0	−8	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	a
M*−σ	−1	0	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	a
M*−σ_fgas−4σ	−1	−4	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	a
Jet	0	0	Jets	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	a
Jet_fgas−4σ	0	−4	Jets	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	a
NoCooling	n/a	n/a	n/a	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	D3A	b
L1_m9_extraDM	0	0	Thermal	1.0	1800^3	1000^3	1.07×10^9	1.41×10^9	14.1	5.70	D3A	d
Planck	0	0	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.72×10^9	22.3	5.70	Planck	a
PlanckNu0p24Var	0	0	Thermal	1.0	1800^3	1000^3	1.06×10^9	5.67×10^9	22.3	5.70	PlanckNu0p24Var	a
PlanckNu0p24Fix	0	0	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.62×10^9	22.3	5.70	PlanckNu0p24Fix	a
PlanckNu0p48Fix	0	0	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.62×10^9	22.3	5.70	PlanckNu0p48Fix	c
LS8	0	0	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	LS8	a
LS8_fgas−8σ	0	−8	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.65×10^9	22.3	5.70	LS8	b
PlanckDCDM12	0	0	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.72×10^9	22.3	5.70	PlanckDCDM12	c
PlanckDCDM24	0	0	Thermal	1.0	1800^3	1000^3	1.07×10^9	5.72×10^9	22.3	5.70	PlanckDCDM24	c

with similar masses and thus reduces the spurious energy transfer between species highlighted by Ludlow et al. (2019, 2021, 2023).

DMO counterparts to all the runs above (except for 'L1_m9_extraDM') are also available. Accompanying these, we also provide the FLAMINGO-10k simulation, which was first used in Pizzati et al. (2024). It features the $(2.8\ \rm cGpc)^3$ volume but at 'm8' resolution. This simulation evolves nearly 1.2×10^{12} particles. The semi-analytic model GALFORM has been run on the output and will soon be available as an additional high-resolution galaxy catalogue.

3. Data products

The FLAMINGO data release comprises a set of data products that are provided consistently across the simulations in the suite, subject to resolution and model-specific differences. A complete description of the file formats, directory layout, dataset definitions, units, and additional metadata is provided in the online documentation.[6] In this section we give a high-level overview of the main data products. Table 4 details the storage requirements for each data product per simulation, including the cumulative volume for each category.

3.1. Snapshots

The simulations originally produced 78 snapshots (79 for L1_m8 and L2p8_m9, 144 for FLAMINGO-10k). With the exception of L1_m9 and L1_m9_DMO, we could not retain all snapshots due to data storage constraints. Therefore, for each simulation we release the full particle snapshots at the following redshifts[7]:

$z = 5, 4, 3, 2, 1.5, 1, 0.75, 0.5, 0.4, 0.3, 0.2, 0.1, 0.$

Snapshots are stored in HDF5 format and follow the standard layout of the SWIFT simulation code. They contain all evolved particle species (dark matter, gas, stars, black holes, and neutrinos), together with extensive metadata describing the cosmology, unit system, and run configuration. Detailed definitions of all fields and their units are given in the online documentation, though this information is also available within the snapshot files themselves. An example of reading snapshot data can be found in Appendix A.2.

Alongside the full particle snapshots, there are two sets of partial snapshots, which are available for all outputs of the original simulation:

- **Reduced snapshots** contain all the particles within R_{100c} of the centres of a selected sample of massive haloes. The haloes are selected by binning in M_{200c}, with bins of width 0.05 dex starting at $M_{200c} = 10^{13}\rm M_\odot$. If a mass bin contains fewer than 200 haloes, all haloes in that bin are included, otherwise 200 haloes are randomly selected.
- **Downsampled snapshots**[8] contain a random 1% of all particles in the snapshot. The particles retained are selected independently for each snapshot and are therefore not consistent across snapshots. All black hole particles are retained.

Interactive visualizations of slices through FLAMINGO simulation snapshots are available on the project web site at https://flamingo.strw.leidenuniv.nl/map_slider.html.

3.2. Halo and galaxy catalogues

The history-based structure finder HBT-HERONS (Han et al., 2018; Forouhar Moreno et al., 2025) was used to identify subhaloes. Halo and

[6] https://flamingo.strw.leidenuniv.nl/data.html.

[7] 32 snapshots have been kept for FLAMINGO-10k.

[8] Downsampled snapshots are not available for the L1m8 simulation.

Table 2
Gravity-only simulations in the FLAMINGO suite. Table replicated from Schaye et al. (2023) and expanded with newer models. The columns list the simulation identifier; the comoving box side length, L; the number of CDM particles, $N_{\rm CDM}$; the number of neutrino particles, N_ν; the mean CDM particle mass, $m_{\rm CDM}$; the Plummer-equivalent comoving gravitational softening length, $\epsilon_{\rm com}$; the maximum proper gravitational softening length, $\epsilon_{\rm prop}$; the assumed cosmology which is specified in Table 3; and the paper reference introducing the simulations: (a) Schaye et al. (2023), (c) Elbers et al. (2025), (d) this paper, (e) Pizzati et al. (2024). Simulation L1_m9_ip_DMO is identical to L1_m9_DMO except that the phases in the initial conditions were inverted. Note that there are no hydrodynamical counterparts to FLAMINGO-10k, L5p6_m10_DMO, L11p2_m11_DMO, PlanckNu0p12Var_DMO, and L1_m9_ip_DMO.

Identifier	L (cGpc)	$N_{\rm CDM}$	N_ν	$m_{\rm CDM}$ ($M_\odot$)	$\epsilon_{\rm com}$ (ckpc)	$\epsilon_{\rm prop}$ (pkpc)	Cosmology	Ref.
L1_m8_DMO	1.0	3600^3	2000^3	8.40×10^8	11.2	2.85	D3A	a
L1_m9_DMO	1.0	1800^3	1000^3	6.72×10^9	22.3	5.70	D3A	a
L1_m10_DMO	1.0	900^3	500^3	5.38×10^{10}	44.6	11.40	D3A	a
FLAMINGO-10k	2.8	$10\,080^3$	5600^3	8.40×10^8	11.2	2.85	D3A	e
L2p8_m9_DMO	2.8	5040^3	2800^3	6.72×10^9	22.3	5.70	D3A	a
L5p6_m10_DMO	5.6	5040^3	2800^3	5.38×10^{10}	44.6	11.40	D3A	a
L11p2_m11_DMO	11.2	5040^3	2800^3	4.30×10^{11}	89.2	22.80	D3A	a
L1_m9_ip_DMO	1.0	1800^3	1000^3	6.72×10^9	22.3	5.70	D3A	d
Planck_DMO	1.0	1800^3	1000^3	6.78×10^9	22.3	5.70	Planck	a
PlanckNu0p12Var_DMO	1.0	1800^3	1000^3	6.74×10^9	22.3	5.70	PlanckNu0p12Var	a
PlanckNu0p24Var_DMO	1.0	1800^3	1000^3	6.73×10^9	22.3	5.70	PlanckNu0p24Var	a
PlanckNu0p24Fix_DMO	1.0	1800^3	1000^3	6.68×10^9	22.3	5.70	PlanckNu0p24Fix	a
PlanckNu0p48Fix_DMO	1.0	1800^3	1000^3	6.68×10^9	22.3	5.70	PlanckNu0p48Fix	c
LS8_DMO	1.0	1800^3	1000^3	6.72×10^9	22.3	5.70	LS8	a
PlanckDCDM12_DMO	1.0	1800^3	1000^3	6.74×10^9	22.3	5.70	PlanckDCDM12	c
PlanckDCDM24_DMO	1.0	1800^3	1000^3	6.73×10^9	22.3	5.70	PlanckDCDM24	c

Table 3
The values of the cosmological parameters used in different simulations. Table replicated from Schaye et al. (2023) and expanded with newer models. The columns list the prefix used to indicate the cosmology in the simulation name; the dimensionless Hubble constant, h; the total matter density parameter, $\Omega_{\rm m}$; the dark energy density parameter, Ω_Λ; the baryonic matter density parameter, $\Omega_{\rm b}$; the sum of the particle masses of the neutrino species, $\sum m_\nu c^2$; the amplitude of the primordial matter power spectrum, $A_{\rm s}$; the power-law index of the primordial matter power spectrum, $n_{\rm s}$; the amplitude of the initial power spectrum parametrized as the r.m.s. mass density fluctuation in spheres of radius 8 h^{-1} Mpc extrapolated to $z=0$ using linear theory, σ_8; the amplitude of the initial power spectrum parametrized as $S_8 \equiv \sigma_8\sqrt{\Omega_{\rm m}/0.3}$; the neutrino matter density parameter, $\Omega_\nu \cong \sum m_\nu c^2/(93.14\ {\rm h^2\ eV})$, and the dark matter decay rate Γ. Note that the values of the Hubble and density parameters are given at $z=0$.

Prefix	h	$\Omega_{\rm m}$	Ω_Λ	$\Omega_{\rm b}$	$\sum m_\nu c^2$ (eV)	$A_{\rm s}$	$n_{\rm s}$	σ_8	S_8	Ω_ν	Γ (H_0/h)
D3A (fiducial)	0.681	0.306	0.694	0.0486	0.06	2.099×10^{-9}	0.967	0.807	0.815	1.39×10^{-3}	–
Planck	0.673	0.316	0.684	0.0494	0.06	2.101×10^{-9}	0.966	0.812	0.833	1.42×10^{-3}	–
PlanckNu0p12Var	0.673	0.316	0.684	0.0496	0.12	2.113×10^{-9}	0.967	0.800	0.821	2.85×10^{-3}	–
PlanckNu0p24Var	0.662	0.328	0.672	0.0510	0.24	2.109×10^{-9}	0.968	0.772	0.807	5.87×10^{-3}	–
PlanckNu0p24Fix	0.673	0.316	0.684	0.0494	0.24	2.101×10^{-9}	0.966	0.769	0.789	5.69×10^{-3}	–
PlanckNu0p48Fix	0.673	0.316	0.684	0.0494	0.48	2.101×10^{-9}	0.966	0.709	0.728	11.4×10^{-3}	–
LS8	0.682	0.305	0.695	0.0473	0.06	1.836×10^{-9}	0.965	0.760	0.766	1.39×10^{-3}	–
PlanckDCDM12	0.673	0.274	0.726	0.0494	0.06	2.101×10^{-9}	0.966	0.794	0.759	1.42×10^{-3}	0.12
PlanckDCDM24	0.673	0.239	0.726	0.0494	0.06	2.101×10^{-9}	0.966	0.777	0.694	1.42×10^{-3}	0.24

Table 4
Approximate sizes of the various data products for different simulations. All values are given in terabytes. The columns list the simulation type (defined by box size, resolution, and whether it is DMO), the size of the $z=0$ snapshot, the size of the $z=0$ SOAP catalogue, the size of a single downsampled HEALPix map (i.e. a single shell or an integrated map), the size of a single halo lightcone, and the size of a single particle lightcone. The final row gives the total data volume of each type of data product over all simulations and all redshifts. The total value for snapshot includes the reduced and downsampled snapshots, the total value for the catalogues includes the HBT-Herons merger trees, and the total values for maps includes the full resolution ($N_{\rm side}=16384$) maps.

Simulation	Snapshot	Halo catalogue	$N_{\rm side}$ = 4096 map	Halo lightcone	Particle lightcone
L1_m8	4.8	0.15	0.015	6	38
L1_m9	0.6	0.03	0.015	0.6	4.4
L1_m10	0.08	0.005	0.015	–	0.5
L2p8_m9	14	0.7	0.015	0.6	27
L1_m8_DMO	1	0.05	0.002	–	–
L1_m9_DMO	0.1	0.008	0.001	–	–
L1_m10_DMO	0.02	0.001	0.001	–	–
FLAMINGO-10k	21	1.8	–	–	–
L2p8_m9_DMO	2.8	0.17	0.001	–	–
L5p6_m10_DMO	2.8	0.21	0.001	–	–
L11p2_m11_DMO	2.8	0.19	–	–	–
Total	1400	180	600	40	160

galaxy catalogues were computed using SOAP (McGibbon et al., 2025), which uses the subhalo centres and gravitational boundness of member particles determined by HBT-HERONS. The SOAP outputs have a file structure similar to that of the snapshots, and can also be read using SWIFTSIMIO. Catalogues are available for each of the original 78 (or 79) simulation snapshots.

SOAP provides halo properties defined according to multiple common conventions, allowing users to select the definition most appropriate for their scientific application:

- **Spherical overdensity properties**: The halo radius is defined as the radius at which the mean enclosed density reaches a specified threshold (e.g. R_{200c}). All particles within this radius are used to compute the corresponding halo properties.
- **Inclusive and exclusive spherical apertures**: Properties are measured within apertures of a fixed physical radius. Inclusive apertures include all particles within the aperture, while exclusive apertures include only particles that are gravitationally bound to the subhalo.
- **Projected apertures**: Properties are computed within two-dimensional projected physical apertures. For each subhalo, projections are provided along the simulation's x, y, and z axes.

For complete merger tree information, the HBT-HERONS catalogues are also provided. The complete documentation can be found on the HBT-HERONS website.[9]

To facilitate matching haloes across simulations, we construct matching catalogues as follows. For each halo in a reference simulation, we consider all particles bound to the halo. Using their unique particle IDs, we then identify the halo in the comparison simulation that contains the largest number of these particles. This matching procedure is then applied in reverse, allowing bijective matches to be identified. Each hydrodynamical simulation is matched to its corresponding DMO counterpart, which in turn enables matching between different feedback variations.

3.3. Lightcones

To facilitate direct comparison with large-area surveys, we release three types of lightcone data products constructed from the simulations: all-sky HEALPix maps, the properties of the particles that contribute to them ("particle lightcones"), and the properties of the haloes contained in them ("halo lightcones").

Particle lightcones record every particle that crossed the observer's past lightcone to some maximum redshift. At high redshift the volume and number of particles in the lightcone become extremely large, so for each particle type there is an upper redshift limit beyond which particles are not output. The construction methodology, redshift sampling, and file formats are described in detail in Appendix A of Schaye et al. (2023) and in the online documentation.

For each 1 cGpc hydrodynamical simulation, two lightcones corresponding to different observer positions are available, with HEALPix maps for one extending to $z = 3$ and for the other to $z = 0.5$ (see Table A1 of Schaye et al. (2023) for the extent of the various particle lightcones). The L2p8 simulation originally produced eight lightcone outputs, so we provide eight sets of maps, although only two particle lightcones remain. For this simulation, each lightcone's HEALPix map extends to $z = 5$. See the online documentation for the redshift ranges covered by the particle lightcones for the different simulations and particle species.

We release halo lightcones that extend to redshift $z = 15$. These are constructed using black hole particles as tracers of the subhalo positions, with halo properties obtained from the SOAP catalogues at the nearest snapshot in redshift.

For the DMO simulations there are HEALPix maps but no particle or halo lightcones.

HEALPix maps

We release full-sky projected maps in concentric spherical shells around the observer. When a particle crosses the observer's lightcone, we determine in which radial shell it lies and then accumulate its contribution to the appropriate maps. These maps use the HEALPix pixelisation scheme (Gorski et al., 2005) to divide the sky into pixels of equal area. The maps are provided at two resolutions. The highest resolution maps have an angular resolution of ≈13 arcseconds ($N_{side} = 16384$). As these maps can be computationally demanding to work with, we also provide downsampled maps with an angular resolution of ≈50 arcseconds ($N_{side} = 4096$).

The widths of the shells used to construct the FLAMINGO HEALPix maps can exceed the size of the simulation box. This means that a photon travelling towards the observer may pass through multiple periodic replications of the same structure, introducing spurious correlations in the maps. We therefore also release a number of integrated maps for which the FLAMINGO shells have been randomly rotated whenever the lightcone diameter exceeds a multiple of the box size. The same rotation has been applied to each of the observables, meaning the different integrated maps can be directly compared and cross correlated.

The following HEALPix maps are currently available:

- **Mass** (Schaye et al., 2023, Appendix A2.3). The total projected mass in each pixel for different particle types. There are individual maps for black hole mass, dark matter mass, stellar mass, gas mass (both with and without smoothing), neutrino mass, and total mass. **[Shells only]**
- **Star formation rate** (Schaye et al., 2023, Appendix A2.3). The total projected star formation rate in each pixel. Each time a gas particle crosses the lightcone its associated star formation rate is accumulated to the pixel containing the particle's position on the sky. **[Shells only]**
- **Weak lensing convergence maps** (Broxterman et al., 2024). Constructed from the total matter lightcone shells using a backward ray-tracing methodology and assuming a simple analytic non-tomographic *Euclid*-like source redshift distribution ($N_{side} = 8192$). **[Integrated only]**
- **CMB lensing** (Yang et al., 2026). Lensing convergence maps using the matter overdensity computed from the total-matter lightcone outputs. The two-dimensional projected overdensity map shells are integrated along the line of sight up to $z = 3.0$ and 5.0 for the 1 and 2.8 cGpc simulation volumes, respectively, weighted by the CMB lensing kernel. **[Integrated only]**
- **Thermal Sunyaev–Zel'dovich effect** (Schaye et al., 2023; Yang et al., 2026). Constructed by accumulating the Compton-y parameter. In addition to the maps corresponding to individual redshift shells, integrated maps are available, where lensing effects are included using PIXELL,[10] applied shell by shell with the integrated convergence map up to the shell of interest. Additionally, a relativistically corrected version of the thermal SZ intensity fluctuation map is available, with the correction function computed using SZPACK (Chluba et al., 2012, 2013). Contributions from gas particles that are star forming or have recently been directly heated[11] by AGN feedback are excluded. **[Integrated and shells]**
- **Kinetic Sunyaev–Zel'dovich effect** (Schaye et al., 2023; Yang et al., 2026). Constructed by accumulating the Doppler-b parameter. Integrated maps are available, where the mapping between the Doppler parameter is given by $\Delta T_{kSZ}/T_{CMB} = -b$, where $T_{CMB} = 2.73$ K. Lensing effects are computed using the same

[9] https://hbt-herons.strw.leidenuniv.nl/.

[10] https://github.com/simonsobs/pixell.

[11] A gas particle is considered recently heated if it has been directly heated by AGN feedback within the last 15 Myr and has a temperature of $10^{-1}\Delta T_{AGN} < T < 10^{0.3}\Delta T_{AGN}$, where, for the fiducial model at m9 resolution, $\Delta T_{AGN} = 10^{7.78}$ K.

procedure as used for lensing of the thermal SZ map. As for the thermal SZ effect, contributions from particles that are star forming or have recently been heated by AGN are excluded. **[Integrated and shells]**

- **Anisotropic screening (optical depth τ)** (Yang et al., 2026). The screening effect is proportional to the line-of-sight integrated optical depth $\tau(\hat{\mathbf{n}})$, which is given by the Thomson scattering cross-section multiplied by the stacked dispersion measure map. As for the thermal SZ effect, contributions from particles that are star forming or have recently been heated by AGN are excluded. **[Integrated and shells]**
- **Dispersion measure** (Schaye et al., 2023, Appendix A2.3). Computed by integrating the free electron number densities in the lightcone outputs. As for the thermal SZ effect, contributions from particles that are star forming or have recently been heated by AGN are excluded. **[Shells only]**
- **Cosmic infrared background** (Yang et al., 2026). Generated from the star formation rate lightcone outputs, with the bolometric infrared luminosity assumed to be proportional to the star formation rate (Kennicutt, 1998). The luminosity at a given frequency is then computed using a greybody radiation spectral energy distribution (SED) for infrared sources (Planck Collaboration et al., 2016), with the SED parameters determined by fitting to the measured 353, 545, and 857 GHz CIB power spectra from Lenz et al. (2019). The same procedure as used for lensing of the thermal SZ map is applied here. **[Integrated only]**
- **Radio point source emission** (Yang et al., 2026). Constructed from the black hole particle lightcone outputs. Radio luminosities are assigned by abundance matching the bolometric AGN luminosity function to the LOFAR 150 MHz luminosity function up to $z = 2.5$ (Kondapally et al., 2022). The lensed source fluxes are extrapolated to higher CMB frequencies using a power-law SED, with the power index fitted to match the measured radio source counts from the SPT survey at 95, 150, and 220 GHz (Everett et al., 2020). The procedure is repeated for subsamples selected by black hole accretion state using different Eddington ratio cuts, $\lambda_{\rm Edd} \equiv L_{\rm bol}/L_{\rm Edd}$: $\lambda_{\rm Edd} < 10^{-2}, 10^{-3}$ and 10^{-6}. **[Integrated only]**
- **Diffuse X-ray emission** (Schaye et al., 2023; McDonald et al., 2026). Constructed by summing the X-ray fluxes from gas particles with temperature $10^5 \leq T/{\rm K} \leq 10^{9.5}$ in three different energy bands: eROSITA high (2.3–8.0 keV), eROSITA low (0.2–2.3 keV) and ROSAT (0.5–2.0 keV). In each band, the X-ray fluxes in each pixel are available in units of photons cm^{-2} s^{-1} and erg cm^{-2} s^{-1}, or photons s^{-1} and erg s^{-1} when convolved with the response matrix of the telescope's energy bands. We utilize the effective area information from the publicly available ROSAT[12] on-axis response function and the survey-averaged eROSITA[13] auxiliary response files. Each particle's X-ray emissivity is computed using tables generated with CLOUDY (Ferland et al., 2017, version 17.02), accounting for the individual elemental abundances (see Braspenning et al., 2024, for details) and is consistent with the radiative cooling rates used during the simulation (Ploeckinger and Schaye, 2020). As for the thermal SZ effect and dispersion measure, X-ray emission from gas particles that are star forming or that have recently been heated by AGN feedback is excluded. **[Integrated and shells]**
- **X-ray emission from AGN** (McDonald et al., 2026). X-ray flux from AGN as predicted directly by FLAMINGO's black hole accretion rates or using the fiducial AGN abundance matching model of McDonald et al. (2026) for the same bands as the diffuse X-ray emission. These maps will become available upon acceptance of McDonald et al. (2026). **[Integrated only]**

Interactive visualizations of the HEALPix maps are available on the FLAMINGO project web site, at https://flamingo.strw.leidenuniv.nl/lightcone_slider_CY.html.

3.4. Power spectra

We release various 3-D power spectra computed on-the-fly at 123 redshifts. For each redshift, we provide the comoving wavenumber, k, and the corresponding shot-noise subtracted power, $P(k)$, for 14 distinct auto- and cross-spectra. These include power spectra for total matter, cold dark matter (CDM), gas, the sum of stars and black holes (starBH), and electron pressure. We also provide cross-power spectra for CDM with gas, neutrinos, and starBH; gas with total matter, neutrinos, and starBH; total matter with electron pressure; starBH with neutrinos; and the cross-correlation between two randomly selected subsets of the neutrino particles, which is done for shot-noise suppression. Note that the baryon response of the power spectrum (i.e., the ratio of the total matter power spectrum in the hydrodynamical simulation to its dark-matter-only counterpart) was already released, including in the form of a Gaussian process emulator embedded in a PYTHON package, by Schaller et al. (2025).

3.5. Initial conditions

The initial conditions for the L1_m9 and L1_m9_DMO simulations are available on the data release website. These are using the SWIFT format for initial conditions based on the `hdf5` storage library. For all the other simulations in the FLAMINGO suite, we intend to provide parameter files for the MONOFONIC code that can be used to reproduce the initial conditions. Further details on the creation of the initial conditions can be found in Section 2.4 and Appendix B of Schaye et al. (2023).

Both MONOFONIC and the SWIFT simulation code (including all physics modules used in the FLAMINGO simulations) are open source. We have also published the parameter files used to run the simulations. This means that, in principle, anyone will be able to reproduce the simulations using the public code and parameters. Users could run their own FLAMINGO variants in different volumes or with different model parameters, for example, or run zoom simulations of individual objects.

3.6. Future data products

We anticipate that additional derived and value-added products will be computed and released over time as the analysis of the simulation continues. New data products will follow the same directory and format conventions as the existing release where possible, and will be documented online. Announcements of new releases and updates will be made via a dedicated mailing list. Instructions for subscribing can be found at https://flamingo.strw.leidenuniv.nl/mailing-list.html.

4. Data access

In this section we describe the system we have set up to provide access to the FLAMINGO simulations.

4.1. Design considerations

The FLAMINGO simulations consist of a large and varied set of outputs. The total size of the data release is in excess of 2.3 PB and includes simulation snapshots, halo catalogues, all-sky HEALPix lightcone maps, and lightcone particle data. The complete data set and even individual simulation snapshots are too large for most users to download in full.We therefore provide a mechanism that allows for selective access

[12] https://heasarc.gsfc.nasa.gov/docs/rosat/ruh/handbook/node122.html#tabeffarea.

[13] Specifically TM8, which excludes TM5 and TM7 as they suffer from light leakage (see Predehl et al., 2021), https://erosita.mpe.mpg.de/dr1/eSASS4DR1/eSASS4DR1_arfrmf/.

to the data to enable scientific exploitation of the simulations by users who might have limited access to computing resources.

The FLAMINGO outputs are stored on the DiRAC Memory Intensive system in Durham, where they are accessed by researchers running HPC jobs. The large size of the outputs means that duplicating them (for example, to cloud storage) is not desirable. This implies that the data must be served directly from the Lustre parallel file system where they are stored, and they must remain in the HDF5 file format required for analysis by local HPC users. This also precludes the use of a relational database, as in the Millennium (Springel et al., 2005b) and EAGLE (Schaye et al., 2015) projects, which would require conversion of the data into the database server's internal representation.

The HDF5 file format was originally designed for local disk storage and is not well suited to remote access. In particular, metadata can be distributed throughout the file in many small chunks. This means that reading HDF5 files using generic tools for remote file system access can be inefficient because it requires many small, random access reads. This problem can be addressed by rewriting the files as "cloud optimized" HDF5 or NetCDF, where HDF5 library parameters are tuned to consolidate metadata as far as possible, or using a format specifically optimized for remote access, such as Zarr.[14] However, for the FLAMINGO data release we wish to serve our existing HDF5 data. We also have some idea of likely access patterns: cutting out regions from the snapshots or patches on the sky from lightcone outputs will involve requesting many slices from one or two dimensional datasets. This can be efficiently implemented as a service which carries out the necessary random access I/O operations on the local file system and produces a compact response for the user to download.

For users who do have enough network bandwidth and storage to download complete outputs, we wish to support full file downloads. The system hosting the data has a Globus[15] endpoint which can be used for bulk data transfers, although this currently requires authentication with a local user account. For simplicity, we would like to also support plain http downloads through the same server that provides selective access to the data.

Additional considerations include:

- From experience with previous simulation data releases, we know that this type of service typically has small numbers of concurrent users. Multiple simultaneous requests must be supported, but scaling beyond a single server is not required.
- The datasets of most interest to users are generally large numeric arrays. Metadata must also be accessible, but fast access to large arrays of fixed size data types is a priority.
- We do not expect the simulation data to change after release, although new data products may be added later.

4.2. Implementation

Given the above constraints, we chose to set up a relatively simple, lightweight service capable of serving the contents of HDF5 files stored on a local file system. Users send http requests specifying a file, a dataset of interest, or possibly a slice within the dataset, and the server reads and returns the corresponding data. This service is designed to re-use existing DiRAC infrastructure as far as possible.

4.2.1. Serialization format

Returning data in HDF5 format poses some practical problems. If we wish to construct and return a new HDF5 file with the requested data, it is difficult to avoid caching the full response on the server before returning any data to the user. This would delay the response and require potentially large and unpredictable amounts of temporary storage. This is not an insurmountable problem (the IllustrisTNG web service is able to construct HDF5 responses, for example) but we prefer to choose a format better suited to streaming so that the user can begin downloading data while it is still being read from the underlying storage. We anticipate that most requests will be for large arrays of numeric data, so the response format must be able to represent these arrays efficiently.

We opted to encode responses using the messagepack[16] serialization format for the following reasons: Messagepack is a compact, binary format that provides all the data types necessary to represent the metadata in our HDF5 simulation outputs. It also includes a raw binary buffer data type which can be used to represent the body of large arrays in an efficient way. This means that dataset values read from HDF5 files can be copied directly to the response stream without translation. Additionally, the messagepack format is much simpler to implement than HDF5 and so many implementations exist.

Note that we do not intend for users to interact with messagepack responses directly. We provide a python module which can unpack responses into `numpy` arrays, which can be either used directly in calculations in memory or written to local HDF5 files using `h5py`.

4.2.2. Parallelization

The FLAMINGO simulation outputs make extensive use of HDF5 filters for data compression. This means that we cannot easily implement our own routines to read data from the files, we must use the HDF5 library. This is a problem because the library does not support thread-based parallelism. Multiple threads may call the library, but these calls are serialized using a single, global lock. Since our service needs to handle multiple concurrent requests, it is necessary to handle each request in a separate process.

We implemented a process pool to handle requests in parallel. When the server receives a request, a process from the pool is assigned to read the data. The process uses the HDF5 library to read the requested data in chunks up to some maximum buffer size. Each chunk is serialized and passed back to the server via shared memory. The server can then write each chunk to the response stream.

To reduce load on the file system, the processes cache recently opened files and datasets, and the server routes requests to processes that already have the requested file and dataset in their cache where possible.

4.2.3. Server implementation

The server is a single instance of the Apache Tomcat[17] application server running on a system with 24 CPU cores, 700 GB memory and a 20 GB/s network connection. While this system will not scale to very large numbers of users, it is capable of providing high throughput for individual requests. Large HDF5 datasets can be streamed at up to 200 MB/s, typically limited by the rate at which a single process can read from the file system, and likely faster than most users can download the response.

The server responds to two types of requests for HDF5 data:

- Recursive serialization of an HDF5 group and its contents, omitting dataset contents larger than some size threshold and subject to a maximum depth for recursion into sub-groups.
- Serialization of all or part of the contents of a dataset.

The first request type can be used to determine what datasets are available, retrieve metadata and efficiently fetch any small datasets. The second type is used to retrieve the actual simulation data. The server also supports requests for file and directory metadata and we have reused code developed for the release of the EAGLE simulations to allow full downloads of files and directories.

[14] https://github.com/zarr-developers/.
[15] https://globus.org.
[16] https://www.msgpack.org.
[17] https://tomcat.apache.org/.

The source code for our server implementation is available on github.[18]

4.2.4. Python client module

We implemented a python client module[19] that can be used to request data from the server. This is designed to resemble the popular `h5py`[20] module for ease of use. HDF5 groups are represented as dict-like containers of datasets and sub-groups. HDF5 datasets are represented by "remote dataset" objects, which are array-like objects that request data from the server when their elements are accessed. These remote datasets support a subset of `numpy` array indexing syntax. Simple slices, specifying a contiguous range of elements in each dimension, generate a request for the corresponding elements of the dataset on the server and return a `numpy` ndarray. A remote dataset may also be indexed with an integer or boolean array in the first dimension. In this case, the index array is translated into a sequence of contiguous slices which are retrieved from the server with a single http request. In Appendix A.1 we illustrate how this module can be used to access FLAMINGO data.

We also adapted the SWIFTSIMIO module (Borrow and Borrisov, 2020) to access FLAMINGO snapshots and halo catalogues using our web service. SWIFTSIMIO is designed to provide an easy-to-use, high-level interface for working with snapshots generated by the SWIFT simulation code (Schaller et al., 2024). The example in Appendix A.2 shows how to open a snapshot on the server using SWIFTSIMIO and Appendix A.3 shows how to download the particles in some region of the simulation volume and save them to a local snapshot file for later analysis. In Appendix A.4 we show how to select a halo from the SOAP catalogue and download the particles bound to it.

We plan to adapt the widely used PYNBODY module (Pontzen et al., 2013) to access FLAMINGO in a similar way.

4.2.5. Web interface

We created a web interface that allows users to browse the contents of the data release. It presents a virtual directory structure containing the simulation data and can also display the contents of HDF5 files. This latter capability is useful because the FLAMINGO outputs contain many descriptive attributes. The web interface also presents full download links for the data and extensive documentation. The data browsing facility is implemented as a javascript client, which makes requests to the server for file and directory contents to display using the same data access API as the python module.

The source code for the web interface, including documentation describing the FLAMINGO outputs, is also published on github.[21]

5. Discussion

5.1. (Dis)agreement with observations and scientific caveats

The FLAMINGO universe is not the real Universe. Nevertheless, the fiducial FLAMINGO model has been shown to reproduce many observations. On the galaxy population side this includes the $z = 0$ galaxy stellar mass function, black hole mass–stellar mass relation, specific star formation rates, and metallicities, though not necessarily over the full mass range covered by the catalogues (Schaye et al., 2023). While all resolutions reproduce the observed $z = 0$ galaxy stellar mass function in the mass range for which they were calibrated, they diverge not only below their resolution limits, but also above the upper mass limit used for the calibration, $M_* = 10^{11.5}\,\mathrm{M}_\odot$, where it also becomes sensitive to the chosen aperture (Schaye et al., 2023).

Only a few studies have so far compared the predicted galaxy populations to observations at higher z. The cosmic star formation history is reproduced up to the redshifts where unresolved low-mass galaxies begin to dominate ($z \approx 5$ for m9 resolution) (Schaye et al., 2023). Lim et al. (2024) and Kumar et al. (2025) showed that FLAMINGO reproduces high-star formation rate objects such as submillimeter galaxies better than previous simulations.

Discrepancies with galaxy observations include $z = 0$ galaxy sizes being slightly too large for $M_* \lesssim 10^{10}\,\mathrm{M}_\odot$ and quenched fractions being too low for the most massive galaxies, $M_* \gtrsim 10^{12}\,\mathrm{M}_\odot$ (Schaye et al., 2023). At high z, Baker et al. (2025) found that, similar to most other simulations, FLAMINGO predicts too few massive, quiescent galaxies.

The quasar luminosity function compares well with observation for $z \lesssim 1$, but its bright end is underestimated at higher redshift (Ding et al., 2026). The clustering of quasars is reproduced for $z \lesssim 3$, but underestimated at $z \approx 4$ (Ding et al., 2026).

Galaxy cluster scaling relations, including their evolution, agree well with the data (e.g. Schaye et al., 2023; Braspenning et al., 2024), even for masses greater than the (box-size dependent) upper halo mass limit used in the calibration ($M_{500c} > 10^{13.73}$ and $10^{14.36}\,\mathrm{M}_\odot$ for the m8 and m9 resolutions, respectively). Cluster thermodynamic radial profiles are also reasonable, also at high redshift, though metallicities (i.e. iron abundances) are substantially too high at small radii (Braspenning et al., 2024).

In the group/low-mass cluster regime ($10^{13} < M_{500c}/\mathrm{M}_\odot < 10^{14}$) the situation is currently unclear. Stacks of kinetic SZ observations combined with weak lensing mass measurements, and stacks of X-ray data from optically-selected groups are both fit much better by the low gas fraction variations of FLAMINGO than the fiducial model (e.g. McCarthy et al., 2025; Hadzhiyska et al., 2025; Popesso et al., 2025; Kovač et al., 2025; Siegel et al., 2026; Bigwood et al., 2025b), suggesting that feedback in the regime of groups or their progenitors needs to be rather strong. However, studies using X-ray observations of individually detected groups, some of which account for selection effects, indicate that the fiducial FLAMINGO model is consistent with the data, but the strong feedback variants are not (Kugel et al., 2023; Eckert et al., 2026). This apparent contradiction could indicate systematic errors or selection effects in some of the data sets, but it could also indicate that the simulations are insufficiently realistic because kinetic SZ and X-ray observations probe different redshifts and radii.

On the large-scale structure side, CMB lensing and its cross correlations with the thermal SZ effect and cosmic shear agree with the simulations (McCarthy et al., 2023; Schaye et al., 2023), while, for the fiducial cosmology, the cosmic shear and thermal SZ power spectra show well-known tensions (reflecting the so-called S_8 tension; McCarthy et al., 2023). However, recent observations of cosmic shear (Wright et al., 2025) and the thermal SZ power spectrum (Efstathiou and McCarthy, 2025) are discrepant with earlier ones and, at least on large scales, are closer to the FLAMINGO predictions. The cross correlation between X-ray emission and cosmic shear is consistent with the data (McDonald et al., 2026). The cosmic infrared background and its cross correlation with CMB lensing also reproduce the data (Yang et al., 2026).

5.2. On tests of numerical convergence

The FLAMINGO simulations include three different numerical resolutions, which differ by factors of 8 (2) in mass (spatial) resolution. For each resolution the subgrid feedback was independently calibrated to match the same observables, though over different galaxy and cluster mass ranges (Kugel et al., 2023).

The resolution limit will generally depend on the property that is being measured and the level of precision required. We advise testing convergence with numerical resolution for each question that is investigated. However, we stress that differences between different resolutions may not (only) reflect direct resolution effects. This is

[18] https://github.com/jchelly/hdfstream-api.
[19] https://hdfstream-python.readthedocs.io/.
[20] https://docs.h5py.org.
[21] https://github.com/jchelly/flamingo-docs.

because the calibration is imperfect and, more importantly, because the calibration is performed only for a few observables. Simulations using different resolutions are effectively different physical models because they resolve different physical processes and because important aspects of the feedback, such as the mass heated/kicked per event, depend directly on the resolution. Because the physical models differ, different resolutions cannot be expected to necessarily produce similar results for observables that were not considered during the calibration. For example, in the left panel of Fig. 8 of Schaye et al. (2023) we can see that the galaxy stellar mass function at a mass of around $10^{12}\mathrm{M}_{\odot}$ is significantly different between m8, m9 and m10 resolutions. Similarly, the stellar half mass radii of active galaxies, shown in Fig. 14 of the same paper, depend on the resolution of the simulation.

5.3. Bugs and issues affecting convergence tests

The interpretation of resolution tests for FLAMINGO is complicated by the fact that we incorporated one improvement and fixed one bug in those simulations that were run in a later phase of the project.

In all intermediate-resolution simulations except for the Jet models, particles with metallicity equal to precisely zero used a star formation threshold density of $n_{\mathrm{H}} = 10\ \mathrm{cm}^{-3}$ instead of $10^{-1}\ \mathrm{cm}^{-3}$. Tests show that this only has significant effects on galaxies with fewer than 10 star particles, where it artificially suppresses the stellar-to-halo mass ratio.

A more important difference between the non-Jet m9 simulations and all other models is related to the black hole repositioning scheme. Because the simulations do not resolve the dynamical friction experienced by black holes, they are moved down the gradient of the gravitational potential by hand. For the purpose of this repositioning of black hole particles, we should[22] subtract the contribution from the black hole to the gravitational potential. As already reported in Schaye et al. (2023), this was, however, only done for the m8 and m10 resolutions and for the Jet variations of the set using m9 resolution. In all other simulations the black hole particles are repositioned less efficiently because they can become trapped in their own potential wells. As shown in Bahé et al. (2022), less efficient repositioning reduces the efficiency of AGN feedback. Tests in smaller volumes show that this is the primary cause of the higher quenched fractions among massive galaxies in the fiducial m8 and the Jet m9 simulations compared to the other m9 models.

5.4. Other bugs and issues

The following significant bugs and issues were already discussed in Schaye et al. (2023),

- For computational efficiency reasons, black hole particles are only repositioned (i.e. moved by hand down the potential gradient to compensate for unresolved dynamical friction) onto gas particles. For gas-poor galaxies, such as low-mass satellites, this can have the consequence that black holes leave their host galaxy, either temporarily or permanently. Care should therefore be taken when studying black holes and/or AGN feedback in satellite galaxies. This issue also affects the galaxy lightcones. Because the locations of black hole particles are used to place galaxies on the lightcone, a fraction of poorly resolved galaxies, in particular satellites, is missing from the galaxy lightcones (see the online documentation for quantitative information).
- AGN feedback is implemented by heating/kicking particles to very high temperatures/velocities, which is necessary to overcome numerical overcooling (e.g. Dalla Vecchia and Schaye, 2012). Because the gas particles subject to energy injection by feedback are selected from the SPH neighbours of black holes/young stars, they tend to be part of the dense interstellar medium. This implies that for a few time steps following energy injection, i.e. until the particles have responded hydrodynamically to the energy injection, such dense and hot gas can artificially distort the observational properties of galaxies, such as their X-ray emission. We therefore advise to test the effect of excluding recently heated/kicked particles, which can be done using the particle property tracking the last time a particle was injected with AGN feedback energy. For some observables (gas and spectroscopic-like temperatures, Compton-y, X-ray) the SOAP catalogues provide versions that exclude particles that were subject to direct AGN heating in the last 15 Myr and whose temperatures are between $10^{-1}\,\Delta T_{\mathrm{AGN}}$ and $10^{0.3}\,\Delta T_{\mathrm{AGN}}$, where ΔT_{AGN} is the AGN heating temperature. However, due to a bug, the SOAP catalogues of all simulations assumed the same value of $\Delta T_{\mathrm{AGN}} = 10^{7.95}\,\mathrm{K}$, which is the AGN heating temperature used in the m9 simulations with the fiducial galaxy formation model, instead of the values listed in Table 1 of Schaye et al. (2023), which vary between $\Delta T_{\mathrm{AGN}} = 10^{7.71}\,\mathrm{K}$ (for fgas+2σ) and $10^{8.40}\,\mathrm{K}$ (for fgas−8σ).
- The original HEALPix maps for the kinetic SZ effect and the dispersion measure were computed using a wrong power of the cosmological scale factor. This was corrected using the expansion factor of the midpoint of each lightcone shell, which has a width of $\Delta z = 0.05$. Note that for those simulations and redshifts for which particle lightcone data is available, the maps can be recomputed if desired (the particle lightcone data has already been corrected).

A comprehensive list of all known issues and bugs, including those mentioned above, as well as additional ones with minor effects, can be found at https://flamingo.strw.leidenuniv.nl/known-issues.html and will be updated as more issues are discovered.

6. Acknowledgement of usage

Publications making use of the public FLAMINGO data are kindly requested to:

- Cite the two papers presenting the FLAMINGO project: Schaye et al. (2023) and Kugel et al. (2023), as well as this data release paper.
- When referring to specific aspects of the methodology, cite the original papers presenting those aspects rather than the overview paper.
- Add the following statement to the acknowledgements: *We acknowledge the Virgo Consortium for making their simulation data available. The* FLAMINGO *simulations were performed using the Durham Memory Intensive system managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC facility (www.dirac.ac.uk).*

7. Conclusions

The FLAMINGO suite of cosmological simulations is a set of 22 hydrodynamical and 16 dark-matter-only simulations in volumes $\geq 1\ \mathrm{cGpc}^3$. These simulations were designed as virtual counterparts to late-time Universe cosmological probes, and thus provide a test bed for the analysis pipelines, especially to understand the effects of systematics induced by baryonic physics. Besides cosmological applications, the FLAMINGO runs provide a theoretical framework for the study of rare objects, such as clusters of galaxies. The fiducial simulations have been carefully calibrated, via machine-learning techniques, to the low-z galaxy stellar mass function and the gas fractions of clusters, offering a plausible realistic baseline scenario. Importantly, the fiducial models

[22] As far as we know, this subtraction was not done by previous cosmological simulations.

are complemented with resolution variations, variations in the baryon calibration, and in the assumed cosmology.

We release here >2.3 petabytes of data from this simulation suite, which we make accessible via a web interface. In addition to the raw data, we provide catalogues of haloes and galaxies, power spectra of various quantities, full-sky tomographic and integrated lightcone maps, and particles data from which additional lightcone maps can be computed. By making these data publicly available, we hope to provide the community with a valuable resource for the theoretical interpretation of cosmology and galaxy evolution datasets.

CRediT authorship contribution statement

John C. Helly: Writing – review & editing, Writing – original draft, Software, Data curation. **Robert J. McGibbon:** Writing – review & editing, Writing – original draft, Software, Data curation. **Joop Schaye:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization. **Matthieu Schaller:** Writing – review & editing, Writing – original draft, Supervision, Software, Project administration, Funding acquisition, Conceptualization. **William McDonald:** Writing – review & editing, Writing – original draft, Software, Methodology, Data curation. **Joey Braspenning:** Software, Methodology, Data curation. **Jeger C. Broxterman:** Writing – review & editing, Methodology, Data curation. **Emily E. Costello:** Writing – review & editing. **Willem Elbers:** Writing – review & editing, Software, Methodology. **Victor J. Forouhar Moreno:** Writing – review & editing, Software. **Carlos S. Frenk:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Adrian Jenkins:** Writing – review & editing, Software, Resources. **Roi Kugel:** Software, Methodology, Investigation. **Ian G. McCarthy:** Writing – review & editing, Supervision, Methodology. **Jaime Salcido:** Writing – review & editing. **Marcel P. van Daalen:** Writing – review & editing, Methodology. **Bert Vandenbroucke:** Writing – review & editing, Software, Data curation. **Tianyi Yang:** Writing – review & editing, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to all FLAMINGO users who helped identify issues with the data and documentation.

JCH is supported by STFC consolidated grant ST/X001075/1. This project has received funding from the Netherlands Organization for Scientific Research (NWO) through research programme Athena 184.034.002. VJFM acknowledges support by NWO through the Dark Universe Science Collaboration (OCENW.XL21.XL21.025). This work was supported by the Science and Technology Facilities Council (grant number ST/Y002733/1). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 769130). This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. For the purpose of open access, the authors have applied a Creative Commons attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Appendix. Usage examples

A.1. Remote access to HDF5 files

In this section we illustrate how to use the FLAMINGO web service to access the contents of HDF5 files on the server. The python module used to access the service can be installed using pip:

```
pip install hdfstream
```

We can connect to the service in a python session as follows:

```
import hdfstream
root_dir = hdfstream.open("cosma", "/")
```

Here, `"cosma"` is an alias for the server to connect to and we are opening the root directory on the server. The result is a `RemoteDirectory` object which represents the service root directory. Files and sub-directories are opened by subscripting the `RemoteDirectory` with a path:

```
flamingo_dir = root_dir["FLAMINGO"]
```

This returns a `RemoteFile` or `RemoteDirectory` representing the file or directory at the specified path. Since a `RemoteDirectory` behaves like a dictionary of files and subdirectories, we can obtain a directory listing by printing its keys:

```
print(list(flamingo_dir))
```

If we already have the location of a file of interest (from the web interface, for example), we can open it by subscripting the root directory object with the full path. For example, a SOAP halo catalogue can be opened with:

```
soap_file = root_dir["FLAMINGO/L1_m10/L1_m10_DMO/SOAP-
    HBT/halo_properties_0077.hdf5"]
```

This opens the file containing the halo catalogue for the redshift $z = 0$ snapshot of the `L1_m10` simulation. It returns a `RemoteFile` object, which is a dict-like container for HDF5 groups and datasets. It behaves similarly to an `h5py` file object, but dataset elements are downloaded on access. We can download the positions of all haloes in the snapshot with:

```
halo_pos = soap_file["BoundSubhalo/CentreOfMass"][...]
```

Here, indexing the dataset with an ellipsis triggers a download of the full dataset contents and the result is a `numpy` array. We can also download parts of datasets:

```
halo_pos = soap_file["BoundSubhalo/CentreOfMass"
    ][0:1000,:]
```

This would download just the first 1000 elements in the first dimension. The units of the dataset can be determined by inspecting its attributes:

```
print(soap_file["BoundSubhalo/CentreOfMass"].attrs)
```

This will return details which include a description of the dataset, a conversion factor to CGS units, and any dependence on the cosmological expansion factor, in the case of co-moving quantities. This approach can be used to access all types of FLAMINGO data products. However, we have also adapted the SWIFTSIMIO module to provide a higher level interface to snapshots and SOAP halo catalogues.

A.2. Accessing snapshots with SWIFTSIMIO

The SWIFTSIMIO python module can be used to access datasets in FLAMINGO snapshots and halo catalogues, either by reading downloaded files directly with `h5py` or by accessing files on the server using the `hdfstream` module. It handles metadata by defining a `numpy.ndarray` sub-class which contains information about the units of each dataset, whether the quantity is stored in co-moving coordinates, and any dependence on the cosmological scale factor. It can be installed with:

```
pip install swiftsimio hdfstream
```

The snapshots and halo catalogues are stored in very similar formats and SWIFTSIMIO is capable of reading both. Here, we use the redshift zero snapshot from the `L1_m10` simulation as an example. The file containing the snapshot can be opened in python with:

```
import hdfstream
root_dir = hdfstream.open("cosma", "/")
snap_file = root_dir["FLAMINGO/L1_m10/L1_m10/snapshots/
    flamingo_0077/flamingo_0077.hdf5"]
```

We can then use SWIFTSIMIO to access the contents of the file:

```
import swiftsimio as sw
snapshot = sw.load(snap_file)
```

This returns a `SWIFTDataset` object, which has attributes corresponding to the particle types in the snapshot:

```
snapshot.gas
snapshot.dark_matter
snapshot.stars
snapshot.black_holes
snapshot.neutrinos
```

Each of these particle types has attributes corresponding to particle properties and accessing an attribute triggers a request to the server for the data. For example, we can download the positions of the star particles with:

```
star_pos = snapshot.stars.coordinates
```

The result is a SWIFTSIMIO cosmo array, which is an array with additional metadata describing the units of the quantity and any dependence on the cosmological scale factor. In this case, it records that coordinates are stored in comoving Mpc:

```
>>> print(star_pos)
[[649.22656783 524.32466783 595.55844783]
 ...
 [987.07910218 995.73126218  88.39935218]] Mpc (
    Comoving)
```

A.3. Downloading regions of interest

The particles in SWIFT snapshots are ordered in a way that makes it possible to read regions from a snapshot without reading the entire file. The simulation volume is divided into grid cells and the particles are sorted according to which cell they fall in. Given a range of coordinates in x, y, and z, SWIFTSIMIO can compute which parts of the file to read or download to find the corresponding particles. The example below shows how to do this:

```
# Open the root directory
import hdfstream
root_dir = hdfstream.open("cosma", "/")

# Open the snapshot file.
snap_file = root_dir["FLAMINGO/L1_m10/L1_m10/snapshots/
    flamingo_0077/flamingo_0077.hdf5"]

# Create a mask object
import swiftsimio as sw
mask = sw.mask(snap_file)

# Define the region to read
from unyt import Mpc
load_region = [[100*Mpc, 150*Mpc], [100*Mpc, 150*Mpc],
    [100*Mpc,150*Mpc]]

# Constrain the region to read
mask.constrain_spatial(load_region)

# Open the snapshot using the mask
snap = sw.load(snap_file, mask=mask)

# Download coordinates of gas particles in the region
gas_pos = snap.gas.coordinates
```

It is also possible to save the region as a new snapshot file. Using the `mask` object from the example above:

```
sw.subset_writer.write_subset("flamingo_region_0077.
    hdf5", mask)
```

This will download all particles in the selected region of the snapshot and save them to a new HDF5 file on the user's local file system.

A.4. Downloading particles in a halo

The python code below shows how to identify a halo of interest in a SOAP halo catalogue and download the particles which belong to that halo, using the most massive halo in the `L1_m10_DMO` simulation at redshift $z = 0$ as an example. This is done by reading all particles in a box around the halo and then discarding particles which are not associated with the halo.

```
import numpy as np
import unyt
import hdfstream
import swiftsimio as sw

# Connect to the hdfstream service and open the root
    directory
root_dir = hdfstream.open("cosma", "/")

# Open the z=0 halo catalogue from the L1_m10_DMO
    simulation
soap_file = root_dir["FLAMINGO/L1_m10/L1_m10_DMO/SOAP-
    HBT/halo_properties_0077.hdf5"]
soap = sw.load(soap_file)

# Get halo positions, masses and indexes
halo_pos = soap.input_halos.halo_centre
halo_m200c = soap.spherical_overdensity_200_crit.
    total_mass
halo_index = soap.input_halos.halo_catalogue_index

# Select the most massive halo
i = np.argmax(halo_m200c)
target_pos = halo_pos[i,:]
target_index = halo_index[i]

# Choose a region to read in. Note that we need to
    choose a radius large
# enough to enclose all halo particles.
radius = 5.0*unyt.Mpc
region = [[x-radius, x+radius] for x in target_pos]

# Open the z=0 snapshot from the L1_m10_DMO simulation
    and select this region
snap_file = root_dir["FLAMINGO/L1_m10/L1_m10_DMO/
    snapshots/flamingo_0077/flamingo_0077.hdf5"]
```

```
mask = sw.mask(snap_file)
mask.constrain_spatial(region)
snap = sw.load(snap_file, mask=mask)

# Read position and halo index of dark matter particles
      in this region
dm_pos = snap.dark_matter.coordinates
dm_halo_index = snap.dark_matter.halo_catalogue_index

# Of the particles we read, identify those which belong
      to the halo
in_halo = (dm_halo_index == target_index)
halo_dm_pos = dm_pos[in_halo,:]
```

The result is an array containing the positions of the particles in the halo. Other particle properties can be read in a similar way.

Data availability

The FLAMINGO simulation data is available at

The FLAMINGO Simulations (Original data) (The FLAMINGO Data Release)

References

Abbott, T.M.C., Aguena, M., Alarcon, A., Allam, S., Alves, O., Amon, A., et al., 2022. Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing. Phys. Rev. D 105 (2), 023520. doi:10.1103/PhysRevD.105.023520, arXiv:2105.13549.

Amon, A., Robertson, N.C., Miyatake, H., Heymans, C., White, M., DeRose, J., et al., 2023. Consistent lensing and clustering in a low-S_8 Universe with BOSS, DES year 3, HSC year 1, and KiDS-1000. MNRAS 518 (1), 477–503. doi:10.1093/mnras/stac2938, arXiv:2202.07440.

Aricò, G., Angulo, R.E., Contreras, S., Ondaro-Mallea, L., Pellejero-Ibañez, M., Zennaro, M., 2021. The BACCO simulation project: a baryonification emulator with neural networks. MNRAS 506 (3), 4070–4082. doi:10.1093/mnras/stab1911, arXiv:2011.15018.

Bagla, J.S., Ray, S., 2003. Performance characteristics of TreePM codes. New A 8, 665–677. doi:10.1016/S1384-1076(03)00056-3, arXiv:astro-ph/0212129.

Bahé, Y.M., Schaye, J., Schaller, M., Bower, R.G., Borrow, J., Chaikin, E., Kugel, R., Nobels, F., Ploeckinger, S., 2022. The importance of black hole repositioning for galaxy formation simulations. MNRAS 516 (1), 167–184. doi:10.1093/mnras/stac1339, arXiv:2109.01489.

Baker, W.M., Lim, S., D'Eugenio, F., Maiolino, R., Ji, Z., Arribas, S., Bunker, A.J., Carniani, S., Charlot, S., de Graaff, A., Hainline, K., Looser, T.J., Lyu, J., Rinaldi, P., Robertson, B., Schaller, M., Schaye, J., Scholtz, J., Übler, H., Williams, C.C., Willmer, C.N.A., Willott, C., Zhu, Y., 2025. The abundance and nature of high-redshift quiescent galaxies from JADES spectroscopy and the FLAMINGO simulations. MNRAS 539 (1), 557–589. doi:10.1093/mnras/staf475, arXiv:2410.14773.

Bigwood, L., Bourne, M.A., Iršič, V., Amon, A., Sijacki, D., 2025a. The case for large-scale AGN feedback in galaxy formation simulations: insights from XFABLE. MNRAS 542 (4), 3206–3230. doi:10.1093/mnras/staf1435, arXiv:2501.16983.

Bigwood, L., Yamamoto, M., Siegel, J., Amon, A., McCarthy, I.G., Dave, R., Salcido, J., Schaller, M., Schaye, J., Yang, T., 2025b. The kinetic Sunyaev Zeldovich effect as a benchmark for AGN feedback models in hydrodynamical simulations: insights from DESI + ACT. doi:10.48550/arXiv.2510.15822, ArXiv E-Prints, arXiv:2510.15822.

Bird, S., Ni, Y., Di Matteo, T., Croft, R., Feng, Y., Chen, N., 2022. The ASTRID simulation: galaxy formation and reionization. MNRAS 512 (3), 3703–3716. doi:10.1093/mnras/stac648, arXiv:2111.01160.

Booth, C.M., Schaye, J., 2009. Cosmological simulations of the growth of supermassive black holes and feedback from active galactic nuclei: method and tests. MNRAS 398 (1), 53–74. doi:10.1111/j.1365-2966.2009.15043.x, arXiv:0904.2572.

Borrow, J., Borrisov, A., 2020. swiftsimio: A Python library for reading SWIFT data. J. Open Source Softw. 5 (52), 2430. doi:10.21105/joss.02430.

Borrow, J., Schaller, M., Bower, R.G., Schaye, J., 2022. SPHENIX: smoothed particle hydrodynamics for the next generation of galaxy formation simulations. MNRAS 511 (2), 2367–2389. doi:10.1093/mnras/stab3166, arXiv:2012.03974.

Braspenning, J., Schaye, J., Schaller, M., McCarthy, I.G., Kay, S.T., Helly, J.C., Kugel, R., Elbers, W., Frenk, C.S., Kwan, J., Salcido, J., van Daalen, M.P., Vandenbroucke, B., 2024. The FLAMINGO project: galaxy clusters in comparison to X-ray observations. MNRAS 533 (3), 2656–2676. doi:10.1093/mnras/stae1436, arXiv:2312.08277.

Broxterman, J.C., Schaller, M., Schaye, J., Hoekstra, H., Kuijken, K., Helly, J.C., Kugel, R., Braspenning, J., Elbers, W., Frenk, C.S., Kwan, J., McCarthy, I.G., Salcido, J., van Daalen, M.P., Vandenbroucke, B., 2024. The FLAMINGO project: baryonic impact on weak gravitational lensing convergence peak counts. MNRAS 529 (3), 2309–2326. doi:10.1093/mnras/stae698, arXiv:2312.08450.

Chaikin, E., Schaye, J., Schaller, M., Benítez-Llambay, A., Nobels, F.S.J., Ploeckinger, S., 2023. A thermal-kinetic subgrid model for supernova feedback in simulations of galaxy formation. MNRAS 523 (3), 3709–3731. doi:10.1093/mnras/stad1626, arXiv:2211.04619.

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., 2026. COLIBRE: calibrating subgrid feedback in cosmological simulations that include a cold gas phase. MNRAS 548 (1), stag300. doi:10.1093/mnras/stag300, arXiv:2509.04067.

Cheng, H., Greengard, L., Rokhlin, V., 1999. A fast adaptive multipole algorithm in three dimensions. J. Comput. Phys. 155 (2), 468–498. doi:10.1006/jcph.1999.6355, URL: http://www.sciencedirect.com/science/article/pii/S0021999199963556.

Chluba, J., Nagai, D., Sazonov, S., Nelson, K., 2012. A fast and accurate method for computing the Sunyaev-Zel'dovich signal of hot galaxy clusters. MNRAS 426 (1), 510–530. doi:10.1111/j.1365-2966.2012.21741.x, arXiv:1205.5778.

Chluba, J., Switzer, E., Nelson, K., Nagai, D., 2013. Sunyaev-Zeldovich signal processing and temperature-velocity moment method for individual clusters. MNRAS 430 (4), 3054–3069. doi:10.1093/mnras/stt110, arXiv:1211.3206.

Crain, R.A., Schaye, J., Bower, R.G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C.S., McCarthy, I.G., Helly, J.C., Jenkins, A., Rosas-Guevara, Y.M., White, S.D.M., Trayford, J.W., 2015. The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations. MNRAS 450 (2), 1937–1961. doi:10.1093/mnras/stv725, arXiv:1501.01311.

Dalla Vecchia, C., Schaye, J., 2012. Simulating galactic outflows with thermal supernova feedback. MNRAS 426 (1), 140–158. doi:10.1111/j.1365-2966.2012.21704.x, arXiv:1203.5667.

Davé, R., Anglés-Alcázar, D., Narayanan, D., Li, Q., Rafieferantsoa, M.H., Appleby, S., 2019. SIMBA: Cosmological simulations with black hole growth and feedback. MNRAS 486 (2), 2827–2849. doi:10.1093/mnras/stz937, arXiv:1901.10203.

Debackere, S.N.B., Schaye, J., Hoekstra, H., 2020. The impact of the observed baryon distribution in haloes on the total matter power spectrum. MNRAS 492 (2), 2285–2307. doi:10.1093/mnras/stz3446, arXiv:1908.05765.

Ding, B., Pizzati, E., Schaye, J., Hennawi, J.F., McDonald, W., Schaller, M., 2026. The luminosity function and clustering of bright quasars in the FLAMINGO cosmological simulations. MNRAS 547 (2), stag308. doi:10.1093/mnras/stag308, arXiv:2510.24283.

Dolag, K., Remus, R.-S., Valenzuela, L.M., Kimmig, L.C., Seidel, B., Fortune, S., Stoiber, J., Ivleva, A., Hoffmann, T., Biffi, V., Marini, I., Popesso, P., Vladutescu-Zopp, S., 2025. Encyclopedia magneticum: Scaling relations from cosmic dawn to present day. doi:10.48550/arXiv.2504.01061, ArXiv E-Prints, arXiv:2504.01061.

Driver, S.P., Bellstedt, S., Robotham, A.S.G., Baldry, I.K., Davies, L.J., Liske, J., et al., 2022. Galaxy And Mass Assembly (GAMA): Data Release 4 and the $z < 0.1$ total and $z < 0.08$ morphological galaxy stellar mass functions. MNRAS 513 (1), 439–467. doi:10.1093/mnras/stac472, arXiv:2203.08539.

Dubois, Y., Pichon, C., Welker, C., Le Borgne, D., Devriendt, J., Laigle, C., et al., 2014. Dancing in the dark: galactic properties trace spin swings along the cosmic web. MNRAS 444 (2), 1453–1468. doi:10.1093/mnras/stu1227, arXiv:1402.1165.

Eckert, D., Seppi, R., Braspenning, J., Finoguenov, A., Gastaldello, F., Lovisari, L., O'Sullivan, E., Ettori, S., Oppenheimer, B.D., Bourne, M.A., Kim, D.-W., Sun, M., Khalil, H., Gozaliasl, G., Bahar, Y.E., Ghirardini, V., Cui, W., Kolokythas, K., McGee, S., 2026. The impact of strong feedback on galaxy group scaling relations. A&A 709, L4. doi:10.1051/0004-6361/202558334, arXiv:2512.04203.

Efstathiou, G., McCarthy, F., 2025. The power spectrum of the thermal Sunyaev–Zeldovich effect. MNRAS 540 (1), 1055–1068. doi:10.1093/mnras/staf709, arXiv:2502.10232.

Elbers, W., Frenk, C.S., Jenkins, A., Li, B., Helly, J.C., Kugel, R., Schaller, M., Schaye, J., Braspenning, J., Kwan, J., McCarthy, I.G., Salcido, J., van Daalen, M.P., Vandenbroucke, B., Pascoli, S., 2025. The FLAMINGO project: the coupling between baryonic feedback and cosmology in light of the S_8 tension. MNRAS 537 (2), 2160–2178. doi:10.1093/mnras/staf093, arXiv:2403.12967.

Elbers, W., Frenk, C.S., Jenkins, A., Li, B., Pascoli, S., 2021. An optimal non-linear method for simulating relic neutrinos. MNRAS 507 (2), 2614–2631. doi:10.1093/mnras/stab2260, arXiv:2010.07321.

Elbers, W., Frenk, C.S., Jenkins, A., Li, B., Pascoli, S., 2022. Higher order initial conditions with massive neutrinos. MNRAS 516 (3), 3821–3836. doi:10.1093/mnras/stac2365, arXiv:2202.00670.

Everett, W.B., Zhang, L., Crawford, T.M., Vieira, J.D., Aravena, M., Archipley, M.A., et al., 2020. Millimeter-wave point sources from the 2500 square degree SPT-SZ Survey: Catalog and population statistics. ApJ 900 (1), 55. doi:10.3847/1538-4357/ab9df7, arXiv:2003.03431.

Feldmann, R., Quataert, E., Faucher-Giguère, C.-A., Hopkins, P.F., Çatmabacak, O., Kereš, D., Bassini, L., Bernardini, M., Bullock, J.S., Cenci, E., Gensior, J., Liang, L., Moreno, J., Wetzel, A., 2023. FIREbox: simulating galaxies at high dynamic range in a cosmological volume. MNRAS 522 (3), 3831–3860. doi:10.1093/mnras/stad1205, arXiv:2205.15325.

Ferland, G.J., Chatzikos, M., Guzmán, F., Lykins, M.L., van Hoof, P.A.M., Williams, R.J.R., Abel, N.P., Badnell, N.R., Keenan, F.P., Porter, R.L., Stancil, P.C., 2017. The 2017 release cloudy. Rev. Mexicana Astron. Astrofis. 53, 385–438, arXiv:1705.10877.

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. MNRAS 543 (2), 1339–1372. doi:10.1093/mnras/staf1478, arXiv:2502.06932.

Frontiere, N., Emberson, J.D., Buehlmann, M., Habib, S., Heitmann, K., Ramachandra, N., Faucher-Giguère, C.-A., 2025. Modeling galaxy formation in cosmological simulations with CRK-HACC. doi:10.48550/arXiv.2511.21921, ArXiv E-Prints, arXiv:2511.21921.

Garaldi, E., Kannan, R., Smith, A., Borrow, J., Vogelsberger, M., Pakmor, R., Springel, V., Hernquist, L., Galárraga-Espinosa, D., Yeh, J.Y.-C., Shen, X., Xu, C., Neyer, M., Spina, B., Almualla, M., Zhao, Y., 2024. The THESAN project: public data release of radiation-hydrodynamic simulations matching reionization-era JWST observations. MNRAS 530 (4), 3765–3786. doi:10.1093/mnras/stae839, arXiv:2309.06475.

Gorski, K.M., Hivon, E., Banday, A.J., Wandelt, B.D., Hansen, F.K., Reinecke, M., Bartelmann, M., 2005. HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. Astrophys. J. 622 (2), 759–771. doi:10.1086/427976.

Grand, R.J.J., Fragkoudi, F., Gómez, F.A., Jenkins, A., Marinacci, F., Pakmor, R., Springel, V., 2024. Overview and public data release of the augmented Auriga Project: cosmological simulations of dwarf and Milky Way-mass galaxies. MNRAS 532 (2), 1814–1831. doi:10.1093/mnras/stae1598, arXiv:2401.08750.

Hadzhiyska, B., Ferraro, S., Farren, G.S., Sailer, N., Zhou, R., 2025. Missing baryons recovered: A measurement of the gas fraction in galaxies and groups with the kinematic Sunyaev-Zel'dovich effect and CMB lensing. Phys. Rev. D 112 (12), 123507. doi:10.1103/mdhz-fgj8, arXiv:2507.14136.

Hahn, O., Rampf, C., Uhlemann, C., 2021. Higher order initial conditions for mixed baryon-CDM simulations. MNRAS 503 (1), 426–445. doi:10.1093/mnras/staa3773, arXiv:2008.09124.

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. MNRAS 474 (1), 604–617. doi:10.1093/mnras/stx2792, arXiv:1708.03646.

Huško, F., Lacey, C.G., Schaye, J., Schaller, M., Nobels, F.S.J., 2022. Spin-driven jet feedback in idealized simulations of galaxy groups and clusters. MNRAS 516 (3), 3750–3772. doi:10.1093/mnras/stac2278, arXiv:2206.06402.

Katz, H., Rosdahl, J., Kimm, T., Blaizot, J., Choustikov, N., Farcy, M., Garel, T., Haehnelt, M.G., Michel-Dansac, L., Ocvirk, P., 2023. The SPHINX public data release: Forward modelling high-redshift JWST observations with cosmological radiation hydrodynamics simulations. Open J. Astrophys. 6, 44. doi:10.21105/astro.2309.03269, arXiv:2309.03269.

Kennicutt, Jr., R.C., 1998. The global Schmidt law in star-forming galaxies. ApJ 498 (2), 541–552. doi:10.1086/305588, arXiv:astro-ph/9712213.

Kondapally, R., Best, P.N., Cochrane, R.K., Sabater, J., Duncan, K.J., Hardcastle, M.J., Haskell, P., Mingo, B., Röttgering, H.J.A., Smith, D.J.B., Williams, W.L., Bonato, M., Calistro Rivera, G., Gao, F., Hale, C.L., Małek, K., Miley, G.K., Prandoni, I., Wang, L., 2022. Cosmic evolution of low-excitation radio galaxies in the LOFAR two-metre sky survey deep fields. MNRAS 513 (3), 3742–3767. doi:10.1093/mnras/stac1128, arXiv:2204.07588.

Kovač, M., Nicola, A., Bucko, J., Schneider, A., Reischke, R., Giri, S.K., Teyssier, R., Schaller, M., Schaye, J., 2025. Baryonification II: constraining feedback with X-ray and kinematic Sunyaev-Zel'dovich observations. J. Cosmology Astropart. Phys. 2025 (11), 046. doi:10.1088/1475-7516/2025/11/046, arXiv:2507.07991.

Kugel, R., Schaye, J., Schaller, M., Helly, J.C., Braspenning, J., Elbers, W., Frenk, C.S., McCarthy, I.G., Kwan, J., Salcido, J., van Daalen, M.P., Vandenbroucke, B., Bahé, Y.M., Borrow, J., Chaikin, E., Huško, F., Jenkins, A., Lacey, C.G., Nobels, F.S.J., Vernon, I., 2023. FLAMINGO: calibrating large cosmological hydrodynamical simulations with machine learning. MNRAS 526 (4), 6103–6127. doi:10.1093/mnras/stad2540, arXiv:2306.05492.

Kumar, A., Artale, M.C., Montero-Dorta, A.D., Guaita, L., Lee, K.-S., Pope, A., Schaye, J., Schaller, M., Gawiser, E., Hwang, H.S., Jeong, W.-S., Lee, J., Padilla, N., Park, C., Ramakrishnan, V., Singh, A., Yang, Y., 2025. Modeling submillimeter galaxies in cosmological simulations: Contribution to the cosmic star formation density and predictions for future surveys. A&A 698, A236. doi:10.1051/0004-6361/202553981, arXiv:2501.19327.

Le Brun, A.M.C., McCarthy, I.G., Schaye, J., Ponman, T.J., 2014. Towards a realistic population of simulated galaxy groups and clusters. MNRAS 441 (2), 1270–1290. doi:10.1093/mnras/stu608, arXiv:1312.5462.

Lemson, G., the Virgo Consortium, 2006. Halo and galaxy formation histories from the millennium simulation: Public release of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the LambdaCDM cosmogony. doi:10.48550/arXiv.astro-ph/0608019, ArXiv E-Prints, astro–ph/0608019, arXiv:astro-ph/0608019.

Lenz, D., Doré, O., Lagache, G., 2019. Large-scale maps of the cosmic infrared background from Planck. ApJ 883 (1), 75. doi:10.3847/1538-4357/ab3c2b, arXiv:1905.00426.

Lim, S., Tacchella, S., Schaye, J., Schaller, M., Helton, J.M., Kugel, R., Maiolino, R., 2024. The FLAMINGO simulation view of cluster progenitors observed in the epoch of reionization with JWST. MNRAS 532 (4), 4551–4569. doi:10.1093/mnras/stae1790, arXiv:2402.17819.

Ludlow, A.D., Fall, S.M., Schaye, J., Obreschkow, D., 2021. Spurious heating of stellar motions in simulated galactic discs by dark matter halo particles. MNRAS 508 (4), 5114–5137. doi:10.1093/mnras/stab2770, arXiv:2105.03561.

Ludlow, A.D., Fall, S.M., Wilkinson, M.J., Schaye, J., Obreschkow, D., 2023. Spurious heating of stellar motions by dark matter particles in cosmological simulations of galaxy formation. MNRAS 525 (4), 5614–5630. doi:10.1093/mnras/stad2615, arXiv:2306.05753.

Ludlow, A.D., Schaye, J., Schaller, M., Richings, J., 2019. Energy equipartition between stellar and dark matter particles in cosmological simulations results in spurious growth of galaxy sizes. MNRAS 488 (1), L123–L128. doi:10.1093/mnrasl/slz110, arXiv:1903.10110.

McAlpine, S., Helly, J.C., Schaller, M., Trayford, J.W., Qu, Y., Furlong, M., et al., 2016. The EAGLE simulations of galaxy formation: Public release of halo and galaxy catalogues. Astron. Comput. 15, 72–89. doi:10.1016/j.ascom.2016.02.004, arXiv:1510.01320.

McCarthy, I.G., Amon, A., Schaye, J., Schaan, E., Angulo, R.E., Salcido, J., Schaller, M., Bigwood, L., Elbers, W., Kugel, R., Helly, J.C., Forouhar Moreno, V.J., Frenk, C.S., McGibbon, R.J., Ondaro-Mallea, L., van Daalen, M.P., 2025. FLAMINGO: combining kinetic SZ effect and galaxy–galaxy lensing measurements to gauge the impact of feedback on large-scale structure. MNRAS 540 (1), 143–163. doi:10.1093/mnras/staf731, arXiv:2410.19905.

McCarthy, I.G., Salcido, J., Schaye, J., Kwan, J., Elbers, W., Kugel, R., Schaller, M., Helly, J.C., Braspenning, J., Frenk, C.S., van Daalen, M.P., Vandenbroucke, B., Conley, J.T., Font, A.S., Upadhye, A., 2023. The FLAMINGO project: revisiting the S_8 tension and the role of baryonic physics. MNRAS 526 (4), 5494–5519. doi:10.1093/mnras/stad3107, arXiv:2309.07959.

McCarthy, I.G., Schaye, J., Bird, S., Le Brun, A.M.C., 2017. The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology. MNRAS 465 (3), 2936–2965. doi:10.1093/mnras/stw2792, arXiv:1603.02702.

McDonald, W., Schaye, J., Kuijken, K., Helly, J., Braspenning, J., Schaller, M., 2026. The FLAMINGO Project: Exploring the X-ray–cosmic-shear cross-correlation as a probe of large-scale structure. doi:10.48550/arXiv.2602.02484, ArXiv E-Prints, arXiv:2602.02484.

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. J. Open Source Softw. 10 (111), 8252. doi:10.21105/joss.08252, arXiv:2507.22669.

Mead, A.J., Peacock, J.A., Heymans, C., Joudaki, S., Heavens, A.F., 2015. An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models. MNRAS 454 (2), 1958–1975. doi:10.1093/mnras/stv2036, arXiv:1505.07833.

Nelson, D., Pillepich, A., Ayromlou, M., Lee, W., Lehle, K., Rohr, E., Truong, N., 2024. Introducing the TNG-Cluster simulation: Overview and the physical properties of the gaseous intracluster medium. A&A 686, A157. doi:10.1051/0004-6361/202348608, arXiv:2311.06338.

Nelson, D., Pillepich, A., Genel, S., Vogelsberger, M., Springel, V., Torrey, P., Rodriguez-Gomez, V., Sijacki, D., Snyder, G.F., Griffen, B., Marinacci, F., Blecha, L., Sales, L., Xu, D., Hernquist, L., 2015. The illustris simulation: Public data release. Astron. Comput. 13, 12–37. doi:10.1016/j.ascom.2015.09.003, arXiv:1504.00362.

Nelson, D., Pillepich, A., Springel, V., Pakmor, R., Weinberger, R., Genel, S., Torrey, P., Vogelsberger, M., Marinacci, F., Hernquist, L., 2019a. First results from the TNG50 simulation: galactic outflows driven by supernovae and black hole feedback. MNRAS 490 (3), 3234–3261. doi:10.1093/mnras/stz2306, arXiv:1902.05554.

Nelson, D., Springel, V., Pillepich, A., Rodriguez-Gomez, V., Torrey, P., Genel, S., Vogelsberger, M., Pakmor, R., Marinacci, F., Weinberger, R., Kelley, L., Lovell, M., Diemer, B., Hernquist, L., 2019b. The IllustrisTNG simulations: public data release. Comput. Astrophys. Cosmol. 6 (1), 2. doi:10.1186/s40668-019-0028-x, arXiv:1812.05609.

Oku, Y., Nagamine, K., 2024. Osaka feedback model. III. Cosmological simulation CROCODILE. ApJ 975 (2), 183. doi:10.3847/1538-4357/ad77d3, arXiv:2401.06324.

Pakmor, R., Springel, V., Coles, J.P., Guillet, T., Pfrommer, C., Bose, S., Barrera, M., Delgado, A.M., Ferlito, F., Frenk, C., Hadzhiyska, B., Hernández-Aguayo, C., Hernquist, L., Kannan, R., White, S.D.M., 2023. The MillenniumTNG Project: the hydrodynamical full physics simulation and a first look at its galaxy clusters. MNRAS 524 (2), 2539–2555. doi:10.1093/mnras/stac3620, arXiv:2210.10060.

Pillepich, A., Nelson, D., Hernquist, L., Springel, V., Pakmor, R., Torrey, P., Weinberger, R., Genel, S., Naiman, J.P., Marinacci, F., Vogelsberger, M., 2018. First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. MNRAS 475 (1), 648–675. doi:10.1093/mnras/stx3112, arXiv:1707.03406.

Pizzati, E., Hennawi, J.F., Schaye, J., Schaller, M., Eilers, A.-C., Wang, F., Frenk, C.S., Elbers, W., Helly, J.C., Mackenzie, R., Matthee, J., Bordoloi, R., Kashino, D., Naidu, R.P., Yue, M., 2024. A unified model for the clustering of quasars and galaxies at $z \approx 6$. MNRAS 534 (4), 3155–3175. doi:10.1093/mnras/stae2307, arXiv:2403.12140.

Planck Collaboration, Ade, P.A.R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., et al., 2016. Planck 2015 results. XV. Gravitational lensing. A&A 594, A15. doi:10.1051/0004-6361/201525941, arXiv:1502.01591.

Ploeckinger, S., Schaye, J., 2020. Radiative cooling rates, ion fractions, molecule abundances, and line emissivities including self-shielding and both local and metagalactic radiation fields. MNRAS 497 (4), 4857–4883. doi:10.1093/mnras/staa2172, arXiv:2006.14322.

Pontzen, A., Roškar, R., Stinson, G., Woods, R., 2013. pynbody: N-Body/SPH analysis for python. Astrophysics Source Code Library, record ascl:1305.002. https://ui.adsabs.harvard.edu/abs/2013ascl.soft05002P.

Popesso, P., Marini, I., Dolag, K., Lamer, G., Csizi, B., Biffi, V., Robothan, A., Bravo, M., Biviano, A., Vladutescu-Zopp, S., Lovisari, L., Ettori, S., Angelinelli, M., Driver, S., Toptun, V., Dev, A., Mazengo, D., Merloni, A., Zhang, Y., Comparat, J., Ponti, G., Mroczkowski, T., Bulbul, E., 2025. Average X-ray properties of galaxy groups: From Milky Way-like halos to massive clusters. A&A 704, A278. doi:10.1051/0004-6361/202453255, arXiv:2411.17120.

Predehl, P., Andritschke, R., Arefiev, V., Babyshkin, V., Batanov, O., Becker, W., et al., 2021. The eROSITA X-ray telescope on SRG. A&A 647, A1. doi:10.1051/0004-6361/202039313, arXiv:2010.03477.

Robertson, N.C., Heymans, C., Zuntz, J., Burger, P., Leonard, C.D., McCarthy, I.G., Paine, J.G., Salcido, J., Šarčević, N., Schaller, M., Schaye, J., van Daalen, M.P., 2026. Impact and interplay of ΛCDM analysis choices for LSST cosmic shear. doi:10.48550/arXiv.2602.21863, ArXiv E-Prints, arXiv:2602.21863.

Schaller, M., Borrow, J., Draper, P.W., Ivkovic, M., McAlpine, S., Vandenbroucke, B., et al., 2024. SWIFT: A modern highly-parallel gravity and smoothed particle hydrodynamics solver for astrophysical and cosmological applications. MNRAS 530 (2), 2378–2419. doi:10.1093/mnras/stae922, arXiv:2305.13380.

Schaller, M., Schaye, J., Kugel, R., Broxterman, J.C., van Daalen, M.P., 2025. The FLAMINGO project: baryon effects on the matter power spectrum. MNRAS 539 (2), 1337–1351. doi:10.1093/mnras/staf569, arXiv:2410.17109.

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., Ludlow, A., Nobels, F.S.J., 2026. The COLIBRE project: cosmological hydrodynamical simulations of galaxy formation and evolution. MNRAS 548 (1), stag375. doi:10.1093/mnras/stag375, arXiv:2508.21126.

Schaye, J., Crain, R.A., Bower, R.G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C.S., McCarthy, I.G., Helly, J.C., Jenkins, A., Rosas-Guevara, Y.M., White, S.D.M., Baes, M., Booth, C.M., Camps, P., Navarro, J.F., Qu, Y., Rahmati, A., Sawala, T., Thomas, P.A., Trayford, J., 2015. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. MNRAS 446 (1), 521–554. doi:10.1093/mnras/stu2058, arXiv:1407.7040.

Schaye, J., Dalla Vecchia, C., 2008. On the relation between the Schmidt and Kennicutt-Schmidt star formation laws and its implications for numerical simulations. MNRAS 383 (3), 1210–1222. doi:10.1111/j.1365-2966.2007.12639.x, arXiv:0709.0292.

Schaye, J., Kugel, R., Schaller, M., Helly, J.C., Braspenning, J., Elbers, W., McCarthy, I.G., van Daalen, M.P., Vandenbroucke, B., Frenk, C.S., Kwan, J., Salcido, J., Bahé, Y.M., Borrow, J., Chaikin, E., Hahn, O., Huško, F., Jenkins, A., Lacey, C.G., Nobels, F.S.J., 2023. The FLAMINGO project: cosmological hydrodynamical simulations for large-scale structure and galaxy cluster surveys. MNRAS 526 (4), 4978–5020. doi:10.1093/mnras/stad2419, arXiv:2306.04024.

Schneider, A., Teyssier, R., 2015. A new method to quantify the effects of baryons on the matter power spectrum. J. Cosmology Astropart. Phys. 2015 (12), 049. doi:10.1088/1475-7516/2015/12/049, arXiv:1510.06034.

Semboloni, E., Hoekstra, H., Schaye, J., 2013. Effect of baryonic feedback on two- and three-point shear statistics: prospects for detection and improved modelling. MNRAS 434 (1), 148–162. doi:10.1093/mnras/stt1013, arXiv:1210.7303.

Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M.P., McCarthy, I.G., 2011. Quantifying the effect of baryon physics on weak lensing tomography. MNRAS 417 (3), 2020–2035. doi:10.1111/j.1365-2966.2011.19385.x, arXiv:1105.1075.

Siegel, J., Bigwood, L., Amon, A., McCullough, J., Yamamoto, M., McCarthy, I.G., Schaller, M., Schaye, J., Schneider, A., 2026. The suppression of the matter power spectrum: strong feedback from X-ray gas mass fractions, kSZ effect profiles, and galaxy-galaxy lensing. MNRAS 549 (4), stag993. doi:10.1093/mnras/stag993, arXiv:2512.02954.

Springel, V., Di Matteo, T., Hernquist, L., 2005a. Black holes in galaxy mergers: The formation of red elliptical galaxies. ApJ 620 (2), L79–L82. doi:10.1086/428772, arXiv:astro-ph/0409436.

Springel, V., White, S.D.M., Jenkins, A., Frenk, C.S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J.A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., Pearce, F., 2005b. Simulations of the formation, evolution and clustering of galaxies and quasars. Nature 435 (7042), 629–636. doi:10.1038/nature03597, arXiv:astro-ph/0504097.

The EAGLE team, 2017. The EAGLE simulations of galaxy formation: Public release of particle data. doi:10.48550/arXiv.1706.09899, ArXiv E-Prints, arXiv:1706.09899.

Villaescusa-Navarro, F., Genel, S., Anglés-Alcázar, D., Perez, L.A., Villanueva-Domingo, P., Wadekar, D., et al., 2023. The CAMELS Project: Public data release. ApJS 265 (2), 54. doi:10.3847/1538-4365/acbf47, arXiv:2201.01300.

Wetzel, A., Hayward, C.C., Sanderson, R.E., Ma, X., Anglés-Alcázar, D., Feldmann, R., et al., 2023. Public data release of the FIRE-2 cosmological zoom-in simulations of galaxy formation. ApJS 265 (2), 44. doi:10.3847/1538-4365/acb99a, arXiv:2202.06969.

Wiersma, R.P.C., Schaye, J., Theuns, T., Dalla Vecchia, C., Tornatore, L., 2009. Chemical enrichment in cosmological, smoothed particle hydrodynamics simulations. MNRAS 399 (2), 574–600. doi:10.1111/j.1365-2966.2009.15331.x, arXiv:0902.1535.

Wright, A.H., Stölzner, B., Asgari, M., Bilicki, M., Giblin, B., Heymans, C., et al., 2025. KiDS-Legacy: Cosmological constraints from cosmic shear with the complete Kilo-Degree Survey. A&A 703, A158. doi:10.1051/0004-6361/202554908, arXiv:2503.19441.

Yang, T., McCarthy, I.G., McCarthy, F., Bolliet, B., Chluba, J., Coulton, W., Helly, J.C., Schaller, M., Schaye, J., 2026. Self-consistent secondary cosmic microwave background anisotropies and extragalactic foregrounds in the FLAMINGO simulations. MNRAS 548 (2), stag625. doi:10.1093/mnras/stag625, arXiv:2512.09891.