

The COLIBRE project: cosmological hydrodynamical simulations of galaxy formation and evolution

Joop Schaye¹,^{1*} Evgenii Chaikin¹, Matthieu Schaller^{1,2}, Sylvia Ploekinger³, Filip Huško¹, Robert J. McGibbon¹, James W. Trayford⁴, Alejandro Benítez-Llambay⁵, Camila Correa¹, Carlos S. Frenk⁶, Alexander J. Richings^{7,8}, Victor J. Forouhar Moreno¹, Yannick M. Bahé^{9,10}, Josh Borrow¹¹, Anna Durrant^{1,12}, Andrea Gebek¹³, John C. Helly⁶, Adrian Jenkins¹⁴, Cedric G. Lacey¹⁵, Aaron Ludlow¹⁴ and Folkert S. J. Nobels^{1,15}

¹Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands

²Lorentz Institute for Theoretical Physics, Leiden University, PO Box 9506, NL-2300 RA Leiden, the Netherlands

³Department of Astrophysics, University of Vienna, Türkenschanzstrasse 17, A-1180 Vienna, Austria

⁴Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Burnaby Road, Portsmouth PO1 3FX, UK

⁵Dipartimento di Fisica G. Occhialini, Università degli Studi di Milano Bicocca, Piazza della Scienza, 3 I-20126 Milano MI, Italy

⁶Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

⁷Centre for Data Science, Artificial Intelligence and Modelling, University of Hull, Cottingham Road, Hull HU6 7RX, UK

⁸E. A. Milne Centre for Astrophysics, University of Hull, Cottingham Road, Hull HU6 7RX, UK

⁹School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK

¹⁰Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland

¹¹Department of Physics and Astronomy, University of Pennsylvania, 209 South 33rd Street, Philadelphia, PA 19104, USA

¹²Astrophysics Research Institute, Liverpool John Moores University, Liverpool L3 5RF, UK

¹³Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, B-9000 Gent, Belgium

¹⁴International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

¹⁵Netherlands Organisation for Applied Scientific Research (TNO), Molengraaffsingel 8, NL-2629 JD Delft, the Netherlands

Accepted 2026 February 20. Received 2026 January 10; in original form 2025 August 27

ABSTRACT

We present the COLIBRE galaxy formation model and the COLIBRE suite of cosmological hydrodynamical simulations. COLIBRE includes new models for radiative cooling, dust grains, star formation, stellar mass loss, turbulent diffusion, pre-supernova stellar feedback, supernova feedback, supermassive black holes, and active galactic nucleus (AGN) feedback. The multiphase interstellar medium is explicitly modelled without a pressure floor. Hydrogen and helium are tracked in non-equilibrium, with their contributions to the free electron density included in metal-line cooling calculations. The chemical network is coupled to a dust model that tracks three grain species and two grain sizes. In addition to the fiducial thermally driven AGN feedback, a subset of simulations uses black hole spin-dependent hybrid jet/thermal AGN feedback. To suppress spurious transfer of energy from dark matter to stars, dark matter is supersampled by a factor 4, yielding similar dark matter and baryonic particle masses. The subgrid feedback model is calibrated to match the observed $z \approx 0$ galaxy stellar mass function, galaxy sizes, and black hole masses in massive galaxies. The COLIBRE suite includes three resolutions, with particle masses of $\sim 10^5$, 10^6 , and $10^7 M_\odot$ in cubic volumes of up to 100, 200, and 400 cMpc on a side, respectively. The largest runs use 136 billion (5×3008^3) particles. We describe the model, assess its strengths and limitations, and present both visual impressions and quantitative results. Comparisons with various low-redshift galaxy observations generally show very good numerical convergence and excellent agreement with the data.

Key words: methods: numerical – galaxies: evolution – galaxies: formation – cosmology: theory.

1 INTRODUCTION

Hydrodynamical simulations following the concurrent formation and evolution of cosmological structures and the galaxies that they contain have become a central part of research

in extragalactic astronomy and cosmology. Starting from redshifts $z \sim 10^2$ using initial conditions (ICs) inferred from the cosmic microwave background (CMB) and surveys of large-scale structure, such simulations follow the growth of density fluctuations and the formation and evolution of galaxies (for recent reviews, see e.g. M. Vogelsberger et al. 2020; R. A. Crain & F. van de Voort 2023; R. Feldmann & R. Bieri 2025). They serve a wide range of purposes. These include gaining insight into

* E-mail: schaye@strw.leidenuniv.nl

astrophysical processes, interpreting observations of galaxies and diffuse gas, investigating the effect of baryon PCs on cosmological probes, testing data analysis techniques, and guiding the design of new observational campaigns.

Contemporary cosmological simulations of galaxy formation in representative volumes (typically $25 \lesssim L/\text{Mpc} \lesssim 300$ on a side) include HorizonAGN (Y. Dubois et al. 2014), EAGLE (J. Schaye et al. 2015; R. A. Crain et al. 2015), IllustrisTNG (A. Pillepich et al. 2018a, b; D. Nelson et al. 2019b), SIMBA (R. Davé et al. 2019), ASTRID (S. Bird et al. 2022), CROCODILE (Y. Oku & K. Nagamine 2024), and (X)FABLE (L. Bigwood et al. 2025). These simulations produce populations of galaxies with properties that agree well in many respects with observational data. Larger volume, but lower resolution simulations such as BAHAMAS (I. G. McCarthy et al. 2017), MillenniumTNG (R. Pakmor et al. 2023), FLAMINGO (J. Schaye et al. 2023; R. Kugel et al. 2023), and Magneticum Pathfinder (K. Dolag et al. 2025) reproduce many (mostly spatially unresolved) galaxy properties, as well as a variety of observations of large-scale structure and clusters of galaxies. Collections of simulations that each zoom into a region centred on a single galaxy or galaxy group, such as APOSTLE (T. Sawala et al. 2016), Auriga (R. J. J. Grand et al. 2017), FIRE (P. F. Hopkins et al. 2018, 2023), NIHAO-UHD (T. Buck et al. 2020), HESTIA (N. I. Libeskind et al. 2020), and ARTEMIS (A. S. Font et al. 2020), can be used to compare with more detailed observations of galaxies. Overall, these models are much more realistic than their earlier counterparts. However, significant discrepancies remain between the data and each of the aforementioned simulations.

While many high-resolution zoom-in cosmological simulations allow gas to cool to temperatures $T \ll 10^4$ K (e.g. P. F. Hopkins et al. 2018, 2023; O. Agertz et al. 2021; Y. Dubois et al. 2021; M. Mina et al. 2021; E. Applebaum et al. 2021; T. A. Gutcke et al. 2022; M. P. Rey et al. 2025), most simulations of representative volumes that have been run to redshift $z = 0$ suppress this cold gas phase by imposing an effective equation of state, or a temperature or pressure floor, on gas with densities typical of the interstellar medium (ISM), i.e. total hydrogen number densities $n_{\text{H}} \gtrsim 10^{-1} \text{ cm}^{-3}$. A notable exception is the 22.1 comoving Mpc (cMpc) FIREbox simulation (R. Feldmann et al. 2023). FIREbox predicts realistic cold gas masses, and it represents an important step towards applying sophisticated physical models to statistically representative volumes. However, even though FIREbox employs the FIRE-2 galaxy formation model, which is consistent with the observationally inferred galaxy mass–halo mass relation when used in higher resolution zoom-in simulations (P. F. Hopkins et al. 2018), FIREbox itself does not predict realistic galaxy stellar masses (see Section 7.1.1). This discrepancy underscores the importance of testing galaxy formation models in representative volumes. FIREbox does not include active galactic nucleus (AGN) feedback, has no model for dust grain evolution, and computes the abundances of atomic and molecular hydrogen using an equilibrium model instead of tracking them directly. Another exception is the 25 cMpc Romulus25 simulation (M. Tremmel et al. 2017). While Romulus25 models many aspects of galaxy formation, and includes a model for the dynamical friction experienced by black holes (BHs) that is more physical than those used in most other simulations including those presented here, it omits some of the dominant cooling channels. In particular, Romulus25 does not account for molecular cooling or dust processes and, for $T > 10^4$ K, it does not include metal-line cooling.

Simulations of representative volumes that stop at high redshift can more easily afford the computational expense associated

with the direct modelling of cold gas. However, studies based on such simulations run the risk of reproducing high- z observations with models that would not work at lower redshift. This risk is a real worry because models that suffer from numerical overcooling, and which thus produce present-day galaxies that are too massive, too old, and too spatially concentrated, will tend to predict high- z galaxies that undergo a more intense phase of ‘compaction’ (see e.g. A. Zolotov et al. 2015, for a simulation study of compaction in high- z galaxies). Such galaxies are more clumpy, have larger masses, reach higher star formation rates (SFRs), but quench earlier, which may, for example, help reproduce observations of submm galaxies (e.g. A. Kumar et al. 2025) and the abundance of massive, quenched galaxies (e.g. A. C. Carnall et al. 2024; W. M. Baker et al. 2025).

The successful reproduction of observations of galaxy populations by cosmological simulations relies on the use of subgrid models for unresolved processes, in particular galactic winds driven by feedback from star formation and AGN feedback from accreting supermassive BHs. The subgrid models for feedback are designed to circumvent ‘numerical overcooling’. This overcooling is understood to result from limited resolution, which causes the energy/momentum injected by stars or AGN to initially be distributed over too much gas mass, leading to too low post-shock temperatures and therefore too short cooling times (C. Dalla Vecchia & J. Schaye 2012).

Methods to mitigate numerical overcooling include decoupling the winds from the hydrodynamics in the ISM (V. Springel & L. Hernquist 2003), which is, for example, used in IllustrisTNG, SIMBA, and ASTRID, or injecting the feedback energy less often but in larger amounts (C. M. Booth & J. Schaye 2009; C. Dalla Vecchia & J. Schaye 2012), as is, for example, used in EAGLE for stellar and AGN feedback and in IllustrisTNG and SIMBA for AGN feedback. Although such subgrid prescriptions can suppress overcooling, they also introduce new artefacts and problems. Wind particles that are temporarily decoupled from hydrodynamical forces cannot heat or inject turbulence in the ISM and ignore the work done by winds in the ISM (e.g. C. Dalla Vecchia & J. Schaye 2008). The alternative of launching particles at high velocities or heating them to high temperatures can cause the feedback to be too bursty, can create overly large bubbles in the ISM, and can result in poor sampling of the feedback processes (e.g. Y. M. Bahé et al. 2016). Using either method, overcooling may still occur downstream of the wind injection, or, at the other extreme, radiative cooling losses may be underestimated.

Subgrid prescriptions for feedback necessarily introduce free parameters that are more numerical than physical. These parameters help control the effectiveness of the feedback, but their values cannot be predicted from first principles. Examples include the initial wind mass loading, velocity, or temperature increase, and, if applicable, the conditions that must be satisfied for wind particles to recouple with the hydrodynamics. The values of at least some of these parameters are calibrated, either explicitly or implicitly, to broadly reproduce a chosen set of observables. For example, EAGLE is calibrated to reproduce the $z \approx 0$ galaxy stellar mass function (SMF), size–mass relation (SMR) of disc galaxies, and BH masses for massive galaxies (R. A. Crain et al. 2015). IllustrisTNG uses these same constraints, as well as the stellar-to-halo-mass relation, the mass–metallicity relation, the mass scale above which star formation is quenched, the halo gas fraction in galaxy groups, and the cosmic star formation history (A. Pillepich et al. 2018a). It is important to note that calibration is not limited to the choice of parameter values, because the results

tend to be sensitive to tacit numerical choices, such as the way the feedback energy is distributed among the gas resolution elements surrounding young stars (e.g. E. Chaikin et al. 2022).

As discussed in J. Schaye et al. (2015), there are multiple reasons why the calibration of the subgrid model for feedback should generally depend on numerical resolution. First, because a higher resolution allows one to directly simulate smaller scales and therefore new phenomena. Secondly, because the implications of keeping the subgrid parameters fixed depend on arbitrary choices in the parametrization. For example, if we use the wind velocity or temperature increase as a parameter and keep it fixed, then a higher mass resolution implies a smaller amount of injected energy per feedback event. If, on the other hand, we use the energy per event as the parameter that is kept fixed, then a higher mass resolution corresponds to a higher velocity or temperature.

The necessity of calibration weakens the *ab initio* predictive power of hydrodynamical simulations relative to, say, dark-matter-only (DMO) simulations. In addition, since baryonic processes and subgrid models are not scale free and simulations cannot resolve all relevant scales, a change in resolution effectively changes the galaxy formation model, regardless of whether the subgrid model is recalibrated, making tests of numerical convergence harder to interpret (e.g. A. D. Ludlow et al. 2020).

A key motivation for preventing the formation of a cold ISM phase is to prevent artificial fragmentation due to the failure to resolve the Jeans scales (e.g. M. R. Bate & A. Burkert 1997; J. Schaye & C. Dalla Vecchia 2008; B. E. Robertson & A. V. Kravtsov 2008). However, in the cold ISM turbulent velocities are expected to be supersonic, which means the effective Jeans scales are typically larger than the thermal ones. Furthermore, since gravity is typically softened on scales greater than the Jeans length in the cold phase, artificial fragmentation is generally not a bottleneck in practice, provided the gravitational and hydrodynamic force resolutions are properly adjusted (S. Ploekinger et al. 2024).

Paradoxically, imposing a pressure floor may in fact make the simulations more sensitive to resolution. This is because in the absence of a cold phase with a small volume-filling factor, the ISM is too smooth. If in reality the cold phase accounts for a large fraction of the interstellar gas mass, then keeping this gas in a more diffuse, volume-filling phase results in the cooling time of the hot phase created by feedback becoming too short, thereby exacerbating (downstream) numerical overcooling.

For Lagrangian methods, it is conventional to quote the gravitational softening length when indicating the spatial resolution. While this common practice may seem sensible for DMO simulations, it is at best questionable for hydrodynamical simulations whose physical approximations break down, or subgrid models kick in, on scales larger than the softening length. For example, the ISM in simulations that do not include a cold, molecular gas phase is less clumpy. One could therefore argue that, irrespective of their formal resolution, the actual resolution of such simulations is similar to the Jeans scale in warm ($T \sim 10^4$ K) gas with density $n_H \sim 10^{-1} \text{ cm}^{-3}$, which is ~ 1 kpc, because the physics is clearly not modelled correctly on smaller scales.

In fact, the softening length is a poor estimate of the spatial resolution even for purely gravitational processes because gravitational interactions between simulation particles and fluctuations in the potential due to particle noise tend to randomize particle angular momenta and drive the system to equipartition, thus transferring energy from high- to low-kinetic energy particles (e.g. J. A. Sellwood 2013, 2015). Since dark matter (DM) particles on halo orbits tend to move faster than the particles orbiting in

a galaxy disc, at least in the direction perpendicular to the disc, in simulations halo particles tend to heat up the disc. The stellar halo similarly heats up the disc, but this is less important because the DM halo contains more energy than its stellar counterpart. Because gravitational softening suppresses fluctuations in the potential due to discreteness effects, a larger softening length can actually improve the effective spatial resolution, for both DMO (A. D. Ludlow, J. Schaye & R. Bower 2019b) and hydrodynamical simulations (A. D. Ludlow et al. 2020).

The spurious transfer of energy is exacerbated by the fact that nearly all cosmological simulations use the same number of DM particles as baryonic particles, or fewer DM particles than baryonic particles. A counterexample is Romulus25 (M. Tremmel et al. 2017), which supersamples the DM by a factor 3.375. Not supersampling the DM implies that the DM particles tend to be at least a factor $(\Omega_m - \Omega_b)/\Omega_b \approx 5.3$ times more massive than the baryonic particles. This difference in particle masses causes energy transfer from the DM to the baryons even if they were to follow the same phase-space distribution (A. D. Ludlow et al. 2019a). Conversely, using equal-mass DM and baryonic particles does not eliminate spurious energy transfer if the two components follow different phase-space distributions, as is generally the case (A. D. Ludlow et al. 2023).

In the real Universe, this type of energy transfer is negligible if the DM particle mass is microscopic, because then the particle sampling noise in the gravitational potential is negligible and the associated relaxation time is very long. Although spurious transfer of energy to gas is typically unimportant because it is radiated away, it can be problematic for the stellar component, particularly for older stars. As a result, galaxies, particularly their inner regions, tend to get puffed up, artificially increasing their size (A. D. Ludlow et al. 2021) and reducing the disc-to-bulge ratio (M. J. Wilkinson et al. 2023). For the half-mass radius, the effect only becomes negligible when there are $\gtrsim 10^5$ DM particles in the halo (A. D. Ludlow et al. 2023). Because softening suppresses the spurious transfer of energy, the effective resolution can worsen when the gravitational softening length is reduced. The most direct way to suppress spurious energy transfer is to increase the number of DM particles.

Motivated by these considerations, we present here the COLIBRE¹ suite of cosmological hydrodynamical simulations. COLIBRE is the latest in the series of simulation projects that includes OWLS (J. Schaye et al. 2010), Cosmo-OWLS (A. M. C. Le Brun et al. 2014), EAGLE (J. Schaye et al. 2015), BAHAMAS (I. G. McCarthy et al. 2017), and FLAMINGO (J. Schaye et al. 2023). Its objectives are most comparable to those of EAGLE, but it improves on that model in many ways. In terms of numerics, the main improvements include the suppression of the spurious transfer of energy discussed above by using four times more DM than baryonic particles, the use of up to 20 times more particles in total (5×3008^3 compared to 2×1504^3 for EAGLE), the use of a new simulation code (SWIFT; M. Schaller et al. 2024), and a new structure finder (HBT-HERONS; V. J. Forouhar Moreno et al. 2025).

¹COLIBRE is a project of the Virgo consortium for cosmological super-computer simulations. The acronym stands for **COL**d **IS**m and **B**etter **RE**solution, where the ‘better resolution’ refers to the larger ratio of DM to baryonic particles compared with most previous simulations, which increases the effective resolution for a fixed baryonic particle mass and gravitational softening length. The explicit simulation of the cold ISM also increases the physical resolution relative to simulations that use the same particle mass and force resolution, but impose a pressure floor on the ISM.

In terms of physics, all subgrid prescriptions are new or have been significantly modified. The most important new features are that cooling below 10^4 K is allowed (i.e. no equation of state is imposed on the ISM), the growth and composition of dust grains are simulated, self-shielding and molecules are included and coupled to the dust physics, pre-supernova (SN) stellar feedback is included, and the sampling of SN and AGN feedback is improved. In addition, besides the fiducial model that uses thermally driven AGN feedback, we include simulations that employ hybrid thermally driven/kinetic jet-driven AGN feedback, where the hybrid model tracks the BH spin. The availability of simulations that achieve similar levels of agreement with observed galaxy populations, but which employ different types of feedback from accreting BHs, can provide insight into the physical mechanisms involved and the systematic uncertainty associated with the critically important but poorly understood process of AGN feedback.

As was the case for EAGLE, the subgrid feedback is calibrated to reproduce as closely as possible the observed $z \approx 0$ galaxy SMF, galaxy half-mass radii, and BH masses of massive galaxies. While this calibration was done by hand for EAGLE, COLIBRE uses machine learning to perform the calibration to the mass function and sizes at our lowest resolution, similar to what was done in R. Kugel et al. (2023) for FLAMINGO. We then make small adjustments by hand when we increase the resolution or switch from thermal to hybrid AGN feedback. The calibration process for the fiducial simulations is detailed in the companion paper (E. Chaikin et al. 2025a), while the adjustments made for the hybrid AGN feedback simulations are presented in F. Huško et al. (2026). In addition to the parameters associated with stellar and AGN energy feedback, the dust and chemistry modules include parameters whose values were chosen based on comparison of test simulations with observations. The effects of varying these parameters are demonstrated in J. W. Trayford et al. (2025) and C. Correa et al. (in preparation), respectively. The radiative cooling module includes parameters associated with local radiation fields and self-shielding whose effect is described in S. Ploekinger et al. (2025). Finally, the effects of the parameter choices associated with the subgrid prescriptions for star formation, pre-SN stellar feedback, and SNIa (type Ia SNe) are discussed in F. S. J. Nobels et al. (2024), A. Benítez-Llambay et al. (2025), and F. S. J. Nobels et al. (in preparation), respectively.

The COLIBRE suite presented here has a ‘wedding cake’ design, with larger volumes using a progressively lower resolution (see Fig. 1 and Section 5 for visual impressions). As we find the model to converge better than EAGLE, we use a wider range of resolutions. The two flagship simulations introduced here each follow 5×3008^3 particles in cubic volumes of 200 and 400 cMpc on a side, yielding a particle mass, for both baryons and DM, of $\sim 10^6$ and $\sim 10^7 M_\odot$, respectively. We also present a higher resolution model with particle mass $\sim 10^5 M_\odot$ in a 25 cMpc box that has reached $z = 0$, while, at the time of writing, 50 and 100 cMpc versions of the same model are still running. We will refer to the high, intermediate, and low COLIBRE resolutions as m5, m6, and m7, respectively.

We view COLIBRE as a stepping stone between very high-resolution simulations of ISM patches (e.g. C.-G. Kim et al. 2023), of isolated galaxies (e.g. J. Rosdahl et al. 2015; V. A. Semenov, A. V. Kravtsov & N. Y. Gnedin 2017; M. C. Smith et al. 2021; A. J. Richings et al. 2022; H. Katz et al. 2022), or of individual (dwarf) galaxies in a cosmological setting (e.g. M. Mina et al. 2021; E. Applebaum et al. 2021; T. A. Gutcke et al. 2022; M. P. Rey et al. 2025), and the previous generation of simulations of

representative volumes. Although COLIBRE cannot resolve individual molecular clouds, it captures key effects of a cold interstellar phase that has a small volume-filling factor while comprising a large fraction of the ISM mass. Including a cold phase also opens up an entirely new set of questions that the simulations can address, such as the evolution and distribution of atomic and molecular hydrogen, and the effect of dust on chemistry and observables on scales exceeding molecular cloud sizes. Although simulations that pressurize the ISM and do not explicitly model the evolution of dust grains have been used to investigate these questions, their analysis requires strong assumptions and relies on post-processing of gas and dust distributions that are unlikely to be particularly realistic.

This paper is organized as follows. Section 2 summarizes the simulation methods, including the gravity and hydrodynamics solver, ICs and cosmological parameters, identification of haloes and galaxies, and data products. Section 3 presents the subgrid prescriptions. Sections 2 and 3 are rather technical sections and can be skipped by readers mostly interested in the results, although we recommend reading at least the brief overview of the subgrid prescriptions at the start of Section 3. Section 4 gives an overview of the simulation runs. In order to give some intuition for what the simulations look like, we present visual impressions in Section 5 and, in Section 6, we characterize the physical conditions using phase diagrams and the properties of the gas from which stars are born. We compare the simulation results with the observations used to calibrate the stellar and AGN feedback in Section 7.1. In Section 7.2, we test predictions for other observations, including the cosmic star formation history; the $z = 0$ relations between galaxy stellar mass and SFRs, quenched fractions, HI masses, molecular masses, gas and stellar metallicities, dust masses and grain sizes; and X-ray luminosities from the circumgalactic medium (CGM) as a function of halo mass. In general, we find very good numerical convergence and excellent agreement with the data. In Section 8, we compare the predictions of the simulations using hybrid and fiducial AGN feedback, finding only small differences and similar levels of agreement with the data as shown in Section 7 for the fiducial simulations. Finally, we summarize and discuss our main results in Section 9.

Images, videos, and interactive visualizations can be found on the COLIBRE website.²

2 SIMULATION METHODS

In this section, we discuss the gravity and hydrodynamics solver (Section 2.1), the ICs and the cosmological parameters (Section 2.2), the (sub)halo finding (Section 2.3), and the main available data products (Section 2.4). The subgrid models for unresolved processes are discussed in the next section (Section 3).

2.1 The gravity and hydrodynamics solvers

The simulations were performed using a version of the SWIFT code³ (M. Schaller et al. 2024) that combines the open-source version 2025.04 with new subgrid modules. SWIFT uses task-based parallelism within compute nodes and non-blocking message passing interface between compute nodes.

²<https://colibre-simulations.org/>

³www.swiftsim.com

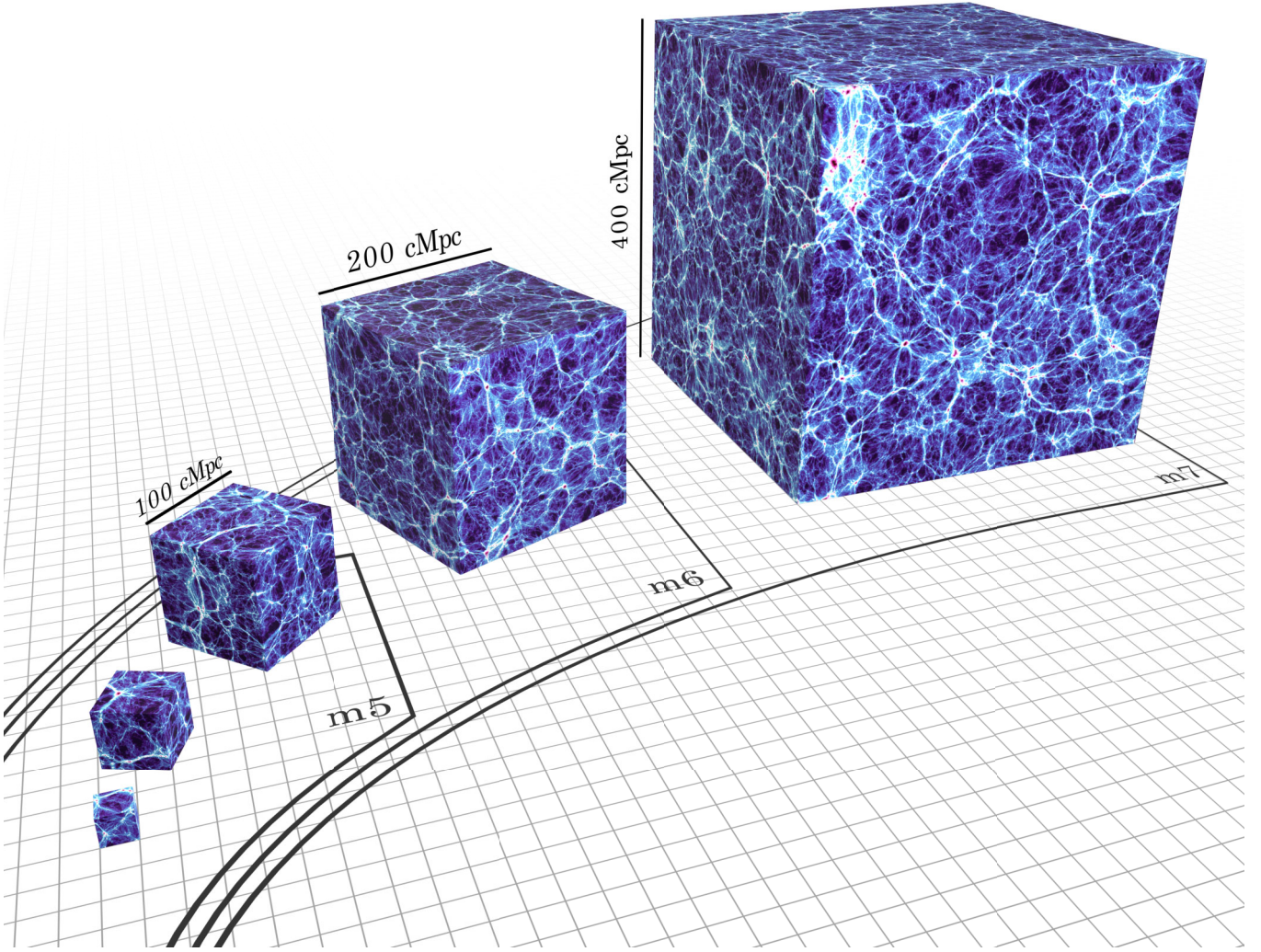


Figure 1. The five cubic COLIBRE boxes, which have side lengths ranging from 25 to 400 cMpc. Colour shows total surface density (in 5 Mpc thick faces) at $z = 0$. The volumes available at high (m5), intermediate (m6), and low (m7) resolution are indicated. Note, however, that at the time of writing the 50 and 100 Mpc high-resolution simulations have not yet reached redshift $z = 0$.

The simulation contains collisionless cold DM (CDM) and baryonic particles. We use four times more CDM than baryonic particles, which helps suppress spurious transfer of energy from CDM to stellar particles and hence improves the effective resolution for the baryonic component (A. D. Ludlow et al. 2019a, 2021, 2023). The baryonic particles are initially all gas particles, but during the simulation they can be converted into stellar or BH particles, which are both collisionless. Stellar and gas particles can, respectively, lose and gain mass due to stellar mass loss (see Section 3.4). While mass loss associated with stellar evolution will not reduce the masses of stellar particles by more than a factor of ≈ 2 , there is no limit to the mass that can be gained by gas particles. To prevent large differences between the masses of baryonic particles, gas particles whose mass exceeds $4m_g$ are split into two nearly cospatial equal mass particles, where m_g is the mean initial gas particle mass. Photons, relativistic neutrinos, and dark energy only affect the expansion history, while non-relativistic neutrinos also influence the growth of structure during the simulation (see below).

The short- and long-range gravitational forces are computed using a fourth-order fast multipole method (Greengard

& Rokhlin 1987; H. Cheng, L. Greengard & V. Rokhlin 1999; W. Dehnen 2014) and a particle-mesh method solved in Fourier space, respectively, following the force splitting approach of J. S. Bagla & S. Ray (2003). The accuracy of the gravity solver is mainly controlled by an adaptive acceptance criterion for the fast multipole method similar to that proposed by W. Dehnen (2014).

The equations of hydrodynamics are solved with smoothed particle hydrodynamics (SPH) using the SPHENIX implementation (J. Borrow et al. 2022). SPHENIX, which was specifically designed for galaxy formation simulations, uses a density-energy formulation of the equations of motion combined with artificial viscosity and conduction terms. Viscosity and conduction limiters prevent spurious energy losses in feedback-generated shocks and radiative cooling events. We use a quartic spline SPH kernel with a smoothing length of $\eta = 1.2348$ in units of the local mean particle separation, $(m_g/\rho_g)^{1/3}$, where ρ_g is the gas particle's density. This choice corresponds to 64.90 weighted neighbours (we use a tolerance of ± 0.02 neighbours). Stellar (and BH) particles also require a smoothing length because they can transfer (accrete) mass and energy to their gas neighbours. To reduce the computational cost, we reduce η to 1.1642 for stellar (but not for

BH) particles, which corresponds to 54 weighted neighbours (and we use a tolerance of ± 1).

We stress that we use the same implementation and parameter values for SPHENIX as in J. Borrow et al. (2022) and refer the interested reader to that paper for the results of a variety of standard hydrodynamical test problems. The performance is satisfactory, although J. Braspennig et al. (2023) show that cloud–wind interactions converge very slowly if the density contrast is high ($\gg 10$). However, at least for the numerical resolution of the simulations presented here, we expect details of the subgrid prescriptions for feedback to dominate over differences due to the implementation of the hydrodynamics solver (e.g. C. Scanapieco et al. 2012; M. Schaller et al. 2015b).

The effects of neutrinos on the expansion rate and on the growth of density fluctuations are modelled using the semilinear method of Y. Ali-Haïmoud & S. Bird (2013), which was also used in the BAHAMAS simulations (I. G. McCarthy et al. 2018) and is compared with other methods in W. Elbers et al. (2021). A linear perturbation integrator is run on the fly to compute the evolution of neutrino density perturbations sourced by the gravitational potential of the baryons and CDM. The gravitational force due to the neutrino perturbations is then included in the total gravitational force exerted on the baryons and CDM. This method self-consistently models the mutual dynamical response of fluctuations in the baryons + CDM and fluctuations in the neutrinos, but it does not capture the effect of neutrino self-gravity on the evolution of neutrino perturbations. Capturing this effect would require the use of neutrino particles, as was done in the large-volume FLAMINGO simulations (J. Schaye et al. 2023), but the effect is negligible on galaxy scales.

Time integration is performed using a standard leapfrog scheme. The individual time-steps of the particles are set by the minimum of the gravity time-step ($\Delta t = (0.025\epsilon/a_{\text{acc}})^{1/2}$), where ϵ is the gravitational softening length and a_{acc} is the acceleration, and the Courant–Friedrichs–Lewy (R. Courant, K. Friedrichs & H. Lewy 1928) condition for hydrodynamics with parameter value 0.2. In addition, the time-steps of young stellar particles (ages < 40 Myr) are limited to be at most 1 Myr (Section 3.4), while gas particles can have their time-step limited by turbulent diffusion (Section 3.5) and BH particles by their accretion rates (Section 3.9). We activate gas particles that are directly heated/accelerated as a result of any kind of energy feedback, i.e. stellar winds, radiation pressure, H II regions, core-collapse supernovae (CCSNe, in either thermal or kinetic form), SNIa, and AGN (in both thermal and kinetic form). Following F. Durier & C. Dalla Vecchia (2012), the time-step of inactive neighbouring gas particles is limited to at most four times the minimum time-step of the neighbours in order to properly evolve the fluid even in extreme shocks. Finally, the time-step is limited by the requirement that the SPH smoothing length does not change by more than a factor of $2^{1/3}$ in successive time-steps.

Gravity is softened using the H. Wendland (1995) C2 kernel on scales smaller than the kernel size 3ϵ , where ϵ is the ‘Plummer-equivalent’ kernel size that we refer to as the softening length. All particles use the same spatially fixed, but time-evolving gravitational softening length (see Table 2). Appendix A contains a discussion of the potential consequences of gravitational softening for the gravitational collapse of gas clouds. We impose a floor on the gas particle SPH smoothing length of $10^{-8}\epsilon$, which is sufficiently small to suppress spurious subkernel instabilities (S. Ploekinger et al. 2024).

2.2 Initial conditions and cosmological parameters

The ICs are generated with MONOFONIC (O. Hahn et al. 2020; M. Michaux et al. 2021) using second-order Lagrangian perturbation theory. CDM and baryon particles are set up in a two-stage process. First, the combined mass-weighted CDM + baryon fluid is initialised as a single fluid, accounting for the presence of neutrinos. This single fluid is then split into separate components with distinct transfer functions. Linear power spectra and transfer functions are computed using CLASS (J. Lesgourgues 2011; Lesgourgues & Tram 2011). To limit cosmic variance without compromising the ability to carry out zoom-in simulations, we use ‘partially fixed ICs’ (R. E. Angulo & A. Pontzen 2016): the amplitudes of modes with $(kL)^2 < 1025$ are set to the mean variance, where k is the wavenumber and L is the side length of the cubic simulation volume. The Gaussian random fields were chosen from subregions of PANPHASIA to allow future zoom-in resimulations (A. Jenkins 2013), see Appendix B for details. The starting redshift, $z = 63$, is a compromise between the desire to limit discreteness errors, which become larger if the simulation starts earlier, and the accuracy of perturbation theory, which worsens if we start later.

The unperturbed particle load for the ICs is created as follows. The entire periodic volume is tiled with uniform cubic cells with the cell centres forming a primitive cubic lattice. An SPH particle is placed at the centre of each of these cubic cells. Equal-mass CDM particles are placed so that there are CDM particles at the vertices and in the centres of the sides of every cubic cell. By symmetry, the gravitational force acting on all particles is zero. Each cell has eight vertices and six sides. Each vertex is shared by eight cells and each side by two cells. The total number of CDM particles per SPH particle in the entire simulation volume is therefore $8/8 + 6/2 = 4$.

Following O. Hahn, C. Rampf & C. Uhlemann (2021), the perturbation theory result is then imprinted in two steps. First, growing mode adiabatic perturbations are realised by perturbing the positions and velocities of the particles using second-order Lagrangian perturbation theory. Both particle species are treated in the same way during this phase. Second, compensated isocurvature perturbations, including leading order corrections to their evolution from decaying modes, are imprinted by altering the masses and relative velocities of baryonic and CDM particles. This procedure results in rms (root-mean-square) particle mass variations of ~ 1 per cent. O. Hahn et al. (2021) show that discreteness errors are suppressed by perturbing the masses of the particles rather than their positions.

Each hydrodynamic simulation is complemented by a corresponding DMO simulation that uses the same initial phases and the same total number of particles. To create the ICs for the DMO simulation, we converted the baryonic particles from the corresponding hydrodynamic run into CDM particles. The DMO simulations therefore also use two-fluid ICs.

The values of the cosmological parameters for all runs presented here are as follows. The dimensionless *Hubble* constant, $h = 0.681$; the density parameter for the total amount of non-relativistic matter, $\Omega_{\text{m}} = 0.306$, with contributions from CDM, $\Omega_{\text{CDM}} = 0.256011$, baryons, $\Omega_{\text{b}} = 0.0486$, and non-relativistic neutrinos, $\Omega_{\nu, \text{NR}} = 0.001389$; the dark energy density parameter, $\Omega_{\Lambda} = 0.693922$; the density parameter for photons, $\Omega_{\gamma} = 0.000053$; the density parameter for relativistic neutrinos, $\Omega_{\nu, \text{R}} = 0.000025$; the amplitude of the primordial matter power spectrum, $A_{\text{s}} = 2.099 \times 10^{-9}$; and the power-law index of the

primordial matter power spectrum, $n_s = 0.967$. The initial baryonic mass fractions of hydrogen and helium are $X = 0.756$ and $Y = 1 - X = 0.244$, respectively.⁴ Note that the values of the *Hubble* and density parameters are given at $z = 0$. The following derived parameters may also be useful. The amplitude of the initial power spectrum parametrized as the rms mass density fluctuation in spheres of radius $8 h^{-1}$ Mpc extrapolated to $z = 0$ using linear theory, $\sigma_8 = 0.807$; the amplitude of the initial power spectrum parametrized as $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3} = 0.815$.

These values of the cosmological parameters are identical to those used for the fiducial FLAMINGO simulations (J. Schaye et al. 2023; R. Kugel et al. 2023). They are the maximum posterior likelihood values from the Dark Energy Survey year three (DES Y3; T. M. C. Abbott et al. 2022) ‘3×2pt + All Ext.’ Λ CDM cosmology. These values assume a spatially flat universe and are based on the combination of constraints from three DES Y3 two-point correlation functions: cosmic shear, galaxy clustering, and galaxy–galaxy lensing, with constraints from external data from baryon acoustic oscillations, redshift-space distortions, type Ia SNe, and *Planck* observations of the CMB (including CMB lensing), big bang nucleosynthesis, and local measurements of the *Hubble* constant (see T. M. C. Abbott et al. 2022 for details). The neutrino density corresponds to $\sum m_\nu c^2 = 0.06$ eV, which is the minimum mass allowed by neutrino oscillation experiments (I. Esteban et al. 2020; P. F. de Salas et al. 2021), and is consistent with the 95 per cent confidence upper limit of 0.13 eV from DES Y3. In this model, the neutrino contribution is provided by one massive and two massless species.

2.3 Structure finding

To identify collapsed structures such as galaxies and clusters of galaxies, the snapshots are post-processed with three different algorithms.

First, a friends-of-friends (FoF) halo finder is run on the DM particles,⁵ which links all DM particles separated by less than the linking length of 0.2 times the mean distance between DM particles. Baryonic particles are then linked to the nearest DM particle within the linking length, if any. FoF groups whose total number of DM, stellar, gas, and BH particles is smaller than 32 are discarded.

Second, the subhalo finder HBT-HERONS⁶ (V. J. Forouhar Moreno et al. 2025) is run, which uses the FoF catalogue and snapshot data as inputs. HBT-HERONS is a modified version of the hierarchical bound-tracing algorithm used in the HBT+ code (J. Han et al. 2012, 2018). The algorithm assumes that structure forms hierarchically, i.e. that every satellite subhalo was once a central subhalo. HBT-HERONS improves on HBT+ in a variety of ways, some of which are critically important for hydrodynamical simulations. V. J. Forouhar Moreno et al. (2025) show that by tracking subhaloes in time, HBT-HERONS outperforms structure finders that only use instantaneous phase-space

information, which tend to have difficulty identifying satellites near halo centres, even if these contain very large numbers of bound particles. Furthermore, satellites found by phase-space finders often undergo unphysically large and non-monotonic changes in mass when they cross pericentre, a problem that is nearly absent for HBT-HERONS (Á. Chandro-Gómez et al. 2025).

HBT-HERONS starts by iteratively identifying self-bound⁷ subhaloes within the highest redshift output and then sequentially processes each subsequent snapshot. Tracer particles, for which we use the 10 most bound particles of DM or stellar type, are subsequently used to track the descendants of the subhaloes and identify their host FoF haloes in the next snapshot, also allowing for the possibility that subhaloes may currently not have an FoF host. If an FoF halo contains multiple subhaloes, then the central subhalo is chosen from among the subhaloes whose masses exceed 0.8 times the mass of its most massive subhalo in the previous snapshot. If more than one subhalo fulfills this criterion, then the one with the lowest specific orbital kinetic energy in the centre-of-mass frame of the FoF halo is chosen as the central. A new subhalo is assigned to any FoF halo that does not yet contain any provided the new subhalo contains at least 20 bound particles and at least 10 tracer (i.e. DM or stellar) particles. Any particle that is no longer bound to the subhalo is removed from it.

Central subhaloes can accrete mass from surrounding diffuse matter as well as from their satellites, but satellite subhaloes can generally only accrete mass that has been part of their own satellites (i.e. satellites of satellites). An exception to this ‘hierarchical’ transfer of mass exists, as gas particles can be transferred between subhaloes without any prior connections under the following specific condition. Before unbinding, HBT-HERONS finds the 10 nearest tracer particles for all gas particles in an FoF halo. If none of these neighbours are part of the same subhalo as the gas particle in question, then the gas particle is reassigned to the subhalo that contains its nearest tracer particle. This re-assignment of gas particles relaxes the assumption of hierarchical assembly for gas, which, given its collisional nature, is not as well motivated as for DM and stars.

Subhaloes that overlap strongly in phase space are merged. Subhaloes containing fewer than 20 bound particles or fewer than 10 tracer particles become ‘orphans’, which may re-appear as bound subhaloes in later snapshots. The location of a subhalo is defined by that of its most bound particle. For further details, see V. J. Forouhar Moreno et al. (2025).

Third, the Spherical Overdensity and Aperture Processor (SOAP) tool (R. McGibbon et al. 2025) is used to compute a large number of (sub)halo properties in a variety of apertures. SOAP takes the subhalo locations and the particle IDs of the subhaloes’ bound particles, which are provided by HBT-HERONS, as input. All galaxy and halo properties, other than their locations and bound particle membership, are taken from SOAP rather than from HBT-HERONS. SOAP computes properties in a variety of 3D and projected apertures, including apertures of fixed proper sizes and mean internal overdensities, and including either all or only bound particles. The spherical overdensity radius $R_{\Delta c}$ is defined as the radius within which the mean internal total

⁴The transfer function used for the ICs assumes fractions of 0.754579 and 0.245421, respectively.

⁵Following convention, we often refer to DM particles, in which case we actually mean CDM particles. Massive neutrinos do not contribute to the mass of CDM particles, which implies that neutrinos are implicitly assumed not to cluster on the scales of interest here. This is a good assumption for galaxies, but it is not quite correct for massive clusters of galaxies.

⁶<https://hbt-herons.strw.leidenuniv.nl/>

⁷HBT-HERONS defines the binding energy of a particle as the difference between its gravitational potential energy and its kinetic energy in the rest frame of the subhalo, where the former is computed from the particles making up the subhalo and the latter includes the thermal energy in the case of gas particles.

matter density is Δ times the redshift-dependent critical density of the Universe. The mass inside $R_{\Delta c}$ is labelled $M_{\Delta c}$. Spherical overdensity apertures are only computed for central subhaloes, whereas properties in apertures of fixed proper size are computed for all subhaloes. In this work, galaxy properties will be computed using the bound particles within 3D apertures of radius 50 proper kpc (pkpc) centred on the most bound particle of the subhalo, a choice motivated by the finding of A. de Graaff et al. (2022) that this aperture reproduces masses inferred from Sérsic fits to mock Sloan Digital Sky Survey (SDSS) images.

Several caveats should be noted. First, the value of the FoF linking length that we use is conventional and is motivated by the virial overdensity predicted by analytic spherical collapse calculations (for a discussion, see S. More et al. 2011). However, the precise value is arbitrary because collapse is not spherically symmetric. Moreover, FoF haloes cannot be determined observationally. This is problematic because even though halo mass is generally expressed as a spherical overdensity mass instead of as an FoF mass, the FoF linking length does determine the number of haloes and hence the number of central galaxies. Indeed, V. J. Forouhar Moreno et al. (2025) show that varying the linking length by 20 per cent leads to ≈ 6 per cent changes in the halo mass function at $M_{200c} = 10^{12} M_{\odot}$. Second, while HBT-HERONS outperforms other structure finders when it comes to identifying satellite subhaloes, it cannot find galaxies that were born without DM if such objects exist. We note that algorithms that do find such objects will generally also fragment galaxies into self-bound subunits such as star clusters and molecular clouds, a problem that becomes more severe in simulations that resolve the multi-phase ISM. Third, HBT-HERONS does not account for the accretion of mass by satellites, except for accretion from satellites of the satellite in question and through gas re-assignment. Fourth, HBT-HERONS does not account for mergers between satellites that were independently accreted by their host halo and therefore lack a hierarchical connection. Most particles from the merged subhalo are therefore assigned to the host halo instead of the descendant satellite, resulting in the presence of a small number of residual substructures in the central subhalo. We find the incidence of such cases to be low and, when present, the substructures are only evident in a small fraction of the stellar component. Fifth, because every HBT-HERONS subhalo must have been identified as a central at some point, subhaloes that were both born and became satellites in between consecutive snapshots are missed. However, we find that the results are converged for the number of snapshots that we use. Sixth, the (sub)halo catalogues extend down to particle numbers for which resolution effects are often important, so it is always necessary to use a convergence test to determine the actual resolution limit for the question of interest.

2.4 Data products

We save the state of the simulation at 128 redshifts between $z = 30$ and 0. This includes 36 full snapshots and 92 reduced snapshots, also referred to as ‘snipsnaps’ (see Supporting Information Table `output_list.txt` for all the output times). Snapshots store many properties of the DM, gas, stellar, and BH particles. While snipsnaps contain all BH properties, they only record a small subset of key gas and stellar properties and we remove the DM properties after completion of the structure finding (see Supporting Information Table `output_properties.pdf` for a list of the particle properties that are saved). For three small-volume simulations (L050m6, L050m6h, and L012.5m5; Table 2),

we save many more snipsnaps (2000 for L050m6 and L050m6h, and 10 000 for L012.5m5) to facilitate video creation and to enable projects requiring particle tracking at very high cadence.

We save auto and cross power spectra at 123 redshifts between $z = 30$ and 0 for matter–matter, CDM–CDM, gas–gas, stars–stars, gas–matter, CDM–gas, gas–(stars+BHs), CDM–(stars+BHs), pressure–pressure, and matter–pressure, where matter is the sum of the densities of CDM, gas, stars, and BHs, but not neutrinos, and where pressure refers to the electron pressure.

To facilitate the production of mock quasar absorption spectra at high cadence without having to read all the data, we frequently save the SPH particles that overlap with 35 randomly positioned lines parallel to the x -axis of the cubic simulation volume and similarly for the y - and z -axes. These ‘line-of-sight’ outputs commence at expansion factor $a = 0.1$ and are produced at intervals of $\Delta a = 0.01a$, yielding 232 output times in total.

As described in Section 2.3, snapshots and snipsnaps are processed using the HBT-HERONS structure finder and the SOAP tool, yielding catalogues of many (sub)halo properties for multiple apertures, as well as merger trees.

3 SUBGRID PRESCRIPTIONS

As is the case for any simulation of a macroscopic system, prescriptions are required to model what happens below the resolution scale (‘subgrid’). In this section, we will summarize our subgrid models for radiative cooling and heating (Section 3.1), dust grains (Section 3.2), star formation (Section 3.3), stellar mass loss and the rates of SNIa (Section 3.4), turbulent diffusion (Section 3.5), stellar winds, radiation pressure from starlight, and H II regions, i.e. stellar feedback processes that start earlier than SN feedback (Section 3.6), SN feedback (Section 3.7), BHs (Section 3.8), and AGN feedback (Section 3.8). The level of detail provided in the subsections depends on whether the subgrid module is described in a dedicated paper and on the importance of the subgrid process for the results and interpretation of the simulations. The calibration of the subgrid feedback parameters to the $z = 0$ galaxy SMF, galaxy sizes, and BH masses is summarized in Section 3.10.

Before describing the individual subgrid models, we provide a brief overview of the main features in the order in which they are discussed in the subsequent subsections:

(i) Radiative cooling is allowed down to ≈ 10 K (S. Ploekinger et al. 2025). Direct modelling of the cold interstellar gas phase necessitates including more physics than is required for simulations that impose an equation of state. In particular, we allow for the formation of molecules in both the gas phase and on the surface of dust grains, self-shielding of the ionizing and photodissociating radiation by gas and dust, photoelectric heating by dust, photoionization, photodissociation, and photoheating by the meta-galactic radiation field, the interstellar radiation field (ISRF), and interstellar cosmic rays. We compute the radiative cooling and heating rates, as well as the changes in the species fractions using our CHIMES (A. J. Richings, J. Schaye & B. D. Oppenheimer 2014a, b) chemical reaction network, which includes the time-dependent evolution of 157 ions and molecules from 11 elements. However, to save computing time, we only treat the hydrogen and helium species fully in non-equilibrium. Cooling rates due to metals are computed element-by-element using tables assuming equilibrium species fractions, but the rates are corrected for the non-equilibrium free electron densities provided by the reaction

network for hydrogen and helium. We use the local (thermal plus turbulent) Jeans length to estimate the shielding column density, the intensity of the ISRF, and the cosmic ray density.

(ii) The formation, growth, and destruction of dust is computed on the fly, using a model consisting of three grain species and two grain sizes (J. W. Trayford et al. 2025). The dust model is coupled to the model for radiative cooling. Reactions and cooling rates involving grains account for the amount and properties of the dust that is tracked in the simulation. Gas metal cooling rates account for the depletion of individual elements onto dust grains.

(iii) Star formation is implemented using a Schmidt law combined with a gravitational instability criterion (F. S. J. Nobels et al. 2024). While the pressure-dependent model used in EAGLE is ideal for simulations without a cold ISM, because it enables the direct implementation of the observed kpc-scale Kennicutt–Schmidt star formation surface density law, simulations like COLIBRE that model the multiphase ISM can be more ambitious and aim to predict the coarse-grained star formation law.

(iv) The nucleosynthetic stellar yields from asymptotic giant branch (AGB) stars, CCSNe, and SNIa use up to date data from the literature (C. Correa et al., in preparation), and the SN Ia delay time distribution is calibrated to the observed evolution of the cosmic SN Ia rate.

(v) Besides the 11 elements that dominate the radiative cooling rates, we track the abundances of the *s*-process elements Ba and Sr, which are produced during the AGB phase, and the *r*-process element Eu, whose production by neutron star mergers, common-envelope jets SNe, and collapsars is modelled (C. Correa et al., in preparation).

(vi) The turbulent diffusion of element mass fractions and the mass fractions of different types of dust grains are modelled (C. Correa et al., in preparation).

(vii) Feedback processes from massive stars that begin before the first CCSNe explode, i.e. stellar winds, radiation pressure, and H II regions, are included (A. Benítez-Llambay et al. 2025).

(viii) The model for CCSN feedback includes not only a stochastic thermal component (as in C. Dalla Vecchia & J. Schaye 2012, but modified to improve the sampling of feedback events in low-density gas) that can drive galactic winds, but also a lower-energy kinetic component that helps drive turbulence (E. Chaikin et al. 2023).

(ix) SN Ia energy feedback is implemented using the same stochastic thermal prescription used for CCSNe in order to mitigate numerical overcooling.

(x) While our fiducial runs use purely thermally-driven AGN feedback (as in C. M. Booth & J. Schaye 2009, but modified to improve the sampling of feedback events for low-mass BHs), we repeat a subset of simulations using the F. Huško et al. (2026) hybrid model that combines thermally-driven wind and kinetic jet feedback. These hybrid AGN feedback simulations track the spin of the BHs.

The remainder of this section is intended to serve as a reference describing the methods that can be skipped by readers interested mostly in the results.

For the equations in this section, we assume the primordial hydrogen mass fraction $X = 0.756$ (e.g. E. Aver et al. 2021) when converting gas mass densities, ρ_g , into total hydrogen number densities, n_H . Note that unless otherwise stated, we use each particle’s actual element mass fractions during the simulation. For equations that depend on the gas particle mass, we assume the mean initial mass for our intermediate-resolution (i.e. m6 resolution; see Section 4) simulations.

3.1 Radiative cooling

Radiative cooling and heating rates are computed using HYBRID-CHIMES, which is detailed in S. Ploeckinger et al. (2025, we use their ‘weak ISRF’ model) and builds on the work of S. Ploeckinger & J. Schaye (2020). The main differences from the subgrid model for radiative cooling used in the simulations presented in S. Ploeckinger et al. (2025) are that in COLIBRE the chemistry and cooling are coupled to the abundances of dust grains that are tracked by each gas particle, as described in Section 3.2.

Cooling is allowed down to temperatures⁸ of 10 K. Rates are computed element by element, accounting for the individual abundances of all the elements that R. P. C. Wiersma, J. Schaye & B. D. Smith (2009a) showed contribute significantly to the cooling: H, He, C, N, O, Ne, Mg, Si, S, Ca, and Fe. The abundances of species, i.e. ion and molecule mass fractions, and cooling rates are computed using an updated version of the CHIMES⁹ reaction network (A. J. Richings et al. 2014a, b) for the radiation field, cosmic ray, and shielding prescriptions from S. Ploeckinger et al. (2025). CHIMES includes 157 ions and molecules. It accounts for the processes of collisional ionization, radiative and di-electronic recombination, grain–surface recombination, charge transfer, collisional excitation, Bremsstrahlung, molecule formation in gas and on dust grains, rovibrational and rotational cooling, Compton heating and cooling (using relativistic corrections, see S. Ploeckinger et al. 2025), cosmic ray ionization and heating, photoionization, photodissociation, gas and dust photoelectric heating, dust depletion, dust cooling, and self-shielding by dust and gas.

The abundances of free electrons and the nine H and He species H I, H II, H⁺, He I, He II, He III, H₂, H₂⁺, and H₃⁺ are tracked individually, i.e. without assuming equilibrium. Reactions involving dust grains, such as the formation of molecular hydrogen on grain surfaces, account for the abundances of grain species tracked in the simulation. Although HYBRID-CHIMES includes the option to boost the rates of reactions on the surface of dust grains, including the formation of H₂, this option is not used in COLIBRE. Cooling due to the HD molecule is added assuming a constant $n_{HD}/n_{H_2} = 1.65 \times 10^{-5}$ (M. Pettini & D. V. Bowen 2001).

The abundances of metal ions and molecules containing heavy elements (CO, H₂O, and OH) are tabulated using CHIMES as a function of density, temperature, metallicity, and redshift, assuming chemical equilibrium (i.e. ionization equilibrium and steady-state chemistry). However, cooling rates due to heavy elements are computed using the non-equilibrium free electron abundance,¹⁰ so the equilibrium assumption is only used for the calculation of the ion and molecule fractions of heavy elements.

Gas cooling rates account for the depletion of heavy elements on dust grains. The dust cooling rate (from recombinations on the grain surface and from grain–gas collisions), the dust photoelectric heating rate, and the formation rate of molecular hydrogen on dust grains account for the grain sizes tracked in the simulation, assuming that the rates scale in proportion with the surface area

⁸To be precise, in SWIFT cooling is allowed down to the internal energy corresponding to neutral gas (mean molecular weight $\mu = 1.224$) with a temperature of 10 K. In addition, CHIMES imposes a minimum temperature of 10 K.

⁹<https://richings.bitbucket.io/chimes/home.html>

¹⁰By ‘non-equilibrium free electron density’, we mean the sum of the non-equilibrium contributions from hydrogen and helium and the (generally much smaller) equilibrium contribution from heavy elements.

to volume ratio, i.e. inversely with the linear grain size for our spherical dust grains.

COLIBRE does not include radiative transfer, but we do model some of its main effects in an approximate manner. While cooling radiation is assumed to escape freely,¹¹ we do account for the attenuation of photoionization, photodissociation, and photoheating rates, which is a critical ingredient for simulations aiming to capture the multiphase ISM. Shielding by dust,¹² H I, H₂, He I, He II, and CO in the gas near the receiving gas particle is modelled. Surface densities are estimated using the local Jeans column density, which is the product of the density of the species of interest and the Jeans length. This approximation, which was used to model the transitions from the ionized to atomic and from the atomic to molecular phases by J. Schaye (2001, 2004), reproduces the ionized to atomic transition in cosmological radiative transfer simulations (A. Rahmati et al. 2013), which, however, lack the resolution and physics to model the atomic to molecular transition. Because the use of the Jeans column density implies the assumption that the gas is self-gravitating, it may overestimate the effect of shielding for pressure-confined clouds. On the other hand, if poorly resolved pressure-confined clouds are embedded in a resolved self-gravitating structure, such as the thick disc of warm gas, which is a situation quite typical for the ISM at the resolution of COLIBRE, then the Jeans column density may still provide a reasonable estimate of the shielding column.

Because we expect supersonic turbulence in the cold phase, we use the maximum of the thermal and turbulent Jeans scales, where the latter assumes a 1D turbulent velocity dispersion of 6 km s^{-1} . At very low densities, corresponding to dynamical time-scales that are long compared to other relevant time-scales, local hydrostatic equilibrium is not expected, and we therefore cap the maximum Jeans length to 10 (50) kpc before (after) reionization. While the shielding of the hydrogen and helium species, whose abundances are computed in non-equilibrium, is modelled using the self-consistently predicted dust grain abundances, the shielding of heavy elements, for which cooling rates are tabulated, assumes a dust-to-metal ratio that depends on the density and temperature but saturates to a constant value in the ISM.

The calculations of species abundances and cooling rates account for the presence of the CMB, the UV(ultraviolet)/X-ray background from galaxies and quasars, an ISRF, and interstellar cosmic rays. We adopt a modified version of the C.-A. Faucher-Giguère (2020) model for the redshift-dependent metagalactic UV/X-ray background. Following S. Ploekinger & J. Schaye (2020), at $z > 7.2$ ($z > 3$), we attenuate the C.-A. Faucher-Giguère (2020) spectrum above the H I (He II) ionization energies using H I (He II) column densities tuned to match the effective photo-ionization and photoheating rates that C.-A. Faucher-Giguère (2020) finds are needed to match observations of the intergalactic medium (IGM) and the optical depth seen by the CMB. Hydrogen and helium reionization complete at $z \approx 7$ and ≈ 3 , respectively. S. Ploekinger et al. (2025) show that the model repro-

duces the observed thermal evolution of the IGM. Due to the non-equilibrium treatment of hydrogen and helium, no extra energy injections around the times of hydrogen and helium reionization (as e.g. used in EAGLE) are needed (e.g. E. Puchwein et al. 2015).

The incident ISRF has the fixed spectral shape from J. H. Black (1987), which combines the radiation field in the solar vicinity from J. S. Mathis, P. G. Mezger & N. Panagia (1983) with the Galactic soft X-ray background from J. N. Bregman & J. P. Harrington (1986). The intensity of the ISRF and the cosmic ray density are assumed to scale with the local SFR surface density, which is assumed to scale with the gas surface density to the power 1.4, as for the observed R. C. Kennicutt (1998) law. The estimate for the local gas surface density is computed in the same way as the self-shielding column density. We use the ‘weak ISRF’ model of S. Ploekinger et al. (2025) because we found that for dwarf galaxies ($M_* \sim 10^9 M_\odot$) their fiducial ISRF, which is normalized to the J. H. Black (1987) amplitude for solar neighbourhood conditions, results in an unnecessarily large fraction of star formation occurring in warm ($T \sim 10^4 \text{ K}$) gas. The amplitude of the ISRF saturates at a total hydrogen column density $N_{\text{H}} = 10^{22} \text{ cm}^{-2}$, because we do not expect the scaling with the Jeans column density to hold inside molecular clouds. The amplitude of the cosmic ray density saturates at $N_{\text{H}} = 10^{21} \text{ cm}^{-2}$, in agreement with the observations of N. Indriolo et al. (2015). The amplitudes of both the ISRF and the cosmic ray density are cut-off below a proper density of $n_{\text{H}} = 10^{-2} \text{ cm}^{-3}$ and an overdensity of 100, because such low (over)densities correspond to conditions typical for the IGM rather than the ISM.

As we cannot resolve the initial phases of SN explosions and AGN winds, species abundances, i.e. ion and molecule fractions, are set to their equilibrium values when the particles are directly heated by thermal SN or thermal AGN feedback, i.e. when their temperature is increased due to the explicit injection of feedback energy (see Sections 3.7 and 3.9).

For each particle at each time-step, we compute the internal energy that the particle is expected to reach at the end of its time-step based on its current net radiative cooling rate. The time derivative of the internal energy is then adjusted so that the particle will drift to this final value over the course of its time-step. Note that if a particle is drifting between time-steps at the time a snapshot is written, then the recorded internal energy is not fully consistent with the recorded ion and molecule abundances, because in CHIMES, the abundances are immediately set to the final values at the end of the time-step. Because the method of drifting the internal energy is only appropriate if the particle is on a time-step that is short compared to the radiative cooling time, we adopt the same modification of this standard prescription as used in the FLAMINGO simulations: if the internal energy is expected to decrease by at least one third during the time-step, then we immediately set the internal energy to the expected final value. This modification also improves consistency between the internal energy and species abundances saved in snapshots for particles that are in between time-steps. A more accurate solution for the treatment of rapidly cooling gas would be to limit the time-step to be small compared with the radiative cooling time, but this would be computationally prohibitively expensive.

3.2 Dust grains

The dust content is a subgrid property of gas particles. The model for the origin and evolution of dust grains is described and tested in J. W. Trayford et al. (2025). Here, we will only give a brief summary.

¹¹For molecular cooling from CO, H₂O, and OH CHIMES accounts for self-absorption of the cooling radiation.

¹²For dust shielding of hydrogen and helium species, we use the dust-to-gas ratio predicted by the simulation. For dust shielding of metal species, we assume a dust-to-gas ratio of 6.6×10^{-3} , as in the CLOUDY ISM grain model (G. J. Ferland et al. 2017). S. Ploekinger et al. (2025) assumed 5.5×10^{-3} in order to match the depletion from E. B. Jenkins (2009). This is not necessary for COLIBRE because depletion and dust-to-gas ratios are treated self-consistently in the live dust model.

Gas particles track the fraction of the mass that is in dust grains of three different chemical compositions and two different grain sizes, i.e. six dust mass fractions in total. The three species are graphites, which are pure carbon, and two kinds of silicates, namely forsterite (Mg_2SiO_4) and fayalite (Fe_2SiO_4), which are both olivine. The grains are assumed to be spherical with radii $r_g = 0.01$ or $0.1 \mu\text{m}$.

Grains are seeded by AGB stars using the metallicity-dependent dust yields of F. Dell’Agli et al. (2017), assuming that stars with zero-age main-sequence (ZAMS) mass $0.85 \leq M/M_\odot \leq 3.5$ and $3.5 < M/M_\odot \leq 5$ produce graphites and silicates, respectively. CCSN dust yields are adapted by combining the dust nucleation yields from S. Zhukovska, H. P. Gail & M. Trieloff (2008) with the total metallicity-dependent yields from C. Correa et al. (in preparation) that are discussed in Section 3.4. The total amount of silicates produced by AGB stars and CCSN depends on the amounts of silicon and oxygen in the ejecta. Forsterite and fayalite are assumed to each account for half of the released silicate mass. Ninety per cent of the dust produced (by mass) is assumed to consist of large (i.e. $0.1 \mu\text{m}$) grains.

The metal yields used to enrich the gas are computed as the difference between the total metal yields (see Section 3.4) and the dust metal yields.

The precise values used for grain seeding, including the assumed fractions of large and small grains, are unimportant because the grain properties are mainly determined by grain evolutionary processes. We note that this situation differs from that in some other simulations (e.g. R. McKinnon, P. Torrey & M. Vogelsberger 2016; R. Davé et al. 2019), which assume the dust yields of E. Dwek (1998) that are an order of magnitude higher than ours. See J. W. Trayford et al. (2025) for further discussion.

Following H. Hirashita & N. V. Voshchinnikov (2014), gas accretion causes grains to grow exponentially in mass with an e -folding time of

$$\tau_{\text{acc}} = \tau_G \left(\frac{r_g}{0.1 \mu\text{m}} \right) \left(\frac{X_j}{X_{j,\odot}} \right)^{-1} \left(\frac{n'_H}{10 \text{ cm}^{-3}} \right)^{-1} \left(\frac{T}{10 \text{ K}} \right)^{-1/2}, \quad (1)$$

where $\tau_G = 180$ and 99.3 Myr for graphites and silicates, respectively, and X_j is the gas-phase mass fraction of the bottleneck element, i.e. the element whose abundance limits would first limit the amount of the grain species that can be formed. The fiducial values correspond to a sticking probability of 0.3 for gas accretion. The mass in each dust component is increased by an exponential factor based on the time-scale computed for that component using this equation.

Because our simulations cannot resolve the dense parts of molecular clouds where grain growth is most efficient, we expect to underestimate the rate of grain growth in dense gas. We therefore increase the density appearing in the gas accretion rate by a subgrid clumping factor, C , i.e. $n'_H = C n_H$. Since we do not wish to boost the rates when the gas densities are sufficiently low that we do not expect significant resolution limitations, and since we do not wish to boost the densities without limit, we adopt the following functional form for the clumping factor

$$C = \begin{cases} 1 & n_H \leq 0.1 \text{ cm}^{-3}, \\ \left(\frac{n_H}{10^{-1} \text{ cm}^{-3}} \right)^{2/3} & 0.1 \text{ cm}^{-3} < n_H \leq 100 \text{ cm}^{-3}, \\ 100 & n_H > 100 \text{ cm}^{-3}. \end{cases} \quad (2)$$

J. W. Trayford et al. (2025) show that for high densities ($n_H \gg 10 \text{ cm}^{-3}$) the results are insensitive to the precise value of the maximum clumping factor, provided it is $\gtrsim 10$, because the dust

depletion fractions saturate (at a value of $2/3$ for graphite, as discussed below, and at levels determined by the abundance of the bottleneck element for silicates).

Thermal sputtering, i.e. destruction by collisions with hot gas particles, causes the grain mass to decline exponentially with an e -folding time of (J. C. Tsai & W. G. Mathews 1995)

$$\tau_{\text{sp}} = 0.85 \text{ Myr} \left(\frac{r_g}{0.1 \mu\text{m}} \right) \left(\frac{n_H}{10 \text{ cm}^{-3}} \right)^{-1} \times \left[1 + \left(\frac{T}{2 \times 10^6 \text{ K}} \right)^{-2.5} \right]. \quad (3)$$

As was the case for gas accretion, we compute the change in mass of each dust component using the e -folding time for that specific component, which for the case of thermal sputtering varies with the grain size but not with the grain composition. Note that the thermal sputtering time-scale depends on the SPH density, i.e. it is not reduced by the clumping factor, because thermal sputtering is due to collisions of grains with hot gas particles instead of with other dust grains. Dust in gas that is converted into stars and in gas directly heated by thermal SN or thermal AGN feedback (see Sections 3.7 and 3.9) is assumed to be fully destroyed by astration and unresolved sputtering, respectively.

Grain–grain collisions cause dust mass to be transferred from large to small grains. This grain shattering process proceeds at an exponential rate with an e -folding time of (S. Aoyama et al. 2017; G. L. Granato et al. 2021)

$$\tau_{\text{sh}} = 54.1 \text{ Myr} \left(\frac{DTG}{0.01} \right)^{-1} \left(\frac{r_g}{0.1 \mu\text{m}} \right) \times \begin{cases} \left(\frac{n_H}{1 \text{ cm}^{-3}} \right)^{-1} & \text{for } n_H < 1 \text{ cm}^{-3}, \\ \left(\frac{n_H}{1 \text{ cm}^{-3}} \right)^{-1/3} & \text{for } n_H \geq 1 \text{ cm}^{-3}. \end{cases} \quad (4)$$

Here DTG is the dust-to-gas ratio, i.e. the fraction of the gas particle mass that is in dust. Grain–grain collisions also lead to transfer of dust mass from small to large grains via coagulation, which proceeds at an exponential rate with an e -folding time based on S. Aoyama et al. (2017):

$$\tau_{\text{co}} = 13.55 \text{ Myr } f_{\text{co}} \left(\frac{DTG}{0.01} \right)^{-1} \left(\frac{r_g}{0.1 \mu\text{m}} \right) \left(\frac{n'_H}{10 \text{ cm}^{-3}} \right)^{-1}. \quad (5)$$

Here, f_{co} is a resolution-dependent factor that we introduced to improve the convergence of the grain sizes with resolution. It is set to $10^{-0.5}$ for our lowest resolution (i.e. m7) simulations and to unity for all other simulations. The shattering and coagulation time-scales assume a material density of 3 g cm^{-3} and a coagulation velocity of 0.2 km s^{-1} . Because shattering is thought to be important in the warm ISM, which we can resolve, while coagulation dominates in the cold and dense ISM (e.g. H. Hirashita & H. Yan 2009), which we cannot fully resolve, we boost the density appearing in the coagulation time-scale by the clumping factor (equation 2), but not the density appearing in the time-scale for shattering. Note that coagulation always dominates over shattering, even if we do not boost the coagulation rate by the clumping factor. Grain growth by gas accretion is faster for smaller grains. Grain mass is then transferred to larger grains because coagulation dominates over shattering.

Motivated by observations of molecular clouds (e.g. A. Fuente et al. 2019), the fraction of carbon in graphite grains is limited to two-thirds of the total carbon content of each gas particle. Without this limit, most of the carbon in the dense ISM would be depleted, whereas in reality, much of the carbon in high-density

gas is observed to be in the form of CO. Given that we lack the resolution to track CO explicitly and that gaseous carbon is an important coolant and diagnostic, we opt to cap the maximum depletion factor. A cap on the dust depletion of oxygen, which is another important coolant in the gas phase, is unnecessary, because its depletion onto silicates is limited by the lower availability of other constituents.

The dust grains are coupled to the radiative cooling and chemistry modules described in Section 3.1. In particular, the cooling rates account for the depletion of heavy elements onto dust grains, for dust cooling and dust photoelectric heating, and for the formation of molecular hydrogen on dust grains, which all depend on the grain size. Finally, we note that dust grains are subject to turbulent diffusion using the prescription described in Section 3.5.

J. W. Trayford et al. (2025) present further details of the dust modelling as well as a suite of simulations that vary the main free parameters and turn individual processes on or off.

3.3 Star formation

The prescription for star formation is taken from F. S. J. Nobels et al. (2024), to which we refer the reader for further details and tests. Briefly, star formation is implemented by stochastically converting gas particles that meet the criterion to be star-forming into collisionless stellar particles. Gas that is locally gravitationally unstable is eligible to form stars at a rate given by a M. Schmidt (1959) law with power-law index 3/2 and an efficiency per free-fall time of $\varepsilon = 0.01$.

Specifically, gas particles are only eligible for star formation, and hence labelled star-forming, if they satisfy the gravitational instability criterion

$$\alpha \equiv \frac{\sigma_{\text{th}}^2 + \sigma_{\text{turb}}^2}{G \langle N_{\text{ngb}} \rangle^{2/3} m_g^{2/3} \rho_g^{1/3}} < \alpha_{\text{crit}}, \quad (6)$$

where we set $\alpha_{\text{crit}} = 1$. Here, σ_{th} and σ_{turb} are the particle's 3D thermal and turbulent velocity dispersions, which are computed from its temperature and the local interparticle velocity (peculiar plus *Hubble*) dispersion, respectively; $\langle N_{\text{ngb}} \rangle$ is the weighted mean number of neighbours in the SPH kernel, which is about 65 for our quartic spline SPH kernel; m_g is the mean, initial gas particle mass; ρ_g is the gas density; and α_{crit} is a constant of order unity whose precise value depends on the assumed geometry. The factor $\langle N_{\text{ngb}} \rangle m_g$ appears because we require a mass equal to that contained within the weighted mean SPH kernel to be gravitationally unstable.

Hence, in higher resolution simulations, i.e. for smaller m_g , star formation requires colder and/or denser gas. Indeed, to take advantage of the fact that higher resolution simulations can model the ISM up to higher densities, we need to move the sub-grid scale, i.e. the scale at which the sub-grid model for star formation is applied, to higher densities as the resolution increases. Note that we therefore do not expect the physical conditions of gas that is labelled as star-forming to converge with the resolution. However, F. S. J. Nobels et al. (2024) show that coarse-grained quantities, such as the observable Kennicutt–Schmidt law on $\sim \text{kpc}$ scales, can still converge with the resolution.

It is instructive to consider the limiting case where the thermal velocity dispersion dominates over the turbulent dispersion. If $\sigma_{\text{th}} \gg \sigma_{\text{turb}}$, then equation (6) yields the temperature-dependent

density threshold¹³

$$n_{\text{H}} > 1.9 \times 10^{-1} \text{ cm}^{-3} \left(\frac{T}{10^4 \text{ K}} \right)^3 \left(\frac{m_g}{1.84 \times 10^6 \text{ M}_{\odot}} \right)^{-2}, \quad (7)$$

where n_{H} is the total hydrogen number density and we assumed a mean molecular mass appropriate for neutral gas, $\mu = 1.3$, when converting between σ_{th} and T . Similarly, if turbulence dominates, $\sigma_{\text{th}} \ll \sigma_{\text{turb}}$, then

$$n_{\text{H}} > 2.7 \times 10^{-2} \text{ cm}^{-3} \left(\frac{\sigma_{\text{turb}}}{10 \text{ km s}^{-1}} \right)^6 \left(\frac{m_g}{1.84 \times 10^6 \text{ M}_{\odot}} \right)^{-2}. \quad (8)$$

The SFR of gas particles that satisfy the star formation criterion (equation 6) is given by the Schmidt law

$$\dot{\rho}_{\star} = \varepsilon \frac{\rho_g}{t_{\text{ff}}} = \varepsilon \left(\frac{32G}{3\pi} \right)^{1/2} \rho_g^{3/2}, \quad (9)$$

where we set the efficiency per free-fall time $\varepsilon = 0.01$. F. S. J. Nobels et al. (2024) and C. d. P. Lagos et al. (2025) show that this value of the efficiency, which is motivated by observations of giant molecular clouds in the Milky Way (e.g. R. Pokhrel et al. 2021; Z. Hu et al. 2022), the Large Magellanic Cloud (e.g. B. B. Ochsendorf et al. 2017), and other nearby galaxies (e.g. D. Utomo et al. 2018; A. K. Leroy et al. 2025), yields good agreement with the observed Kennicutt–Schmidt law.

It is instructive to rewrite this as the gas consumption time-scale

$$\frac{\rho_g}{\dot{\rho}_{\star}} \approx 1.4 \times 10^{10} \text{ yr} \left(\frac{\varepsilon}{0.01} \right)^{-1} \left(\frac{n_{\text{H}}}{10^{-1} \text{ cm}^{-3}} \right)^{-1/2}, \quad (10)$$

which indicates that densities $n_{\text{H}} > 10^{-1} \text{ cm}^{-3}$ are required for a star-forming gas particle to be expected to be converted to a star particle within a *Hubble* time at $z = 0$. The Schmidt law corresponds to an SFR per gas particle of

$$\dot{m}_{\star} = m_g \frac{\dot{\rho}_{\star}}{\rho_g}, \quad (11)$$

$$\approx 1.3 \times 10^{-4} \text{ M}_{\odot} \text{ yr}^{-1} \left(\frac{\varepsilon}{0.01} \right) \left(\frac{m_g}{1.84 \times 10^6 \text{ M}_{\odot}} \right) \left(\frac{n_{\text{H}}}{10^{-1} \text{ cm}^{-3}} \right)^{1/2}, \quad (12)$$

which, combined with the star formation criterion, gives an estimate of the minimum possible SFR.

3.4 Stellar mass loss and chemical enrichment

Our model for stellar mass loss and chemical enrichment is based on the prescription developed by R. P. C. Wiersma et al. (2009b) for OWLS, a slightly modified version of which was used in EAGLE. It includes mass loss by winds from AGB and massive stars, CCSNe, and SNIa. As in R. P. C. Wiersma et al. (2009b), we track the 11 elements¹⁴ that R. P. C. Wiersma et al. (2009a) found to be important for the radiative cooling rates: H, He, C, N, O, Ne, Mg, Si, S, Ca, and Fe.

For COLIBRE, we introduce several improvements:

- (i) Updated nucleosynthetic yields for all enrichment channels.
- (ii) An updated SNIa delay time distribution.

¹³Eq. (29) of F. S. J. Nobels et al. (2024) expresses this as a density-dependent temperature ceiling, but the exponent of the density term was erroneously omitted from that equation. In addition, the solid lines of fig. 1 of F. S. J. Nobels et al. (2024), which use this equation, are normalized too low by about 0.35 dex.

¹⁴S and Ca are assumed to track Si, i.e. their abundance relative to that of Si is assumed to be solar (M. Asplund et al. 2009).

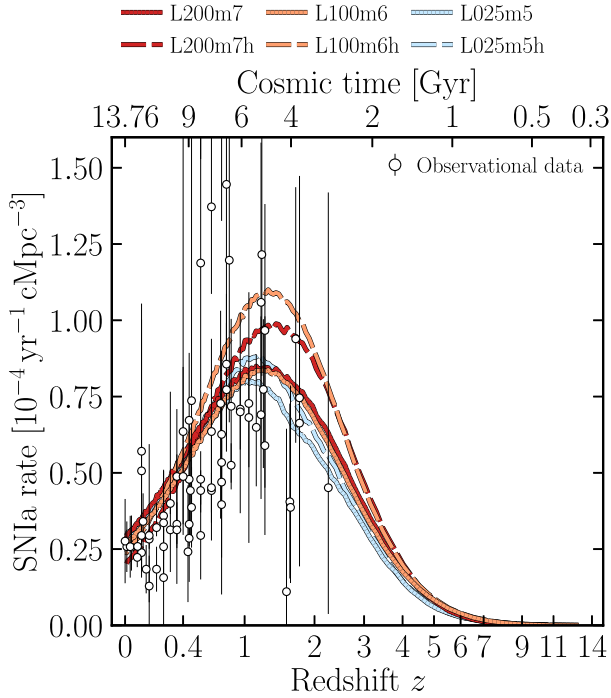


Figure 2. The evolution of the cosmic rate of SNIa per unit time and comoving volume for the fiducial (solid) and hybrid AGN feedback (dashed) simulations of different resolutions (different colours). The data points correspond to observed rates at $z < 0.2$ (E. Cappellaro, R. Evans & M. Turatto 1999; D. Hardin et al. 2000; D. S. Madgwick et al. 2003; L.-G. Strolger 2003; G. Blanc et al. 2004; F. Mannucci et al. 2005; A. Horesh et al. 2008; B. Dilday et al. 2010; W. Li et al. 2011; R. M. Quimby et al. 2012; C. Frohmaier et al. 2019), $0.2 < z < 0.75$ (R. Pain et al. 2002; B. J. Barris & J. L. Tonry 2006; J. D. Neill et al. 2006; M. T. Botticella et al. 2008; J. Melinder et al. 2012; E. Cappellaro et al. 2015) and at $z > 0.75$ (J. L. Tonry et al. 2003; T. Dahlen et al. 2004; T. Dahlen, L.-G. Strolger & A. G. Riess 2008; D. Poznanski et al. 2007; N. Kuznetsova et al. 2008; S. A. Rodney & J. L. Tonry 2010; O. Graur et al. 2011; K. Perrett et al. 2012; J. E. Okumura et al. 2014; S. A. Rodney et al. 2014). All simulations are in reasonable agreement with the compilation of observations.

(iii) Tracking of the s -process elements Sr and Ba, which are produced by AGB stars.

(iv) Tracking of the r -process element Eu, with separate contributions from neutron star mergers, common envelope jets SNe, and collapsars.

(v) A model for unresolved small-scale mixing by turbulent diffusion (see Section 3.5).

The model is motivated, detailed, tested, and calibrated in C. Correa et al. (in preparation). Here, we only provide a brief summary.

Each stellar particle constitutes a simple stellar population (SSP), i.e. a coeval population of stars of uniform chemical composition, whose ZAMS mass distribution follows the G. Chabrier (2003) initial mass function (IMF) in the range $0.1\text{--}100 M_{\odot}$. Stellar particles inherit their chemical composition from their parent gas particles, where heavy elements in gaseous and solid (i.e. dust) phases are summed.

At each enrichment time-step, which, as we discuss below, can be longer than the dynamical time-step for stellar particles older than 100 Myr, we determine which ZAMS masses transition to the AGB or CCSN stage using the metallicity-dependent stellar

lifetimes of L. Portinari, C. Chiosi & A. Bressan (1998). Elemental mass loss, including both individually tracked elements and a total metal mass tracer (which accounts for untracked elements), is computed by interpolating nucleosynthetic yield tables as a function of the ZAMS mass and metallicity. We distinguish between the mass of an element produced or destroyed within the star, and the mass of the element that the star was born with and that simply passed through the star. While the latter depends on individual element abundances, the former depends only on the total metallicity, as the yield tables taken from the literature assume that the abundances of heavy elements relative to each other are solar.

Intermediate-mass stars, i.e. stars with ZAMS masses of $1\text{--}8 M_{\odot}$, lose mass during the AGB phase. The AGB yields are taken from A. I. Karakas (2010), C. K. Fishlock et al. (2014), C. L. Doherty et al. (2014), A. I. Karakas & M. Lugaro (2016), and G. C. Cinquegrana & A. I. Karakas (2022). However, based on the findings of C. Kobayashi, A. I. Karakas & M. Lugaro (2020), we decrease the A. I. Karakas & M. Lugaro (2016) Ba yields by a factor of 2. Massive stars, i.e. stars with ZAMS masses of $8\text{--}40 M_{\odot}$, lose mass as a result of both stellar winds and CCSNe. CCSN yields are taken from K. Nomoto, C. Kobayashi & N. Tominaga (2013), while yields representing pre-SN mass loss are derived from C. Kobayashi et al. (2006). The massive star yields of C and Mg were multiplied by a factor of 1.5 to improve agreement with their abundance ratios relative to Fe in Milky Way stars from the Apache Point Observatory Galactic Evolution Experiment (APOGEE) survey (H. Jönsson et al. 2018) in the metallicity range $[\text{Fe}/\text{H}] \geq -1$. Such adjustments are commonly made, motivated by the fact that yields computed by different studies often differ at the factor of 2 level, even for a fixed IMF and ignoring complications due to stellar rotation and binary evolution (e.g. R. P. C. Wiersma et al. 2009b; D. Romano et al. 2010; T. Buck et al. 2021; J. Liang, E. Gjergo & X. Fan 2023). Very massive stars, i.e. stars with ZAMS masses $> 40 M_{\odot}$, are assumed to not enrich the ISM because their ejecta, including pre-SN mass loss, fall back into a BH. Note that part of the metal yields of AGB and CCSNe are released in the form of dust grains (see Section 3.2 and J. W. Trayford et al. 2025).

The rate of SNIa per unit ZAMS stellar mass for an SSP of age t is given by the empirical delay time distribution function

$$\dot{\mathcal{N}}_{\text{SNIa}}(t) = \frac{\nu}{\tau} e^{-(t-t_{\text{delay}})/\tau} \Theta(t - t_{\text{delay}}), \quad (13)$$

where $\nu = 1.54 \times 10^{-3} M_{\odot}^{-1}$, $\tau = 2$ Gyr, $t_{\text{delay}} = 40$ Myr, and $\Theta(x)$ is the Heaviside step function. While the value of t_{delay} is motivated by the lifetime of stars that end their lives as CCSNe, the values of ν and τ have been calibrated by F. S. J. Nobels et al. (in preparation) by comparing small-volume test simulations with observations. Fig. 2 shows that the model is in reasonable agreement with the observed evolution of the cosmic SNIa rate density compiled from (often discrepant) measurements taken from the literature. F. S. J. Nobels et al. (in preparation) show that the model also reproduces observations of the evolution of the specific SNIa rate as a function of galaxy stellar mass.

We adopt the SNIa yields from the W7 model of S.-C. Leung & K. Nomoto (2018) for near-Chandrasekhar-mass white dwarfs undergoing turbulent deflagration modelled with a deflagration-to-detonation transition. This is an updated version of the W7 model of K. Iwamoto et al. (1999).

We consider three production channels for the r -process element Eu: neutron star mergers, common-envelope jets SNe, and

collapsars. The rates of these events are modelled using calibrated delay-time distributions, each with functional forms that differ from the one adopted for SNIa. To more accurately capture the rarity and stochastic nature of the r -process events, a stochastic approach is implemented. We refer to C. Correa et al. (in preparation) for a detailed description.

While the yields for intermediate-mass and massive stars depend on ZAMS mass and metallicity, the yields for SNIa and the r -process production channels are assumed to be constant. Because the contributions of the s - and r -process elements to the cooling rates are neglected, and because their contributions to the overall metallicity (on which stellar lifetimes and yields depend) are negligible, their yields can be re-scaled in post-processing without breaking the self-consistency of the simulations. For Eu, this can be done separately for the different production channels, because we track the contribution of each channel.

The stellar mass loss of an SSP is a steeply declining function of its age. About half of the total mass lost in a *Hubble* time, which is a little over 40 per cent of the initial mass, is due to massive stars and hence occurs during the first ≈ 40 Myr (see e.g. fig. 1 of M. C. Segers et al. 2016a). To ensure that this early phase is well sampled, we limit the time-step (not just for enrichment, but also for the dynamics) of stellar particles younger than 40 Myr to be no longer than 1 Myr.

Because enrichment (specifically the neighbour finding) is computationally expensive, stellar particles older than 100 Myr only transfer mass after time intervals of 0.05 times their current age. All the mass lost during such a time interval is then transferred in a single time-step. This enrichment time-step is also applied to the SNIa energy feedback. Note that the dynamical time-steps, which for collisionless particles such as stars are set by the gravity solver, are not changed for these older stellar particles. We have verified that this restricted sampling of the enrichment by older stars does not have a significant effect on the particle metallicity distribution functions.

3.5 Turbulent diffusion

In SPH simulations, gas particles do not exchange mass, which implies that there is no mixing on mass scales smaller than the particle mass. This is not unreasonable because mixing on sub-particle mass scales is by definition unresolved. However, there are situations in which this lack of small-scale mixing can have undesirable consequences. An example, illustrated and discussed in R. P. C. Wiersma et al. (2009b), is the dispersal of heavy elements by galactic winds. A young star particle may enrich $\sim 10^2$ neighbouring gas particles with CCSN ejecta. If those particles are then moved to large distances by a galactic wind, the number of particles inside the radius reached by the outflow may be orders of magnitude larger, yet only the initial $\sim 10^2$ particles are enriched by the star particle. Such inhomogeneous enrichment would be unrealistic if unresolved small-scale mixing processes are in reality efficient, though in practice it may not be important if there are many enrichment events by many star particles. Since radiative cooling rates depend on the abundances at the particle/cell scale, the treatment of small-scale mixing may also affect the dynamics.

We note that grid codes tend to have the opposite problem as Lagrangian codes. Because in grid codes gas flows are fully mixed within a cell, such simulations may suffer from numerical overmixing.

We currently have a poor understanding of small-scale mixing processes (for a review, see C.-A. Faucher-Giguère & S. P. Oh 2023), partly because such processes are difficult to simulate, even for idealized cloud–wind interactions at very high resolution (e.g. J. Braspenning et al. 2023). Nevertheless, it is reasonable to assume that unresolved turbulence would enhance small-scale mixing. Following early work on chemodynamics (e.g. F. J. Martínez-Serrano et al. 2008; Greif et al. 2009; S. Shen, J. Wadsley & G. Stinson 2010), we therefore implement turbulent mixing as a diffusion process,

$$\frac{dX}{dt} = \frac{1}{\rho_g} \nabla \cdot (D \nabla X), \quad (14)$$

where X is an abundance, e.g. the mass fraction of a particular element, d/dt is the Lagrangian time derivative, and the diffusion coefficient D is taken to be proportional to the product of the Frobenius norm of the velocity shear tensor, the gas density, the SPH smoothing kernel squared, and a dimensionless free parameter, C_{diff} . The diffusion coefficient therefore has dimensions $\text{mass length}^{-1} \text{time}^{-1}$. Note that this is equivalent to using a diffusion coefficient $D' = D/\rho_g$ that does not depend on the gas density and has the more familiar dimensions of area over time, if we replace equation (14) by $\frac{dX}{dt} = \frac{1}{\rho_g} \nabla \cdot (D' \rho_g \nabla X)$. Our treatment of diffusion is motivated by the Kolmogorov turbulence cascade. We refer the reader to C. Correa et al. (in preparation) for further details.

The value of C_{diff} varies in the literature. Based on the C. Correa et al. (in preparation) comparison of $[\text{Mg}/\text{Fe}]$ as a function of $[\text{Fe}/\text{H}]$ in small test simulations and for observations of Milky Way stars, we adopt $C_{\text{diff}} = 0.01$. C. Correa et al. (in preparation) show that for $[\text{Fe}/\text{H}] > -1$, diffusion has only a limited impact on the element abundances of star particles within the disc of Milky Way-type galaxies. However, for lower metallicities, larger values of C_{diff} boost the alpha enhancement. C. Correa et al. (in preparation) also show that the effect of diffusion becomes more important at lower resolution. Including diffusion leads to better convergence of chemical abundances with numerical resolution, particularly on sub-galactic scales, but it does not have a significant effect on other galaxy properties. The fact that diffusion is more important at lower metallicities and at lower resolution is expected, as both tend to correspond to fewer enrichment events.

Turbulent diffusion is implemented for the elemental mass fractions, the mass fractions of dust grains of different chemical compositions and grain sizes, and for diagnostic tracers that track the mass fractions produced by different mass-loss channels (AGB stars, massive stars and CCSNe, SNIa, and different r -process channels).

Finally, we note that turbulent diffusion can limit the time-steps of gas particles, $(\Delta t)_i \leq 0.2 h_i^2 / D_i$, where h_i is the SPH smoothing length of particle i .

3.6 Stellar winds, radiation pressure, and H II regions

Besides the radiative feedback from the local ISRF described in Section 3.1, we include three types of stellar feedback that are active before CCSNe start exploding: stellar winds, radiation pressure, and H II regions. The implementation of these processes is detailed in A. Benítez-Llambay et al. (2025), here we will only give a brief summary. We refer to the collective effects of these processes as ‘pre-CCSN’ or ‘early’ stellar feedback, but note that they continue to operate after the first CCSN of an SSP explodes.

Early stellar feedback not only provides feedback channels additional to CCSNe, but can also change the effectiveness of CCSN feedback. On the one hand, early stellar feedback can disperse the high-density stellar birth cloud (e.g. M. Chevance et al. 2022), which can make CCSN feedback more effective by reducing radiative cooling losses for CCSNe produced by SSPs that have not yet moved out of their birth clouds. On the other hand, by smoothing the dense star-forming gas, the clustering of CCSNe is reduced, which can weaken the efficiency of the associated feedback (e.g. J. Rosdahl et al. 2015; M. C. Smith et al. 2021).

Stellar winds and radiation are assumed to inject momentum at a rate taken from the BPASS stellar population synthesis models (version 2.2.1; J. J. Eldridge et al. 2017; E. R. Stanway & J. J. Eldridge 2018) for the G. Chabrier (2003) IMF with mass limits 0.1 and 100 $M_{\odot}$ that is used throughout. The rate depends on the metallicity and age of the SSP. The momentum flux due to radiation pressure is determined by convolving the emitted energy as a function of wavelength, which depends on the SSP metallicity and age, and is taken from BPASS, with the wavelength-dependent fraction that is absorbed locally by gas or dust. The column densities of gas and dust of the gas surrounding the SSP are computed in the same manner as the column densities used in the radiative cooling module (see Section 3.1), i.e. they are based on the local Jeans length, and the fraction of the radiation that is absorbed is computed using CHIMES. Feedback due to stellar winds and radiation pressure is implemented kinetically, using the stochastic method of C. Dalla Vecchia & J. Schaye (2008) and a kick velocity of 50 km s⁻¹.

While the spatially varying ISRF accounts for the diffuse UV radiation field that is, for example, important for photoelectric dust heating, the ionizing radiation emitted by massive stars has a much shorter mean-free path. In fact, the H II regions created by young SSPs are unresolved in our simulations, unless the density is much lower than typical of the ISM. The ionization of gas by young SSPs is implemented by stochastically labelling gas particle neighbours of young stellar particles as H II regions. Hydrogen and helium in H II regions are assumed to be singly ionized. The same applies to heavy elements in case of full non-equilibrium cooling, but in the fiducial COLIBRE setup heavy element ion fractions are assumed to be in equilibrium (see Section 3.1). Gas particles in H II regions are prevented from cooling below $T = 10^4$ K and from forming stars.

Every 2 Myr, we compute the expected gas mass inside the Strömgen sphere of the H II region centred on the stellar particle powered by the ionizing luminosity of any SSP younger than 50 Myr. The mass in the H II region depends on the ionizing luminosity, which is a function of the age and metallicity of the SSP and is taken from BPASS, and the gas density, which we take to be the current density at the location of the stellar particle. The ratio of the mass in the H II region and the mass in the SPH kernel then yields the probability that a given gas neighbour is tagged as part of an H II region for the duration of the 2 Myr H II region time-step. Overlap between different H II regions is ignored. After the 2 Myr have elapsed the particle's temperature and ion fractions can evolve freely, unless the particle is again tagged as part of an H II region.

3.7 Supernova feedback

In COLIBRE, galactic winds driven by stellar feedback are powered mainly by CCSNe rather than by the early stellar feedback discussed in the previous section. The pre-CCSN feedback can

make the CCSN feedback more efficient by reducing the gas density in the environment of young stars, which in turn reduces energy losses due to radiative cooling.

It has long been known that simulations that distribute the SN energy released by stellar particles at every time-step among their gas neighbours suffer from overcooling, with inefficient feedback resulting in excessively large amounts of stellar mass (e.g. N. Katz, D. H. Weinberg & L. Hernquist 1996). As discussed in C. Dalla Vecchia & J. Schaye (2012), if the sound-crossing time across the resolution element in which SN energy is injected is longer than the (post-shock) radiative cooling time, then the energy-driven expansion phase will be cut short, making feedback inefficient. There are two reasons why this problem is more severe at lower resolution. First, a lower numerical resolution implies a longer sound-crossing time because the resolution elements are larger. Second, and more importantly, if we decrease the resolution, then the radiative cooling time of the heated gas tends to become shorter due to the larger amount of mass over which the SN energy is distributed, which leads to lower peak temperatures. Hence, at the resolutions achievable for cosmological simulations of representative volumes, catastrophic overcooling will occur unless excessive radiative losses are prevented.

Commonly used approaches to suppress numerical overcooling include using multiphase particles and injecting the energy in a hot and dilute subgrid phase (S. Marri & S. D. M. White 2003), injecting the energy in kinetic form and temporarily decoupling wind particles from the hydrodynamics (V. Springel & L. Hernquist 2003), temporarily turning off radiative cooling (e.g. R. J. Thacker & H. M. P. Couchman 2000), or combining the energy from many SNe in order to inject it in amounts sufficient to raise the temperature to values large enough for radiative cooling to initially be inefficient (C. Dalla Vecchia & J. Schaye 2012). COLIBRE employs the last option, variants of which are sometimes referred to as ‘superbubble feedback’ (B. W. Keller et al. 2014).

As in EAGLE, we follow C. Dalla Vecchia & J. Schaye (2012) and inject CCSN energy in thermal form using a stochastic treatment that has as its parameters the injected CCSN energy in units of 10^{51} erg, f_E , and the desired temperature increase of the heated gas, ΔT_{SN} .

The amount of energy from CCSNe that becomes available for a stellar particle of initial (i.e. before stellar mass loss) mass m_* in the time-step from t to $t + \Delta t$ is

$$\Delta E_{\text{CCSN}} = 10^{51} \text{ erg } f_E m_* \int_{m_d(t+\Delta t)}^{m_d(t)} \Phi(m) dm, \quad (15)$$

where $\Phi(m)$ is the IMF, which we assume to be G. Chabrier (2003) and to cover the mass range 0.1–100 $M_{\odot}$, and $m_d(t)$ denotes the mass of the stars that explode as CCSNe at age t . We use the metallicity-dependent stellar lifetime tables from L. Portinari et al. (1998) to compute $m_d(t)$. The function $m_d(t)$ is non-zero only for ZAMS masses between $m_{\text{min,CCSN}} = 8$ and $m_{\text{max,CCSN}} = 100 M_{\odot}$, which correspond to stellar ages of ≈ 40 and ≈ 3 Myr, respectively. Note that EAGLE uses $m_{\text{min,CCSN}} = 6 M_{\odot}$, which yields ≈ 50 per cent more CCSN energy in total.

For each SPH neighbour of a given stellar particle and for each time-step, the probability that the gas particle is heated is equal to the ratio of the energy available for thermal feedback, $(1 - f_{\text{kin}})\Delta E_{\text{CCSN}}$, and the energy required to increase the temperature of all the stellar particle's SPH neighbours by ΔT_{SN} . Here, f_{kin} is the fraction of the CCSN energy injected in kinetic form (see below).

Table 1. Fiducial values of subgrid parameters that vary between models.

Resolution	AGN	$P_{\text{E,pivot}}/k_B$ (K cm ⁻³)	$n_{\text{H,pivot}}$ (cm ⁻³)	$m_{\text{BH,seed}}$ (M _⊙)	$f_{\text{E,min}}$	$\Delta T_{\text{SN,min}}$ (K)	$\Delta T_{\text{SN,max}}$ (K)	$\Delta T_{\text{AGN,max}}$ (K)	$M_{\text{FoF,seed}}$ (M _⊙)	ϵ_f	f_{co}
m5	Thermal	1.5×10^4	1.00	2×10^4	0.8	$10^{7.00}$	$10^{8.0}$	10^{10}	1×10^{10}	0.05	1
m6	Thermal	1.0×10^4	0.50	3×10^4	0.3	$10^{6.75}$	$10^{8.0}$	10^{10}	1×10^{10}	0.05	1
m7	Thermal	8.0×10^3	0.60	$10^{5.5}$	0.1	$10^{6.50}$	$10^{7.5}$	$10^{9.5}$	5×10^{10}	0.10	$10^{-0.5}$
m5	Hybrid	1.5×10^4	1.20	4×10^4	0.8	$10^{7.00}$	$10^{8.0}$	10^{10}	1×10^{10}	0.02	1
m6	Hybrid	1.0×10^4	0.75	$10^{4.75}$	0.3	$10^{6.75}$	$10^{8.0}$	10^{10}	1×10^{10}	0.02	1
m7	Hybrid	8.0×10^3	1.50	5×10^5	0.1	$10^{6.50}$	$10^{7.5}$	$10^{9.5}$	5×10^{10}	0.03	$10^{-0.5}$

Notes. The columns list the simulation resolution; the type of AGN feedback; the pivot pressure for the pressure-dependent CCSN energy fraction, $P_{\text{E,pivot}}/k_B$ (equation 20); the pivot density for the density-dependent SN temperature increment, $n_{\text{H,pivot}}$ (equation 18); the BH seed mass, $m_{\text{BH,seed}}$; the minimum CCSN energy fraction, $f_{\text{E,min}}$; the minimum SN temperature increment, $\Delta T_{\text{SN,min}}$; the maximum SN temperature increment, $\Delta T_{\text{SN,max}}$; the maximum AGN temperature increment, $\Delta T_{\text{AGN,max}}$; the FoF halo mass for BH seeding, $M_{\text{FoF,seed}}$; the efficiency for thermal AGN feedback, and ϵ_f ; the pre-factor for the dust coagulation time (equation 5), f_{co} .

Eq. (18) of C. Dalla Vecchia & J. Schaye (2012) gives the density above which thermal feedback is expected to be ineffective due to catastrophic radiative losses,

$$n_{\text{H,tc}} = 2 \times 10^2 \text{ cm}^{-3} f_t^{-3/2} \left(\frac{m_g}{1.84 \times 10^6 M_\odot} \right)^{-1/2} \left(\frac{\Delta T_{\text{SN}}}{10^{7.5} \text{ K}} \right)^{3/2}, \quad (16)$$

where f_t is the ratio of the radiative cooling time of the heated gas element to the sound-crossing time across the particle's SPH kernel¹⁵ and we assumed the temperature after heat injection to be large compared to the temperature before injection, the gas to be ionized, and the hydrogen mass fraction to be the primordial one.

To avoid numerical overcooling in dense gas, one may wish to use very large values of ΔT_{SN} , but this would lead to poor sampling of feedback events, and it may lead to excessively powerful explosions when feedback does occur. The expectation value for the number of gas particles that are heated over the lifetime of a stellar particle is (eq. 4 of E. Chaikin et al. 2023)

$$\langle N_{\text{heat,tot}} \rangle \approx 0.91 (1 - f_{\text{kin}}) f_E \left(\frac{m_*}{m_g} \right) \left(\frac{\Delta T_{\text{SN}}}{10^{7.5} \text{ K}} \right)^{-1}, \quad (17)$$

where m_g is the (average) mass of gas particles in the simulation. Assuming that $m_* \approx m_g$, $f_E \sim 1$, $f_{\text{kin}} \ll 1$ and requiring that each star particle on average deposits at least one energy injection in its lifetime (i.e. $\langle N_{\text{heat,tot}} \rangle \gtrsim 1$), gives a constraint on the heating temperature $\Delta T_{\text{SN}} \lesssim 10^{7.5} \text{ K}$. This is why EAGLE, which used $f_{\text{kin}} = 0$ and $f_E \sim 1$, adopted $\Delta T_{\text{SN}} = 10^{7.5} \text{ K}$.

However, heating just one gas particle per stellar particle does not sample the feedback well, particularly in less well-resolved galaxies that do not have many young stellar particles. Furthermore, releasing the energy in very energetic, sporadic explosions may create overly large bubbles in the ISM (e.g. Y. M. Bahé et al. 2016). To reduce these potential problems, we make two modifications relative to EAGLE.

First, following E. Chaikin et al. (2023), we inject a fraction $f_{\text{kin}} = 0.1$ of the energy in the form of low-velocity kicks. As detailed in E. Chaikin et al. (2023), the kinetic feedback is implemented by kicking particles in pairs in random but opposite directions in a manner that conserves energy, linear momentum,

¹⁵Note that C. Dalla Vecchia & J. Schaye (2012) adopt $f_t = 10$, which is more conservative than the value of unity to which the coefficient in equation (16) corresponds.

and angular momentum. In this scheme, the actual post-kick velocity of the wind particles relative to the star particle will differ somewhat from the target kick velocity if the particles are moving with respect to the star particles or if the two target particles have different masses. We adopt a target velocity of 50 km s^{-1} , which energy-wise is equivalent to a temperature increment of only $\Delta T \sim 10^5 \text{ K}$, guaranteeing good sampling even for $f_{\text{kin}} = 0.1$. E. Chaikin et al. (2023) show that while the thermal feedback generates a hot phase of the ISM that modulates the SFR by driving galactic outflows, the low-energy kicks increase the turbulent velocity dispersion in the neutral ISM, which in turn helps stabilize the gas against star formation (see equation 6).

Second, instead of using a constant ΔT_{SN} , we make it an increasing function of the gas density,

$$\Delta T_{\text{SN}}(n_{\text{H,SN}}) = 10^{6.5} \text{ K} \left(\frac{n_{\text{H,SN}}}{n_{\text{H,pivot}}} \right)^{2/3}, \quad (18)$$

Here $n_{\text{H,pivot}}$ is a free parameter with resolution-dependent values in the range $0.5 - 1.0 \text{ cm}^{-3}$ for the thermal AGN feedback and $0.75 - 1.5 \text{ cm}^{-3}$ for the hybrid AGN feedback model, respectively (see Section 3.10). The logarithmic slope of $2/3$ is motivated by the cooling argument from C. Dalla Vecchia & J. Schaye (2012), our equation (16). The gas density is computed using the SPH formalism at the location of the stellar particle each time the stellar particle can inject CCSN energy,

$$n_{\text{H,SN}} = \frac{X}{m_p} \sum_{i=1}^{N_{\text{ngb}}} m_{g,i} W(|\mathbf{r}_* - \mathbf{r}_i|, h_*), \quad (19)$$

where X is the primordial hydrogen mass fraction, m_p the proton mass, $m_{g,i}$ is the mass of gas particle i , $\mathbf{r}_*$ and $\mathbf{r}_i$ are the coordinates of the stellar particle and gas particle i , respectively, and W is the SPH kernel function with the stellar particle's smoothing length h_* , and the sum runs over all gas particles within the stellar kernel.

The use of a variable ΔT_{SN} enables better sampling and prevents overly large bubbles in low-density gas, where overcooling is less likely to be an issue. However, to prevent very poor sampling in high-density gas and catastrophic cooling in low-density gas, we limit the allowed temperature increment to the range $\Delta T_{\text{SN,min}} < \Delta T_{\text{SN}} < \Delta T_{\text{SN,max}}$. The extrema were chosen on the basis of a few small test runs. Their values increase with the resolution (see Table 1), but are always in the ranges $10^{6.5} \text{ K} \leq \Delta T_{\text{SN,min}} \leq 10^7 \text{ K}$ and $10^{7.5} \text{ K} \leq \Delta T_{\text{SN,max}} \leq 10^8 \text{ K}$.

A third improvement relative to EAGLE is that we do not use particle weighting, which is similar to mass weighting if the

particles have similar masses, to distribute the energy among the SPH neighbours of a stellar particle. Instead, we use the isotropic implementation of E. Chaikin et al. (2022), who demonstrated that isotropic weighting results in significantly reduced cooling losses compared to mass weighting because the mass tends to be concentrated in denser gas. We consider isotropic weighting to be more physical than mass weighting, since there is no reason why SN energy would be preferentially directed towards denser gas.

Any simulation that uses a subgrid prescription to suppress excessive radiative losses will need to calibrate its parameters (or functional forms) on a chosen set of observables. This is because numerical overcooling can in practice usually not be completely overcome, because radiative losses will occur in nature, and because the most appropriate values of the subgrid parameters can typically not be determined directly from observations or from models. In EAGLE, the parameter f_E increases with decreasing metallicity and with increasing density. The density dependence was critical to overcome overcooling in massive galaxies (R. A. Crain et al. 2015), because ISM densities much higher than n_{H,t_c} are sampled in such galaxies at the resolution of EAGLE. Because COLIBRE includes a multiphase ISM, a dependence on density could introduce a sudden sharp change in energy at constant pressure if a particle transitions from the warm to the cold interstellar gas phase.

We therefore opt to make f_E a function of the thermal gas pressure. Because we do not wish to change the energy budget of a stellar particle as it ages, we make f_E a function of the stellar birth pressure, P_{birth} , i.e. the thermal pressure of its parent gas particle at the time it was converted into the stellar particle.¹⁶ The functional form for $f_E(P_{\text{birth}})$ is the same as that for $f_E(n_{H,\text{birth}})$ in EAGLE,¹⁷

$$f_E(P_{\text{birth}}) = f_{E,\text{min}} + \frac{f_{E,\text{max}} - f_{E,\text{min}}}{1 + \exp\left(-\frac{\log_{10} P_{\text{birth}} - \log_{10} P_{E,\text{pivot}}}{\sigma_P}\right)}, \quad (20)$$

where $f_{E,\text{min}}$ and $f_{E,\text{max}}$ are, respectively, the minimum and maximum energies that can be injected by a single SN, in units of 10^{51} erg, $P_{E,\text{pivot}}$ is the birth pressure for which $f_E = (f_{E,\text{min}} + f_{E,\text{max}})/2$, and the parameter σ_P determines the width of the transition from $f_{E,\text{min}}$ to $f_{E,\text{max}}$. Note that, different from EAGLE, f_E is independent of the metallicity. Note also that because EAGLE assumes stars with mass $6 - 8 M_\odot$ explode as (electron capture) SNe, the EAGLE values of f_E need to be multiplied by a factor 1.5 for a fair comparison with COLIBRE.

Equation (20) implies that stellar particles born in higher gas-pressure environments will inject more CCSN energy. This can be interpreted as an ingredient necessary to overcome residual numerical overcooling, which can occur despite our use of stochastic thermal feedback with a variable ΔT_{SN} because densities can exceed n_{H,t_c} by large factors. It can also be interpreted as a reflection of the expectation that physical cooling losses will be smaller in superbubbles resulting from the clustering of many SNe (e.g. P. Sharma et al. 2014). Values $f_E > 1$ can be justified on numerical grounds, i.e. because such values are needed to overcome numerical overcooling, or on physical grounds, e.g. because some SNe inject kinetic energies greater than 10^{51} erg (e.g. P. A.

Mazzali et al. 2014) or because the IMF may become more top-heavy in starbursts (e.g. Gunawardhana et al. 2011), where gas pressures are expected to be high. Values $f_E < 1$ can be justified because the use of stochastic thermal feedback can result in an underestimate of the physical radiative losses.

As discussed in Section 3.10, we set $f_{E,\text{max}} = 4$ and $\sigma_P = 0.3$. The parameter $f_{E,\text{min}}$ increases with increasing resolution, but is limited to the range 0.1–0.8 for the simulations used here. The value of $P_{E,\text{pivot}}$ also increases with the simulation resolution, with $8 \times 10^3 \text{ K cm}^{-3} < P_{E,\text{pivot}}/k_B < 1.5 \times 10^4 \text{ K cm}^{-3}$ for the simulations presented here (see Table 1), where k_B is the Boltzmann constant.

Finally, we note that energy feedback from SNIa is implemented as for CCSNe, but with $f_E = 1$, $f_{\text{kin}} = 0$, and with equation (15) replaced by

$$\Delta E_{\text{SNIa}} = 10^{51} \text{ erg } f_E m_* \int_t^{t+\Delta t} \dot{\mathcal{N}}_{\text{SNIa}}(t - t_*) dt, \quad (21)$$

where $\dot{\mathcal{N}}_{\text{SNIa}}(t - t_*)$ is the rate of SNIa per unit initial stellar mass for an SSP born at time $t_* < t$, see equation (13). Given that compared to CCSN feedback, SNIa feedback is unimportant for the observables that we calibrate the subgrid feedback to, we set $f_E = 1$ and $f_{\text{kin}} = 0$ for simplicity. Another reason for setting $f_E = 1$ is that residual numerical overcooling is less likely for SNIa than for CCSN feedback, since SNIa tend to explode long after the stars have moved out of the dense environments in which they were born. Our treatment of SNIa feedback differs from EAGLE, where stochastic thermal feedback was used for CCSNe, but the SNIa energy was instead dumped thermally at each time-step, causing the SNIa feedback to be ineffective due to numerical overcooling. Nevertheless, in COLIBRE SNIa feedback remains much less important than CCSNe (for comparisons, see F. S. J. Nobels et al. in preparation).

3.8 Supermassive black holes

The prescription for supermassive BHs for the thermal AGN model is a modified version of that presented in F. S. J. Nobels et al. (2022) and Y. M. Bahé et al. (2022), which is itself based on V. Springel, T. Di Matteo & L. Hernquist (2005), T. Di Matteo et al. (2008), and C. M. Booth & J. Schaye (2009). We will first describe the model ingredients, i.e. BH seeding, gas accretion, mergers, and repositioning, and then discuss the modifications specific to the hybrid AGN feedback simulations.

3.8.1 Seeding

Starting from scale factor $a = 1/(1+z) = 0.05$, after every¹⁸ $\Delta a = 0.00751a$, an FoF halo finder is run on the fly as described in Section 2.3. Haloes more massive than $M_{\text{FoF,seed}}$ that do not yet contain a BH are seeded with a BH of mass $m_{\text{BH,seed}}$. The densest gas particle is then converted into a collisionless BH particle. Because the desired BH seed mass is generally small compared to the mass of the BH particle, $m_{\text{BH,part}}$, BH particles store a subgrid BH mass, $m_{\text{BH}} \leq m_{\text{BH,part}}$. The subgrid BH mass is the mass that controls the physical processes related to BHs, while the BH particle mass is used in the gravity solver.

¹⁶For this purpose, we define $P_{\text{birth}} = \rho_{\text{birth}} k_B T_{\text{birth}} / m_p$.

¹⁷For EAGLE, this function was expressed in the mathematically equivalent form $f_E = f_{E,\text{min}} + (f_{E,\text{max}} - f_{E,\text{min}}) / [1 + (P_{\text{birth}}/P_{E,\text{pivot}})^{n_P}]$ with $n_P = -1/(\sigma_P \ln 10)$.

¹⁸The cadence of the on-the-fly FoF finder used for seeding is the same as in the FLAMINGO simulation, but the value of Δa was misquoted in J. Schaye et al. (2023).

The BH seed mass varies between $m_{\text{BH,seed}} = 2 \times 10^4 M_\odot$ for m5 and $\approx 3 \times 10^5 M_\odot$ for m7 resolution. Such values correspond to the ‘heavy seed’ scenario, where BHs are formed by the direct collapse of a massive gas cloud or by the rapid merging of stars and/or low-mass BHs (e.g. J. Regan & M. Volonteri 2024). The value of $m_{\text{BH,seed}}$ is set during calibration (see Section 3.10). It has some effect on the onset of the rapid BH growth phase through gas accretion, which can lead to quenching of star formation. BHs are seeded in haloes with $M_{\text{FoF,seed}} = 5 \times 10^{10} M_\odot$ for m7 resolution and $M_{\text{FoF,seed}} = 1 \times 10^{10} M_\odot$ for the m6 and m5 resolutions. The minimum value of $M_{\text{FoF,seed}} = 1 \times 10^{10} M_\odot$ guarantees that the haloes are well resolved, i.e. they contain $\gg 10^2$ particles for all resolutions. At m7 resolution, we use a five times higher value to ensure that the subgrid BH mass converges to the BH particle mass at large masses ($m_{\text{BH}} \gg m_g$). A lower value of $M_{\text{FoF,seed}}$ yields more seed BHs and at m7 resolution, the increase in BH particle mass during mergers would then be too large for the subgrid BH mass to be able to catch up to BH particle mass via gas accretion.

3.8.2 Gas accretion

BHs can grow through gas accretion and mergers with other BHs, but growth by the accretion of stars or DM is neglected. Following M. R. M. R. Krumholz, McKee & Klein (2006), the gas accretion rate is given by the Bondi–Hoyle–Lyttleton (BHL) accretion rate (H. Bondi 1952; F. Hoyle & R. A. Lyttleton 1939)

$$\dot{m}_{\text{BHL}} = 4\pi G^2 \frac{\rho_g}{c_s^3} m_{\text{BH}}^2, \quad (22)$$

modified to account for turbulence and vorticity,

$$\dot{m}_{\text{accr}} = \dot{m}_{\text{BHL}} \frac{f_{\text{turb}} f_{\text{ang}}}{(f_{\text{turb}}^2 + f_{\text{ang}}^2)^{1/2}}, \quad (23)$$

which interpolates smoothly between the asymptotic regimes where the accretion rate is limited by turbulence ($\dot{m}_{\text{accr}} \rightarrow f_{\text{turb}} \dot{m}_{\text{BHL}}$ if $f_{\text{turb}} \ll f_{\text{ang}}$) or by vorticity ($\dot{m}_{\text{accr}} \rightarrow f_{\text{ang}} \dot{m}_{\text{BHL}}$ if $f_{\text{turb}} \gg f_{\text{ang}}$).

The turbulence limiter is (M. Ruffert & D. Arnett 1994)

$$f_{\text{turb}} = \left[\frac{\lambda^2 + \mathcal{M}^2}{(1 + \mathcal{M}^2)^4} \right]^{1/2}, \quad (24)$$

with $\lambda = 1.1$ (M. R. M. R. Krumholz et al. 2006) and where $\mathcal{M} = v/c_s$ is the Mach number, with c_s the sound speed of the ambient gas. We take $v^2 = v_{\text{BH}}^2 + \sigma_{\text{turb}}^2$, where v_{BH} and σ_{turb} are, respectively, the bulk velocity relative to the BH and the velocity dispersion of the ambient gas in the frame of the BH. If $\mathcal{M} \ll 1$ then the turbulence-limited accretion rate asymptotes to λ times the BHL rate.

The vorticity limiter is (Krumholz, McKee & Klein 2005)

$$f_{\text{ang}} = 0.34 f(\omega_*), \quad (25)$$

where $\omega_* = \omega r_{\text{B}}/c_s$, $\omega = |\nabla \times \mathbf{v}|$ is the vorticity of the ambient gas, $r_{\text{B}} = Gm_{\text{BH}}/c_s^2 \approx 4 \text{ kpc} (m_{\text{BH}}/10^8 M_\odot)(c_s/10 \text{ km s}^{-1})^{-2}$ is the Bondi radius, and $f(\omega_*) = (1 + \omega_*^{0.9})^{-1}$. If the vorticity is small, i.e. $\omega \ll c_s/r_{\text{B}}$, then the vorticity-limited accretion rate asymptotes to 0.34 times the BHL rate. The reason why the accretion rate asymptotes to a value slightly smaller than the BHL rate is that we assume that even when vorticity is negligible at the Bondi radius, the accreting gas circularizes before it reaches the Schwarzschild radius (Krumholz et al. 2005).

This prescription for gas accretion differs from that of EAGLE, which also used BHL accretion, but followed a different approach to account for the effect of relative motions between the BH and the gas and for the angular momentum of the gas (Y. M. Rosas-Guevara et al. 2015).

The Eddington accretion rate is the accretion rate for which the AGN luminosity equals the Eddington luminosity, i.e. the luminosity for which the radiation pressure due to electron scattering balances the gravity of the BH. Previous simulations such as EAGLE did not allow super-Eddington accretion. However, enforcing the Eddington limit may be inappropriate, e.g. because accretion may be aspherical, photons may be trapped, or accretion originating from larger scales may result in a large inward momentum flux (e.g. D. Izquierdo-Villalba et al. 2023). F. Huško et al. (2025) found that allowing super-Eddington accretion can have a large impact on galaxy evolution at high redshift. In COLIBRE, the gas accretion rate is therefore not limited to the Eddington rate. We do limit it to 100 times the Eddington rate to prevent very large increases in BH mass on possibly poorly resolved time-scales.

The Eddington rate is given by

$$\dot{m}_{\text{Edd}} = \frac{4\pi G m_p m_{\text{BH}}}{\epsilon_r \sigma_{\text{T}} c} \approx 2.2 M_\odot \text{ yr}^{-1} \left(\frac{\epsilon_r}{0.1} \right)^{-1} \left(\frac{m_{\text{BH}}}{10^8 M_\odot} \right), \quad (26)$$

where σ_{T} is the Thomson cross-section, $\epsilon_r = 0.1$ is the assumed radiative efficiency of the subgrid accretion disc (e.g. similar to N. I. Shakura & R. A. Sunyaev 1973), and c is the speed of light. Note that the BHL rate exceeds the Eddington rate for densities greater than

$$n_{\text{H}}(\dot{m}_{\text{BHL}} = \dot{m}_{\text{Edd}}) = \frac{X c_s^3}{\sigma_{\text{T}} G c \epsilon_r m_{\text{BH}}}, \quad (27)$$

$$\approx 2.8 \times 10^{-2} \text{ cm}^{-3} \left(\frac{c_s}{10 \text{ km s}^{-1}} \right)^3 \left(\frac{\epsilon_r}{0.1} \right)^{-1} \left(\frac{m_{\text{BH}}}{10^8 M_\odot} \right)^{-1}, \quad (28)$$

To reach accretion rates 100 times the Eddington rate, 100 times higher densities are required.

Accounting for the energy that is radiated away as the gas spirals in through a subgrid accretion disc, gas accretion causes the BH to gain mass at the rate

$$\dot{m}_{\text{BH}} = (1 - \epsilon_r) \dot{m}_{\text{accr}}, \quad (29)$$

and the BH particle mass to decrease by

$$\dot{m}_{\text{BH,part}} = -\epsilon_r \dot{m}_{\text{accr}}. \quad (30)$$

If gas accretion results in $m_{\text{BH}} > m_{\text{BH,part}}$ then, in order to conserve mass-energy, mass is transferred from the surrounding gas particles to the BH particle mass. The mass and momentum taken from each SPH neighbour are proportional to that neighbour’s contribution to the gas density at the location of the BH. Neighbours with mass less than half the initial baryonic particle mass are excluded. Y. M. Bahé et al. (2022) find that this method of allowing the BH to ‘nibble’ on its neighbours lets the BH subgrid mass track the BH particle mass more closely than if, as in EAGLE, entire gas particles are stochastically swallowed by the BH. Nibbling also prevents the artificially large velocity kicks that occur due to momentum conservation when entire particles are swallowed.

3.8.3 Mergers

BHs merge if three conditions are met. First, they must be separated by less than three gravitational softening lengths, $r <$

3ϵ , where r is their spatial separation. Second, their relative velocity must satisfy $\Delta v < [2G(m_{\text{BH,part},1} + m_{\text{BH,part},2})/r]^{1/2}$, where $m_{\text{BH,part},i}$ is the particle mass of BH i . This second condition differs from EAGLE, which used the C. M. Booth & J. Schaye (2009) criterion $\Delta v < \sqrt{Gm_{\text{BH}}/h}$, where h is the SPH smoothing length of the more massive BH. As discussed in Y. M. Bahé et al. (2022), using r instead of h is more appropriate because merging is a gravitational process and should therefore not depend on the gas density. We use the particle BH mass rather than the subgrid BH mass used in EAGLE and Y. M. Bahé et al. (2022) because in the simulation the particle BH masses determine the particle orbits. Third, the less massive BH must be within the SPH smoothing length of the more massive BH, $r < h$. This last condition enables the code to find neighbouring BHs that are candidates for merging.

The energy lost in gravitational waves during the merging process is assumed to be that for two non-spinning BHs with masses $m_{\text{BH},1}$ and $m_{\text{BH},2}$ and is subtracted from the rest mass energy of the descendant BH, i.e. from the sum of the rest mass energies of the merging BHs, for both the subgrid and the particle masses. It is taken from E. Barausse, V. Morozova & L. Rezzolla (2012),

$$\frac{E_{\text{GW}}}{(m_{\text{BH},1} + m_{\text{BH},2})c^2} \approx 0.057191\nu + (0.037763 + 0.506)\nu^2 \quad (31)$$

where $\nu = m_{\text{BH},1}m_{\text{BH},2}/(m_{\text{BH},1} + m_{\text{BH},2})^2$ and terms $\mathcal{O}(\nu^3)$ are neglected. Here, the $0.506\nu^2$ term accounts for the plunge, merger, and ringdown phases, while the other terms account for the inspiral phase. The energy loss is maximum for equal mass BHs, in which case it is ≈ 4.8 per cent of the total BH rest mass energy. Note that this energy loss was not accounted for in EAGLE.

When two BHs merge, we conserve linear momentum. Gravitational wave recoil is not accounted for. We remove the particle hosting the lower mass BH from the simulation.

3.8.4 Repositioning

BHs are affected by dynamical friction, which causes their orbits to shrink towards the galaxy nucleus. There are several reasons why cosmological simulations will generally underestimate the effects of dynamical friction. First, except for very massive BHs in relatively high-resolution simulations, the BH particle masses are equal to or similar to the masses of other particles, whereas in reality they are many orders of magnitude higher than the masses of DM and atomic particles. Second, the contribution of small-impact parameter scatterings is underestimated due to gravitational softening (e.g. M. Tremmel et al. 2015). Third, nuclear star clusters will dominate the dynamical friction if they are more massive than the BH that they host, as is expected for low-mass galaxies (e.g. N. Neumayer, A. Seth & T. Böker 2020), but such clusters are unresolved (e.g. L. Ma et al. 2021).

To compensate for the inability to numerically resolve dynamical friction, most simulations either move the BH down the potential gradient or pin it to the centre of its halo. Y. M. Bahé et al. (2022) demonstrated that implementing a BH repositioning scheme is crucial, because without any repositioning, BH growth is dramatically reduced, which in turn causes AGN feedback to become ineffective. Our treatment of BH repositioning follows the recommendations of Y. M. Bahé et al. (2022). At each time-step, the BH is moved instantaneously to the location of the SPH neighbour within three gravitational softening lengths that has the lowest gravitational potential, provided it is lower than the potential at the current location of the BH. For the purpose of

repositioning, we subtract the contribution of the BH to the potential in order to prevent the BH from becoming trapped in its own local potential well. The velocity of the BH is not altered when it is repositioned. However, in practice the BH will lose kinetic energy in the subhalo's rest frame as it repeatedly moves up the gravitational potential before being repositioned.

Compared to EAGLE, the most important differences in the repositioning scheme are that EAGLE did not correct the potential for the contribution of the BH and that EAGLE also allowed repositioning on to stellar and DM particles. The choice to only allow BHs to move to the locations of gas particles was motivated by the considerable computational expense associated with finding collisionless neighbours. However, this restriction can have undesired consequences. For example, in gas-poor galaxies, the BH may not track the minimum of the gravitational potential as well as in gas-rich galaxies. BHs hosted by satellites that are being stripped of their gas may move out of their galaxy, though the fact that the BH velocities are unaltered during repositioning will generally cause them to return as long as the satellites retain some of their gas (Y. M. Bahé et al. 2022). Nevertheless, satellites losing their BHs as a result of repositioning is a concern and for future simulations the computational expense associated with improved and more physical prescriptions for dynamical friction (e.g. M. Tremmel et al. 2015; A. Genina, V. Springel & A. Rantala 2024) may be considered worthwhile. However, a fully satisfactory solution to the important problem of modelling dynamical friction in cosmological simulations does not exist, particularly for BHs with $m_{\text{BH}} \lesssim m_{\text{CDM}}$.

3.8.5 BH growth in the hybrid thermal/jet model

As described in Section 3.9, a subset of our runs implement AGN feedback as a combination of thermal feedback and kinetic jet feedback. These runs include a subgrid accretion disc model that is used to track BH spins and accretion disc states, which together determine the nature and efficiency of the AGN feedback. The model, aspects of which are described in F. Huško et al. (2022, 2024, 2025) is motivated and detailed in F. Huško et al. (2026). Here, we will only give a brief overview.

The subgrid accretion disc model for the evolution of the BH spin vector accounts for gas accretion, BH mergers, jet-induced spindown, radiation torques, and Lense–Thirring torques. Equation (31) for the energy lost in gravitational waves during BH mergers is replaced by a spin-dependent version. The rate of growth of the BH through gas accretion, equation (29) for the thermal model, is replaced by

$$\dot{m}_{\text{BH}} = (1 - \epsilon_{\text{r}} - \epsilon_{\text{jet}} - \epsilon_{\text{wind}})\epsilon_{\text{accr}}\dot{m}_{\text{accr}}, \quad (32)$$

where ϵ_{jet} is the jet efficiency, ϵ_{wind} is the accretion disc wind efficiency, and ϵ_{accr} is the accretion efficiency. The latter is the fraction of the accretion rate onto the disc, $\dot{m}_{\text{accr}}$, that is not ejected and actually ends up in the BH (if we ignore the loss of rest mass energy through feedback, i.e. the term in parentheses in equation 32), which is assumed to be unity in the thermal model.

The type of accretion disc depends on the Eddington ratio (at the outer accretion disc scale, i.e. not reduced by the factor ϵ_{accr}),

$$f_{\text{Edd}} \equiv \frac{\dot{m}_{\text{accr}}}{\dot{m}_{\text{Edd}}}. \quad (33)$$

The radiative efficiency, ϵ_{r} , that appears in the denominator of the Eddington accretion rate (equation 26), is replaced by the expression from I. D. Novikov & K. S. Thorne (1973) for a relativistic

Keplerian accretion disc. In this model, the radiative efficiency increases with the BH spin to account for the decrease in the radius of the innermost stable circular orbit with the spin. Hence, the Eddington ratio decreases with BH spin.

The accretion disc can be in one of three different states:

(i) *Thick disc*, $f_{\text{Edd}} < 0.01$. The disc is geometrically thick and optically thin. The accretion efficiency increases with the Eddington ratio and depends on the size of the thick component of the disc. The relation for this size includes a free parameter, $r_{\text{tr},0}$, which corresponds to the dimensionless size, i.e. the size relative to the gravitational radius of the BH, for $f_{\text{Edd}} = 0.01$ and is calibrated to the $z = 0$ AGN bolometric luminosity function.¹⁹

(ii) *Thin disc*, $0.01 < f_{\text{Edd}} < 1$. The disc is geometrically thin and optically thick. The accretion efficiency $\epsilon_{\text{accr}} = 1$.

(iii) *Slim disc*, $f_{\text{Edd}} > 1$. The disc is slim, a state intermediate between the thick and thin discs. The accretion efficiency is set to $\epsilon_{\text{accr}} = 0.01$.

The fraction of BHs in the slim disc state is small for all BH masses and redshifts and is only non-negligible for intermediate-mass ($10^6 \lesssim m_{\text{BH}}/M_{\odot} \lesssim 10^8$) BHs at high redshift ($z > 3$). The thick (thin) disc fraction decreases (increases) with redshift. At a fixed redshift, it is minimum (maximum) for $m_{\text{BH}} \sim 10^7 M_{\odot}$. At $z = 0$, the thick disc state is dominant at all masses, while for $z > 2$ BHs with $m_{\text{BH}} > 10^6 M_{\odot}$ are mostly in the thin disc state.

The seed BH mass is $m_{\text{BH,seed}} = 4 \times 10^4 M_{\odot}$, $10^{4.75} M_{\odot} \approx 5.6 \times 10^4 M_{\odot}$, and $5 \times 10^5 M_{\odot}$ for m5, m6, and m7 resolution, respectively. Together with the value for the pivot density for SN feedback (equation 18), its value is calibrated to reproduce the observed $z = 0$ galaxy SMF and SMR (see Section 3.10).

3.9 AGN feedback

In the fiducial simulations, AGN feedback is implemented using a purely thermally driven mode. Thermal AGN feedback is implemented using a method based on C. M. Booth & J. Schaye (2009). To prevent numerical overcooling, the energy is not injected at every time-step but in quantities sufficient to increase the temperature of a gas particle by a pre-defined amount specified by the parameter ΔT_{AGN} . This is similar to the subgrid recipe for SN feedback (see the discussion in Section 3.7) and to the treatment of AGN feedback in EAGLE, except that the prescription is not stochastic. Instead, as in C. M. Booth & J. Schaye (2009), each BH saves its AGN feedback energy in a reservoir until there is sufficient energy for a single feedback event, i.e. to increase the temperature of at least one gas particle by²⁰ ΔT_{AGN} . The energy is then injected into the SPH neighbour(s) nearest to the BH. This differs from the isotropic scheme used to distribute the SN energy (see Section 3.7), but E. Chaikin et al. (2022) show that the use of the nearest particle results in a similar feedback efficiency and a similar degree of isotropy as the isotropic scheme. For AGN feed-

back, we prefer to use the closest particle because the distance to SPH neighbours can become large in gas-poor galaxies.

The rate at which AGN feedback energy becomes available is given by

$$\dot{E}_{\text{AGN}} = \epsilon_f \epsilon_r \dot{m}_{\text{accr}} c^2, \quad (34)$$

while the observable AGN luminosity is

$$L_{\text{AGN}} = (1 - \epsilon_f) \epsilon_r \dot{m}_{\text{accr}} c^2, \quad (35)$$

where ϵ_f is the fraction of the intrinsic luminosity, $\epsilon_r \dot{m}_{\text{accr}} c^2$, that is assumed to couple to the ambient gas. Due to self-regulation, the value of ϵ_f strongly affects the masses of BHs in galaxies that are sufficiently massive for their growth to be regulated by AGN feedback (C. M. Booth & J. Schaye 2009, 2010). The value of ϵ_f is difficult to predict from first principles and was chosen to yield reasonable BH masses in high-mass galaxies at $z = 0$. We use $\epsilon_f = 0.1$ for our m7-resolution runs, but 0.05 for the m6 and m5 resolutions.

The energy required for a single feedback event is

$$\Delta E_{\text{AGN,thermal}} = \frac{m_g k_B \Delta T_{\text{AGN}}}{(\gamma - 1) \mu m_p}, \quad (36)$$

where γ is the ratio of specific heats, which we assume to be the value for a monatomic gas, $\gamma = 5/3$, and μ is the mean molecular weight, which we will assume to be 0.6, corresponding to an ionized gas of primordial composition. The time interval between feedback events is

$$\begin{aligned} \Delta t_{\text{AGN}} &= \Delta E / \dot{E}_{\text{AGN}}, \\ &= 3.8 \times 10^4 \text{ yr} \left(\frac{\epsilon_f}{0.05} \right)^{-1} \left(\frac{m_g}{1.84 \times 10^6 M_{\odot}} \right) \\ &\quad \times \left(\frac{m_{\text{BH}}}{10^8 M_{\odot}} \right)^{-1} \left(\frac{\Delta T_{\text{AGN}}}{10^9 \text{ K}} \right) \left(\frac{\dot{m}_{\text{accr}}}{\dot{m}_{\text{Edd}}} \right)^{-1}. \end{aligned} \quad (37)$$

Unlike EAGLE, we opt to scale ΔT_{AGN} with the BH mass,

$$\Delta T_{\text{AGN}} = 10^9 \text{ K} \left(\frac{m_{\text{BH}}}{10^8 M_{\odot}} \right). \quad (38)$$

This functional form makes Δt_{AGN} , and hence the time sampling of AGN feedback, independent of BH mass if the accretion rate is proportional to the Eddington rate, i.e. for a fixed Eddington ratio. The use of a variable heating temperature improves the sampling of AGN feedback for low BH masses, which would otherwise have very long Δt_{AGN} , particularly given that the accretion rate is typically sub-Eddington. Because the change to a variable ΔT_{AGN} was made after the calibration of the subgrid model using a constant $\Delta T_{\text{AGN}} = 10^9 \text{ K}$ (E. Chaikin et al. 2025a), the normalization of equation (38) was chosen to coincide with the value used in the calibration for the mass of BHs typical of galaxies transitioning from a state of active star formation to passive evolution.

To prevent catastrophic overcooling at low BH masses, as well as poor sampling of AGN feedback for lower accretion rates at high BH masses, we limit $10^{6.5} \text{ K} \leq \Delta T_{\text{AGN}} \leq \Delta T_{\text{AGN,max}}$, where $\Delta T_{\text{AGN,max}} = 10^{9.5} \text{ K}$ in m7-resolution simulations and 10^{10} K for the m6 and m5 resolutions. We opted for a higher maximum value for the higher resolutions after some experimentation revealed that this was necessary to keep the AGN feedback efficient, because higher gas densities, which would otherwise imply greater radiative losses, are reached in the higher resolution simulations. Note that the energy injected per feedback event scales as $m_g \Delta T_{\text{AGN}}$ and therefore decreases from m7 to m6 and from m6

¹⁹Although the thick disc tends to produce too low bolometric luminosities to affect the comparison to the data, the accretion efficiency affects the efficiency of feedback and hence the accretion rates and thus the fraction of BHs in the thin disc state that dominates the higher luminosities.

²⁰More than one particle can be heated if the BH accretes a sufficient amount of gas during a single time-step. BHs are allowed to heat a maximum of 50 gas neighbours in a given time-step; if this number is insufficient to use up the available AGN feedback energy, then ΔT_{AGN} is increased.

to m5 resolution. Hence, the sampling of AGN feedback improves if the resolution increases.

The efficiency of AGN feedback on resolved scales, i.e. its impact on the resolved properties of galaxies and the gas around them, is mainly determined by the value of ΔT_{AGN} . Higher values of ΔT_{AGN} generally imply stronger feedback, provided the value is still low enough for the feedback to be well sampled. While the value of ϵ_f had little effect on the emerging strength of AGN feedback in the EAGLE simulations, which can be understood in terms of self-regulated AGN feedback²¹ (C. M. Booth & J. Schaye 2010), in COLIBRE, it has a more significant effect. This is because ΔT_{AGN} now depends on the mass of the BH, which, in turn, is sensitive to ϵ_f .

Because BHs are collisionless particles, their time-steps are determined by the gravity solver and can be long compared to Δt_{AGN} . To ensure that the sampling of AGN feedback is not compromised, which could lead to runaway BH growth, we use Δt_{AGN} as a time-step limiter for BHs, but with a floor of 100 yr. In addition, we limit the BH time-step to at most four times the minimum time-step of the BH's SPH neighbours. Note that this does not prevent time-stepping issues altogether because within its time-step, the BH cannot respond to changes in the physical conditions of the ambient gas. An improvement for future simulations would be to also use the hydro time-step of the gas around the BH to wake up drifting BH particles.

3.9.1 AGN feedback in the hybrid thermal/jet model

The hybrid model for AGN feedback is detailed and tested in F. Huško et al. (2026). Here, we will only give a brief summary.

In the hybrid AGN feedback model, kinetic jets are assumed to be present in each of the three accretion disc regimes discussed in the previous section (thick, thin, and slim). In the thick and slim disc states, there is in addition an accretion disc wind, while for thin discs, there is instead the radiation pressure driven wind that may originate from the accretion disc and/or from larger (but still unresolved) scales than the accretion disc wind. The radiatively driven and accretion disc winds are both injected as thermal energy using the method described for the thermal model but with different efficiencies. The energy from jets is injected in kinetic form in directions that are nearly parallel and antiparallel to the BH spin axis. There are separate energy reservoirs for thermal and jet feedback. The dominant feedback mechanisms in the hybrid model are thermal feedback from the thin disc (important at all redshifts), jets from the thin disc (important at $z \gtrsim 1$) and jets from the thick disc (important at $z \lesssim 1$).

The total AGN feedback power is

$$\dot{E}_{\text{AGN}} = (\epsilon_f \epsilon_r + \epsilon_{\text{jet}} + \epsilon_{\text{wind}}) \epsilon_{\text{accr}} \dot{m}_{\text{accr}} c^2. \quad (39)$$

The model assumes a magnetically arrested disc (MAD; R. Narayan, I. V. Igumenshchev & M. A. Abramowicz 2003) with jets powered by the R. D. Blandford & R. L. Znajek (1977) mechanism. The efficiencies ϵ_r , ϵ_{jet} , and ϵ_{wind} are not free parameters. They are functions of the BH spin, which is tracked by the spin evolution model, and, except for ϵ_r , the magnetization of the disc, where the magnetic fields are determined by the

accretion rate and BH spin under the MAD assumption. For the thin disc, the accretion disc wind efficiency $\epsilon_{\text{wind}} = 0$, while for the thick and slim discs, the coupling efficiency for radiative feedback $\epsilon_r = 0$. For the thin disc, ϵ_r is a free parameter that is calibrated to fit the observed $z = 0$ relation between BH mass and galaxy stellar mass. We use $\epsilon_r = 0.02$ for the m5 and m6 resolutions, and $\epsilon_r = 0.03$ for the m7 resolution.

The observable AGN luminosity is

$$L_{\text{AGN}} = (1 - \epsilon_f) \epsilon_r \epsilon_{\text{accr}} \dot{m}_{\text{accr}} c^2. \quad (40)$$

For the thick and slim discs $\epsilon_r = 0$ and this becomes $L_{\text{AGN}} = \epsilon_f \epsilon_{\text{accr}} \dot{m}_{\text{accr}} c^2$. For the thin disc, which dominates the AGN luminosity function for detectable luminosities, we have $\epsilon_{\text{accr}} = 1$ and thus $L_{\text{AGN}} = (1 - \epsilon_f) \epsilon_r \dot{m}_{\text{accr}} c^2$. While the expression for the thin disc luminosity is the same as for the thermal AGN feedback model (equation 35), we emphasize that in the hybrid model the radiative efficiency depends on spin, while it is constant in the thermal model.

Jets are implemented by kicking two particles in opposite directions. The jet direction is limited to be within 7.5 deg from the BH spin axis, but is otherwise random. Equations (36) and (38) are, respectively, replaced by

$$\Delta E_{\text{AGN,jet}} = 2 \times \frac{1}{2} m_g v_{\text{jet}}^2, \quad (41)$$

and

$$v_{\text{jet}} = v_{\text{jet},0} \left(\frac{m_{\text{BH}}}{10^9 M_\odot} \right)^{1/2}, \quad (42)$$

where $v_{\text{jet},0} = 10^{4.5} \text{ km s}^{-1}$ and the jet velocity is limited to the range $10^{2.5} \text{ km s}^{-1} \leq v_{\text{jet}} \leq v_{\text{jet},0}$. To ensure conservation of energy, the actual change in the velocity of a particle kicked into a jet differs from v_{jet} if the particle is not at rest in the frame of the BH.

There are three BH time-step limiters in addition to the one for thermal AGN feedback. The first is analogous to the one for thermal feedback, but applied to jet feedback. The second and third limiters ensure that the relative changes in both the magnitude and the direction of the spin are small.

3.10 Calibration of the feedback

The calibration of the feedback is detailed in E. Chaikin et al. (2025a). Briefly, we first created a Gaussian process emulator using a two-level Latin hypercube of 48 L050m7 simulations, varying four subgrid parameters. The four subgrid parameters that are sampled in the hypercube are the BH seed mass, $m_{\text{BH,seed}}$, which affects the onset of AGN feedback, and the following three stellar feedback parameters: the fraction of energy injected as low-velocity (50 km s^{-1}) kinetic feedback, f_{kin} ; the gas density for which the temperature of gas that is directly heated by SNe is increased by $\Delta T_{\text{SN}} = 10^{6.5} \text{ K}$ (see equation 18), $n_{\text{H,pivot}}$; and the thermal stellar birth pressure for which the injected CCSN energy is halfway between the minimum and maximum allowed values (see equation 20), $P_{\text{E,pivot}}$.

The emulator was then used to fit to the observed S. P. Driver et al. (2022) $z = 0$ galaxy SMF from the Galaxy And Mass Assembly (GAMA) survey and the $z < 0.05$ galaxy size–SMR from J. A. Hardwick et al. (2022), where size refers to the projected half-mass radius. We chose these particular data sets because they are recent and span the required mass range. An advantage of the S. P. Driver et al. (2022) mass function, which they show is in good agreement with other measurements, is that it is provided precisely at $z = 0$. While most observational studies report half-

²¹Self-regulated feedback results in outflows balancing inflows when averaged over suitable length and time-scales. A higher (lower) value of ϵ_f means that less (more) mass needs to be accreted by the BH to power the outflow rate that balances the galaxy-scale inflow rate. This then results in smaller (larger) BH masses, but a similar outflow rate.

light radii, J. A. Hardwick et al. (2022) present half-mass radii, which allow for a direct comparison with the simulations without having to produce virtual observations accounting for dust attenuation. These observables were both fit at $z = 0$ over the galaxy stellar mass range $9.0 < \log_{10} M_*/M_\odot < 11.3$, where the lower and upper limits are motivated by the resolution and box size, respectively. These mass limits are well matched to the mass range spanned by the size measurements of J. A. Hardwick et al. (2022), $9.11 < \log_{10} M_*/M_\odot < 11.26$, while the S. P. Driver et al. (2022) SMF extends to both lower and higher masses. Equal weight was assigned to the SMF and SMR. The rounded best-fitting subgrid parameter values were adopted for the fiducial m7 model.

For the higher resolutions, the parameters $n_{\text{H,pivot}}$, $P_{\text{E,pivot}}$, and $m_{\text{BH,seed}}$, but not f_{kin} , were then slightly adjusted by trial and error to achieve convergence for these observables in small simulation volumes (mainly L025m6 and L012.5m5). Some adjustments were also made to $\Delta T_{\text{SN,min}}$, $\Delta T_{\text{SN,max}}$, $\Delta T_{\text{AGN,max}}$, and $f_{\text{E,min}}$, which tend to increase with increasing resolution. In addition, the m5 and m6 resolutions use a five-times smaller value of $M_{\text{FoF,seed}}$ than m7. Finally, for m7 resolution, we reduced the dust coagulation time to improve convergence for the dust grain sizes. Although the higher resolution simulations can be used to probe lower galaxy masses, we did not consider observations of galaxy sizes below the lowest mass covered by the J. A. Hardwick et al. (2022) data, $\log_{10} M_*/M_\odot = 9.11$. Note that because we recalibrate after changing the resolution, we are aiming for ‘weak convergence’ in the language of J. Schaye et al. (2015).

See Table 1 for the values of the parameters that change with resolution. Only three of these parameters, $P_{\text{E,pivot}}$, $n_{\text{H,pivot}}$, and $m_{\text{BH,seed}}$, are critically important for the calibration to the $z = 0$ SMF and SMR. Because there are degeneracies between some of the parameters in the table, this choice of critical parameters is not unique. Because the degeneracies are only partial, adjustments in some of the remaining parameters can further improve the fits. E. Chaikin et al. (2025a) show the effects that changing the values of individual parameters has on the calibration observables for m7 resolution.

E. Chaikin et al. (2025a) demonstrate that simpler versions of the model for CCSN feedback than our fiducial prescription cannot simultaneously fit the SMF and the SMR. A constant energy (i.e. constant f_{E}), purely thermal model with fixed $\Delta T_{\text{SN}} = 10^{7.5}$ K (the fiducial value of ΔT_{SN} from C. Dalla Vecchia & J. Schaye 2012 used in EAGLE) cannot fit the SMF for any value of f_{E} . Injecting a fraction of the energy in kinetic form, with f_{kin} a free parameter and using a target velocity of 50 km s^{-1} , can fit either the SMF or the SMR, but not both simultaneously. Adding a density-dependent ΔT_{SN} improves the fits, but good fits are only obtained when the available energy, i.e. f_{E} , is made to vary, which we choose to do as a function of the thermal pressure (see Section 3.7).

The efficiency of the thermal AGN feedback, ϵ_f , which is the fraction of the energy radiated by the BH accretion disc that is used to heat the ambient gas, is calibrated to give a reasonable match to the observed relation between BH and stellar masses for massive galaxies in the local Universe (see Section 7.1.3). Previous simulations using the same prescription and similar resolution found that values of order 10 per cent yield acceptable results (e.g. C. M. Booth & J. Schaye 2009; J. Schaye et al. 2015). We therefore set $\epsilon_f = 0.1$ and found no reason to change it for m7 resolution. Its value was halved for the higher resolution simulation to improve convergence.

Similarly to when we change the resolution, when we switch to the hybrid AGN feedback model, we use the best-fitting values of the thermal models as the starting point. Two of these parameters, $n_{\text{H,pivot}}$ and $m_{\text{BH,seed}}$, as well as some parameters specific to the hybrid AGN model, were adjusted by trial and error using small-volume test simulations. The calibration of the simulations using hybrid AGN feedback is discussed in detail in F. Huško et al. (2026).

4 THE SIMULATIONS

The COLIBRE simulation suite includes two types of models, simulations with thermally driven AGN feedback and simulations with hybrid, i.e. thermal plus jet, AGN feedback. There are three resolutions, corresponding to mean particle masses (both for baryons and CDM) of $\sim 10^5$, 10^6 , and $10^7 M_\odot$. Comoving box sizes start at 25 cMpc and increase by powers of 2. The maximum box sizes are 100, 200, and 400 cMpc for the high-, intermediate-, and low-resolution runs, respectively. All but the 50 and 100 cMpc high-resolution simulations have already reached $z = 0$.

Table 2 lists all hydrodynamic simulation runs. The run identifier indicates the box size in cMpc (prefix ‘L’), the rounded log base 10 of the initial mean particle mass in solar masses (prefix ‘m’) and the type of AGN feedback (default is thermal, ‘h’ indicates hybrid). For example, run L200m6 is a box of 200 cMpc on a side with a particle mass $\sim 10^6 M_\odot$ and using purely thermally driven AGN feedback.

Due to their computational expense,²² the largest runs, which use $5 \times 3008^3 \approx 1.4 \times 10^{11}$ particles, are run only with thermal AGN feedback. The largest hybrid AGN feedback runs use 5×1504^3 particles, i.e. eight times fewer. All runs start with four times more CDM than baryonic particles.

The gravitational softening length is the same for baryons and CDM. It is initially held constant in comoving units at $\epsilon_{\text{com}} = 0.9, 1.8$, and 3.6 ckpc for m5, m6, and m7, respectively. This corresponds to $\approx 1/37$ and $\approx 1/23$ times the mean interparticle spacing of baryons and CDM, respectively. At redshift $z < 1.57$, the softening length is kept constant in proper units, which yields a maximum physical softening length of $\epsilon_{\text{prop}} = 0.35, 0.7$, and 1.4 pkpc for m5, m6, and m7, respectively.

For each hydrodynamic simulation, there is a corresponding DMO simulation for which the ICs were created by converting the baryonic particles into CDM particles. Hence, the DMO and hydrodynamic simulations use the same total number of particles and identical ICs. This allows for more accurate comparisons between hydrodynamic and DMO simulations than the usual approach of simply omitting the baryonic component and scaling up the CDM particle masses accordingly because we maintain the separate transfer functions and do not change the particle masses.

5 VISUAL IMPRESSIONS

In this section, we present images that give visual impressions of the COLIBRE simulations.

The top-left panel of Fig. 3 shows a projection of the L025m5 simulation at $z = 0.1$. The colour coding indicates the baryon surface density. The image shows the cosmic web with haloes distributed mostly along filaments. All other panels show stellar

²²Runs L200m6 and L400m7 took, respectively, 72 and 34 million core hours on the Cosma8 computer in Durham, UK. Both simulations were run on 160 nodes that each have 128 cores.

Table 2. COLIBRE hydrodynamical simulations for which results are presented here. The simulations are grouped by type of AGN feedback and sorted by decreasing resolution and then by decreasing box size.

Identifier	L (cMpc)	N_b	N_{CDM}	m_g ($M_\odot$)	m_{CDM} ($M_\odot$)	ϵ_{com} (ckpc)	ϵ_{prop} (pkpc)	z
Fiducial AGN feedback								
L100m5	100	3008 ³	4×3008^3	2.30×10^5	3.03×10^5	0.9	0.35	6.0
L050m5	50	1504 ³	4×1504^3	2.30×10^5	3.03×10^5	0.9	0.35	1.1
L025m5	25	752 ³	4×752^3	2.30×10^5	3.03×10^5	0.9	0.35	0
L200m6	200	3008 ³	4×3008^3	1.84×10^6	2.42×10^6	1.8	0.7	0
L100m6	100	1504 ³	4×1504^3	1.84×10^6	2.42×10^6	1.8	0.7	0
L050m6	50	752 ³	4×752^3	1.84×10^6	2.42×10^6	1.8	0.7	0
L025m6	25	376 ³	4×376^3	1.84×10^6	2.42×10^6	1.8	0.7	0
L400m7	400	3008 ³	4×3008^3	1.47×10^7	1.94×10^7	3.6	1.4	0
L200m7	200	1504 ³	4×1504^3	1.47×10^7	1.94×10^7	3.6	1.4	0
L100m7	100	752 ³	4×752^3	1.47×10^7	1.94×10^7	3.6	1.4	0
L050m7	50	376 ³	4×376^3	1.47×10^7	1.94×10^7	3.6	1.4	0
L025m7	25	188 ³	4×188^3	1.47×10^7	1.94×10^7	3.6	1.4	0
Hybrid AGN feedback								
L025m5h	25	752 ³	4×752^3	2.30×10^5	3.03×10^5	0.9	0.35	0
L050m5h	50	1504 ³	4×1504^3	2.30×10^5	3.03×10^5	0.9	0.35	2.5
L100m6h	100	1504 ³	4×1504^3	1.84×10^6	2.42×10^6	1.8	0.7	0
L050m6h	50	752 ³	4×752^3	1.84×10^6	2.42×10^6	1.8	0.7	0
L025m6h	25	376 ³	4×376^3	1.84×10^6	2.42×10^6	1.8	0.7	0
L200m7h	200	1504 ³	4×1504^3	1.47×10^7	1.94×10^7	3.6	1.4	0
L100m7h	100	752 ³	4×752^3	1.47×10^7	1.94×10^7	3.6	1.4	0
L050m7h	50	376 ³	4×376^3	1.47×10^7	1.94×10^7	3.6	1.4	0
L025m7h	25	188 ³	4×188^3	1.47×10^7	1.94×10^7	3.6	1.4	0

Notes. The columns list the simulation identifier [where ‘L’ is followed by the side length of the cubic simulation volume in cMpc; ‘m’ is followed by the rounded $\log_{10}$ of the mean initial particle mass in solar masses (for both baryons and DM); ‘h’ indicates hybrid AGN feedback]; the comoving box side length, L ; the number of baryonic particles, N_b ; the number of CDM particles, N_{CDM} ; the initial mean baryonic particle mass, m_g ; the mean CDM particle mass, m_{CDM} ; the comoving gravitational softening length, ϵ_{com} ; the maximum proper gravitational softening length, ϵ_{prop} ; the redshift which the simulation had reached at the time of writing. Most combinations of resolution and AGN feedback type are also available in a L012.5 volume.

light accounting for attenuation by dust, based on the predicted spatial distribution, chemical composition, and sizes of the dust grains. The insets on the right show successive zooms into two galaxy groups. The bottom two rows show face- and edge-on views of four galaxies with mass $M_* \sim 10^{11} M_\odot$. These galaxies are numbered and their locations are indicated in the main panel. Galaxies 1–3 are spirals with mass $M_* \approx 10^{10.8} M_\odot$. Galaxies 1 and 2 are relatively isolated, but galaxy 3 is in the process of merging with a companion galaxy. Galaxy 4 is the most massive object in the group shown in the middle-right panel. It is a gas-poor galaxy with mass $M_* \approx 10^{11.3} M_\odot$ and an ellipsoidal shape. For the disc galaxies, dust lanes are clearly visible in the edge-on views as well as in the spiral arms for face-on views. Fig. 3 illustrates that the L025m5 simulation resolves the detailed internal structure of realistic-looking galaxies in a variety of cosmic environments.

The procedure for making stellar-light images is detailed in F. Huško et al. (in preparation). Briefly, luminosities are generated using the package Flexible Stellar Population Synthesis (FSPS; C. Conroy, J. E. Gunn & M. White 2009; C. Conroy & J. E. Gunn 2010) based on the stellar particle ages, metallicities, and redshift. The sampling of stars with ages smaller than 10 Myr is improved by using star-forming gas particles. A 3D luminosity grid is then combined with a 3D grid of the dust distribution. Dust attenuation accounts for absorption and scattering, and for the size

and chemical composition of the dust grains. A red giant branch image is created following R. Lupton et al. (2004) for the *Hubble Space Telescope* (HST) ACS/WFC F475W (blue), F625W (green), and F775W (red) bands.

Fig. 4 compares the spatial distributions of different components of the ISM in face- and edge-on views of a spiral galaxy at $z = 0$ in the L025m5 simulation. The galaxy has mass $M_* = 8 \times 10^{10} M_\odot$ and $\text{SFR} = 2.7 M_\odot \text{yr}^{-1}$. From left to right, the different panels show stellar light, H I column density, H₂ column density times 2, and dust surface density. The central few kpc host a stellar bar and are poor in gas and dust. This nuclear region is surrounded by a ring of dusty, largely molecular gas. Compared with the atomic gas, the molecular gas and, to a lesser extent, the dust are more concentrated toward the centre and the spiral arms.

Fig. 5 illustrates the effect of simulation resolution on galaxy images. We compare HST colour images of the stellar light of the corresponding $z = 0$ galaxies in the L025m5 (top), L025m6 (middle), and L025m7 (bottom) simulations. From left to right, the different columns show galaxies with stellar mass $\log_{10} M_*/M_\odot \approx 8.5, 9.5, 10.3, 10.8$, and 11.3 in L025m6, with each column showing the same galaxy matched across three resolutions. These values vary by ≈ 0.1 dex between the different resolutions. Such small differences are consistent with the typical amplitudes of random deviations found when repeating cosmological simula-

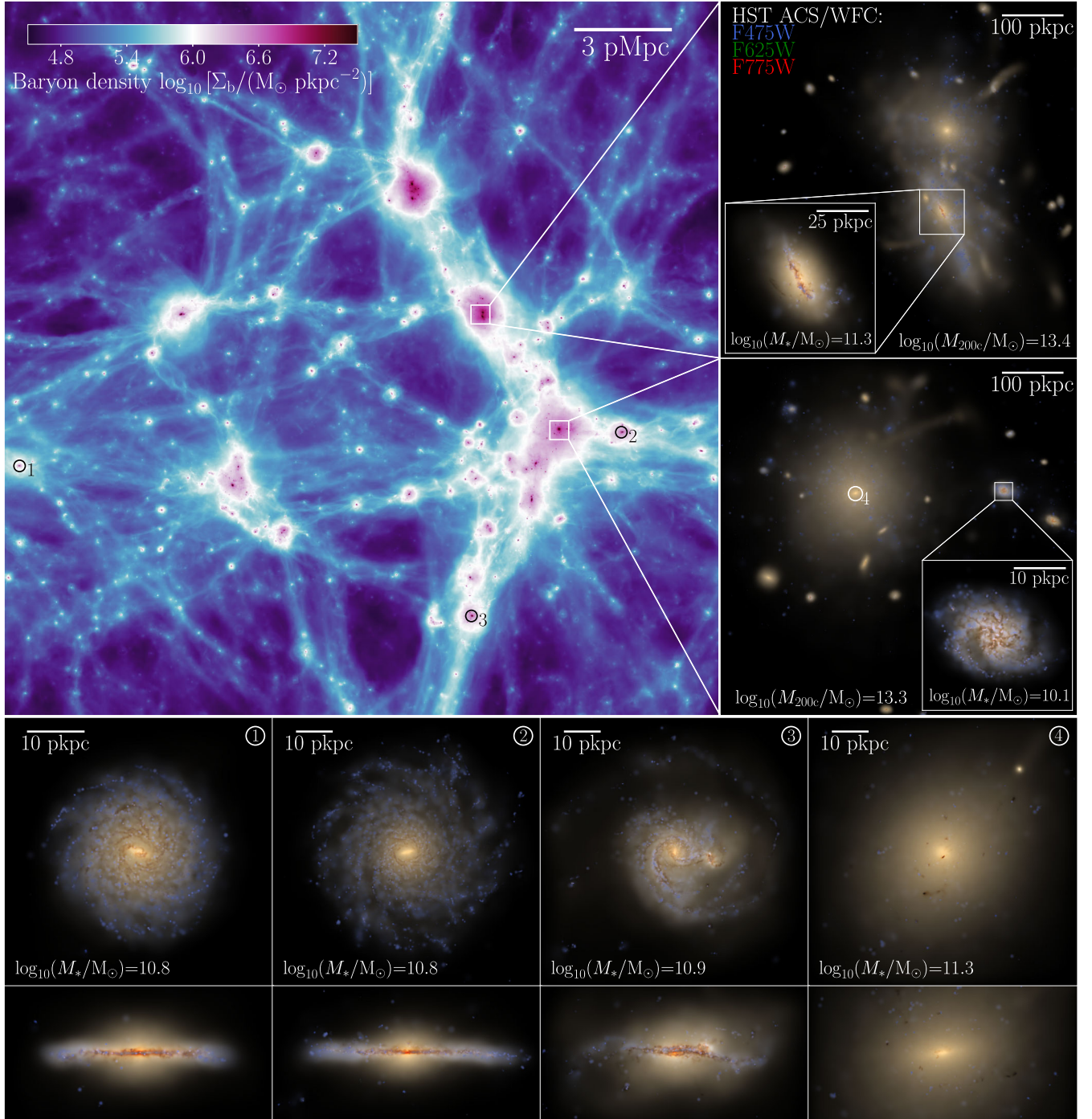


Figure 3. Visual impression of the dynamic range in the high-resolution COLIBRE simulation L025m5 at $z = 0.1$. The top-left panel shows a projection of the entire simulation with the colour encoding baryon surface density. The other panels zoom into different regions and show the stellar light in *HST* colours accounting for attenuation by dust. The top-right and middle-right panels zoom into 800 pkpc cubic regions of groups of mass $M_{200c} = 10^{13.4}$ and $10^{13.3} M_{\odot}$, respectively. The insets in these two panels show 75 pkpc $\times$ 75 pkpc and 30 pkpc $\times$ 30 pkpc zoom-in images of galaxies with stellar mass $M_* = 10^{11.3}$ and $10^{10.1} M_{\odot}$, respectively. The orientations of these two galaxies are the same as in the other images of the same regions. The two bottom rows show face- and edge-on views of four galaxies with stellar masses in the range $10^{10.8} \leq M_*/M_{\odot} \leq 10^{11.3}$. These galaxies are numbered and their locations are indicated in the top-left or middle-right panels. Galaxy zooms show regions of 75 pkpc $\times$ 75 pkpc for face-on orientations and 75 pkpc $\times$ 30 pkpc for edge-on orientations, except for the zoom of galaxy 1 on the bottom left, which is 50 pkpc $\times$ 50 pkpc and 50 pkpc $\times$ 20 pkpc for the face- and edge-on orientations, respectively. COLIBRE simultaneously models the large-scale cosmic web and the internal structure of galaxies.

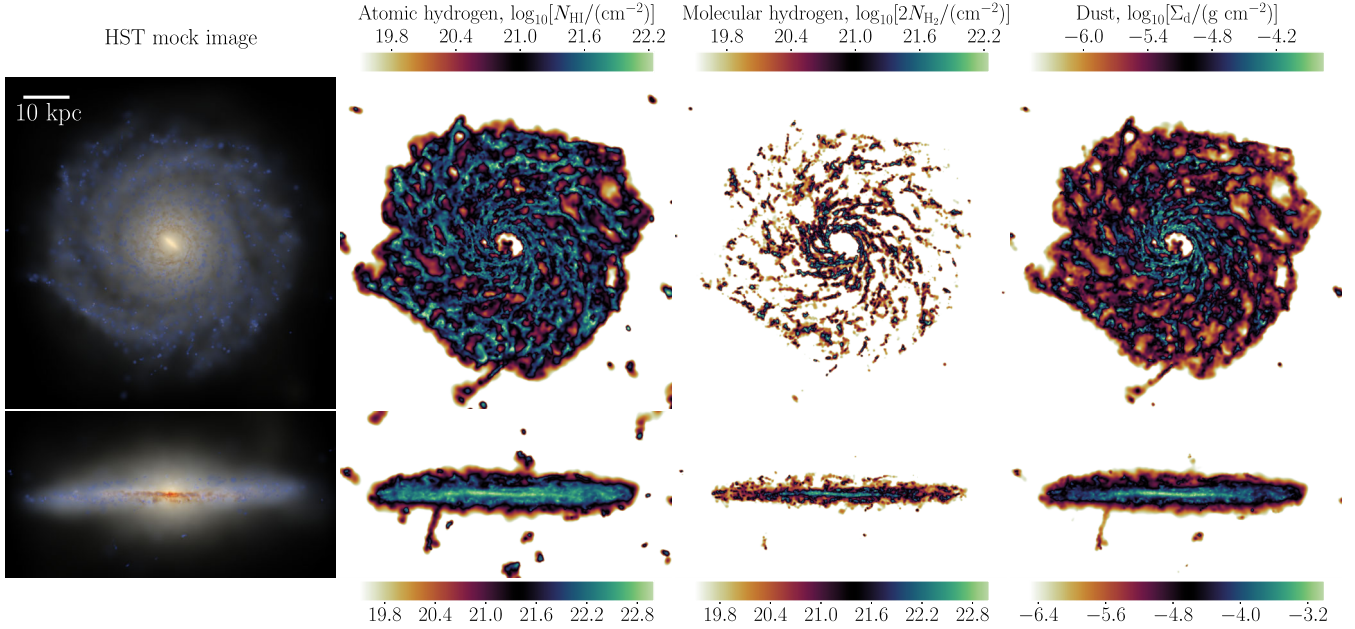


Figure 4. Face-on (top) and edge-on (bottom) view of a $z = 0$ disc galaxy of mass $M_* = 8 \times 10^{10} M_\odot$ and $\text{SFR} = 2.5 M_\odot \text{ yr}^{-1}$ in the L025m5 simulation. From left to right, the columns show stellar light, H I column density, H₂ column density times 2, and dust surface density. The image sizes are 75 kpc $\times$ 75 kpc for face-on and 75 kpc $\times$ 30 kpc for edge-on orientations. Beyond the gas-poor centre, molecular hydrogen is more concentrated towards the centre and in the spiral arms than atomic hydrogen, while the dust distribution is intermediate between that of atomic and molecular hydrogen (compare e.g. the green in the H₂ map with the red in the dust map).

tions that, like COLIBRE, include stochastic subgrid models (J. Borrow et al. 2023). Except for the lowest mass galaxy (the two lowest mass galaxies), resolution m6 (m7) suffices to classify the galaxy morphology, determine its orientation, and measure spatially integrated properties such as the mass and SFR. However, higher resolution simulations reveal increasingly detailed substructure and produce thinner discs.

Fig. 6 gives a visual impression of the L200m6 simulation. The four main panels show 10 Mpc thick slices, with the hue (brightness) indicating the temperature (gas surface density). Clockwise from top left the different same-size panels show images of 200, 30, 3, and 0.3 Mpc on a side, successively zooming into the outskirts of a galaxy cluster. The two small images in the bottom left panel show stellar light in *Euclid* bands, accounting for dust attenuation. As the figure shows, L200m6 has a volume that is sufficiently large to include many galaxy clusters and to capture the large-scale distribution of galaxies and clusters, while having sufficient resolution to spatially resolve the structure of galaxies.

Fig. 7 provides a visual impression of the effect of AGN jet activity on the gas around a star-forming galaxy of stellar mass $10^{10.4} M_\odot$ in the hybrid AGN feedback simulation L100m6h at $z = 0$. The main image and the first zoom-in show the gas temperature, while the final zoom-in shows stellar light in *HST* bands, accounting for dust attenuation. We can clearly see the shock fronts resulting from two different episodes of AGN activity. We have verified, using jet tracers, that these bubbles are driven mainly by jet rather than thermal AGN feedback. Below the image centre, we can see a lower mass galaxy ($M_* \approx 10^{9.9} M_\odot$) whose CGM has also been shaped by past jet activity. We refrain from comparing with the same region in the thermal AGN feedback simulation because the stochastic nature of AGN feedback episodes implies that such comparisons are of limited use for individual galaxies. However, we note that the isotropic thermal AGN feedback used in the fiducial simulations also frequently results in bipolar out-

flows because galactic winds tend to escape perpendicular to the galaxy disc along the path of least resistance, though the outflows tend to be less extended and less collimated than for jet AGN feedback (see fig. 12 of F. Huško et al. 2026).

More images, also for L400m7, as well as videos and interactive visualizations, can be found on the COLIBRE website.²³

6 PHASE DIAGRAMS AND STELLAR BIRTH PROPERTIES

In this section, we provide some physical intuition for the behaviour, outcome, and limitations of the simulations. In Section 6.1, we discuss the distribution and properties of gas at $z = 0$ in temperature/pressure–density space, limited to the L200m6 simulation for brevity. In Section 6.2, we explore the physical properties of the gas from which stars are born and in which CCSNe explode, and their dependence on numerical resolution.

6.1 Phase diagrams

Fig. 8 shows the distribution of gas particles in temperature–density (left panel) and pressure–density (right panel) space. For densities $n_H \gtrsim 10^{-1} \text{ cm}^{-3}$, the left panel shows concentrations of gas corresponding to the three main phases of the ISM, i.e. cold ($T \lesssim 10^2 \text{ K}$), warm ($T \sim 10^4 \text{ K}$), and hot ($T \gg 10^4 \text{ K}$) gas. The right panel shows that these different phases of the ISM can have the same pressures. Note that in the previous generation of simulations of representative volumes, such as EAGLE, there is no cold phase and the temperature of the warm phase becomes

²³<https://colibre-simulations.org/>

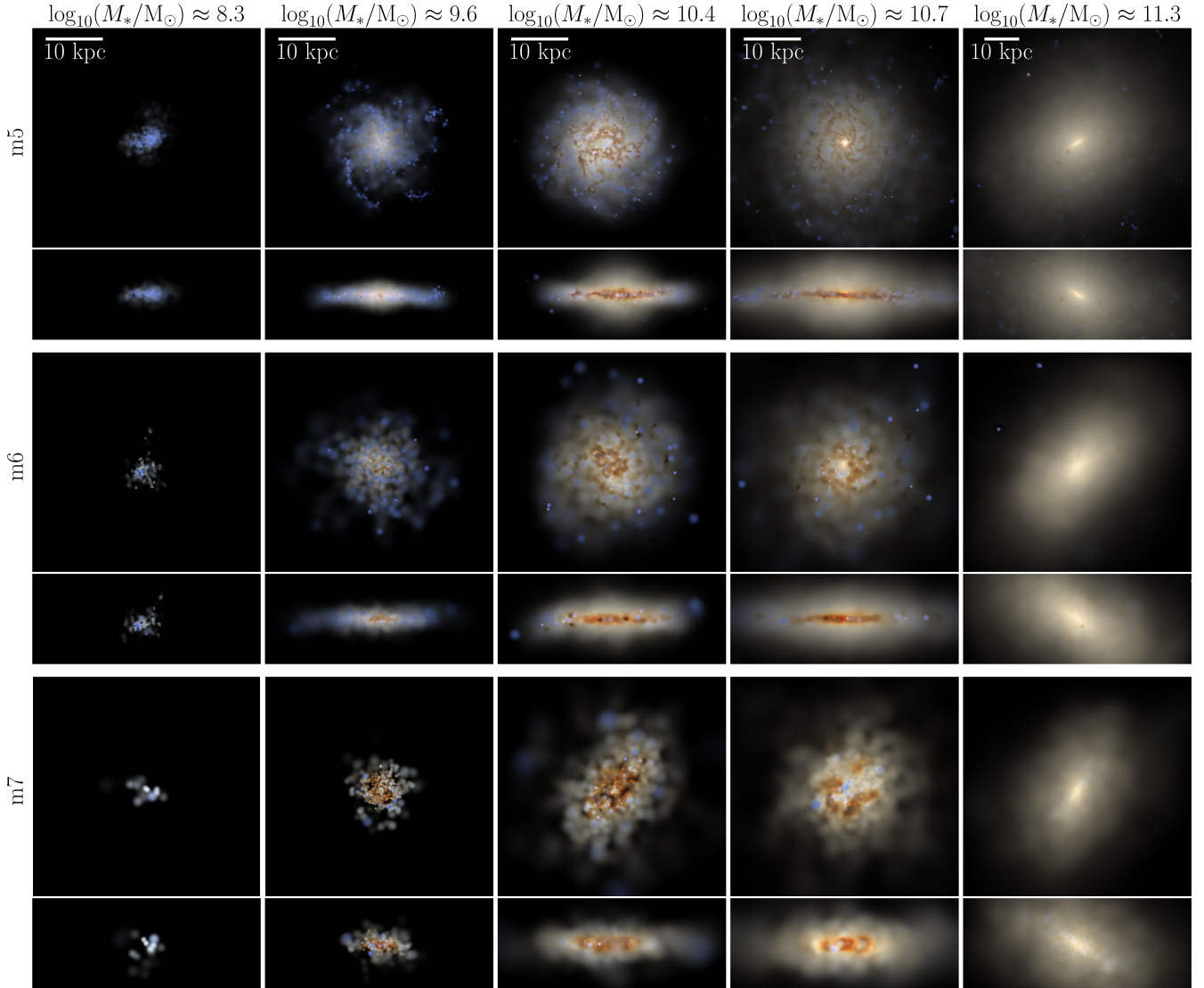


Figure 5. Comparison of galaxy morphologies across different COLIBRE resolutions. Each column shows face-on and edge-on images of the same galaxy in the L025m5 (top), L025m6 (middle), and L025m7 (bottom) simulations at $z = 0$. The different columns show different galaxies with stellar mass increasing from left to right. The stellar masses for m6 resolution are indicated above each column. These can differ by ± 0.1 dex for the other resolutions. The images show stellar light in *HST* colours and account for attenuation by dust. They are 50 kpc across except for the last column, which is 75 kpc across. Higher resolution simulations reveal more detail and predict thinner discs. Except for the lowest mass galaxy, the galaxy shape can be inferred at all resolutions.

unrealistically high because a (density-dependent) entropy floor is imposed on the ISM.

Fig. 9 shows six different phase diagrams for the L200m6 simulation at $z = 0$. The six panels show the mean atomic hydrogen mass fraction (top left), the mean molecular hydrogen mass fraction (top right), the mean metallicity (middle left), the fraction of particles in which thermal CCSN or AGN feedback was injected in the last 100 Myr (middle right), the mean dust-to-metal mass ratio (bottom left), and the mean mass ratio of small-to-large dust grains (bottom right).

The top panels of Fig. 9 show that the cold interstellar phase is neutral (i.e. H I or H₂), with the denser parts ($n_{\text{H}} \gtrsim 10^2 \text{ cm}^{-3}$) being substantially molecular. Comparing with the bottom-left panel, we see that the temperature–density relation of the bulk of the cold ISM, i.e. track I in the left panel of Fig. 8, corresponds to

features traced by both the molecular hydrogen fraction and the dust-to-metal ratio. This is expected because molecules primarily form on dust grains, and both dust and molecules play a key role in regulating the thermal balance of the ISM. The bottom-right panel shows that the higher gas densities favour larger dust grain sizes, which is due to accretion of metals on to dust grains in the cold and dense ISM.

Region E in the left panel of Fig. 8, which is the track with the high concentration of particles that emerges above $n_{\text{H}} \sim 10^{-3} \text{ cm}^{-3}$ and approximately follows the $T = 10^4 \text{ K}$ isotherm, traces the thermal equilibrium temperature of low-metallicity gas (see the middle-left panel of Fig. 9). Above $n_{\text{H}} \sim 10^{-0.5} \text{ cm}^{-3}$, the track rapidly transfers to another track at $T \sim 10^2 \text{ K}$, after which its temperature declines gradually with increasing density (region I). The middle-left panel of Fig. 9 also shows that the

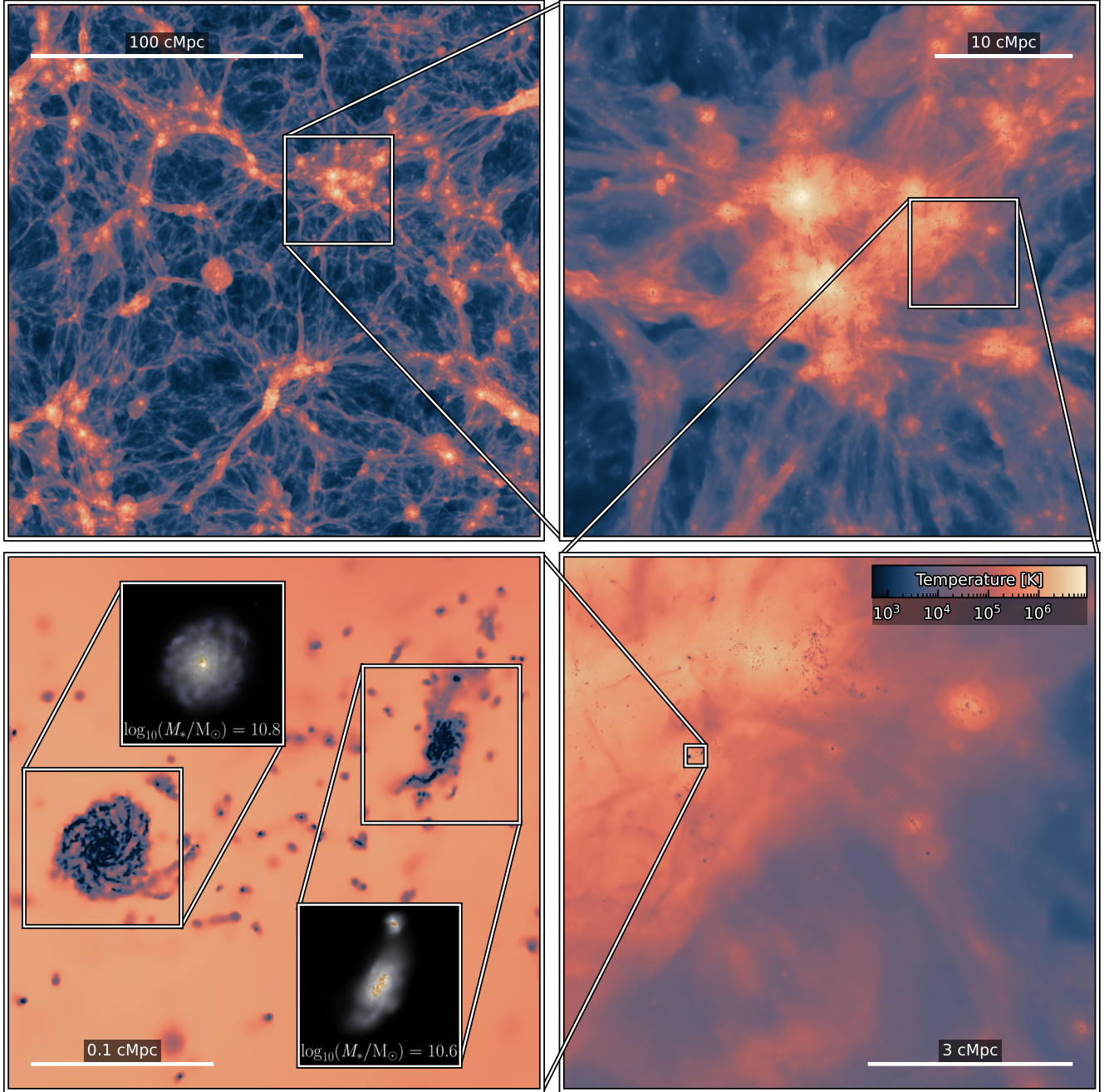


Figure 6. Visual impression of the dynamic range in the COLIBRE simulation L200m6 at $z = 0.1$. The top-left panel shows a 25 cMpc thick slice of the temperature (mass-weighted mean of $\log_{10} T$) of the gas. The three other main panels show consecutive zooms into the outskirts of a galaxy cluster, keeping the same 25 cMpc depth. The two smaller images in the bottom-left panel show the stellar light in *Euclid* colours and account for attenuation by dust. COLIBRE simultaneously models the large-scale cosmic web, the multiphase gas in the intracluster and interstellar media, and the internal structure of galaxies.

transition to the cold ISM occurs at higher densities in gas with lower metallicity.

At densities $10^{-3} \lesssim n_{\text{H}}/\text{cm}^{-3} \lesssim 10^2$, there is a small fraction of gas at temperatures below the track corresponding to the thermal equilibrium curve of the cold ISM (region F in the left panel of Fig. 8). From Fig. 9, we see that this gas has a much lower molecular fraction and dust-to-metal ratio than the main cold ISM track. This is likely gas that overshot the equilibrium temperature as it

cooled rapidly from high ISM temperatures ($T \gg 10^4$ K). This is a non-equilibrium effect resulting from the temporary availability of residual free electrons due to the recombination lag captured by our time-dependent treatment of hydrogen and helium (see A. J. Richings et al. 2014a, for a demonstration of this effect). The middle-right panel shows that a large fraction of this gas recently received thermal feedback energy through direct injection. This recently cooled gas has not yet had the time to (re)grow dust grains and (re)form molecules that, if they were present

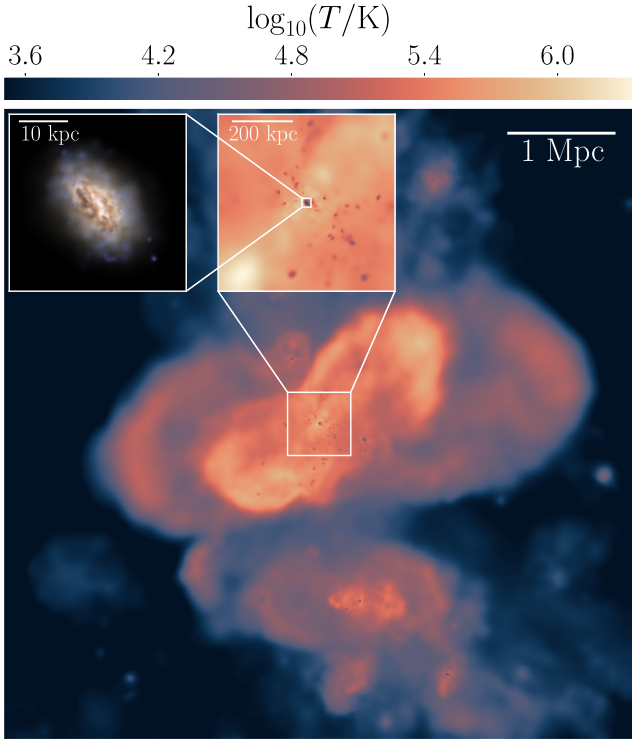


Figure 7. Visual impression of the effect of AGN jets on the gas around two star-forming galaxies. The main image shows the gas temperature (mass-weighted mean of $\log_{10} T$) in a 6 Mpc cubic region in the hybrid AGN feedback L100m6h simulation at $z = 0$. The image is centred on a halo of mass $M_{200c} \approx 10^{11.9} M_{\odot}$ with a central galaxy of stellar mass $10^{10.4} M_{\odot}$, SFR $1.6 M_{\odot} \text{ yr}^{-1}$, and nuclear BH mass $10^{7.0} M_{\odot}$. The other panels show consecutive zooms of the temperature in a 0.6 Mpc cubic region and of stellar light in *HST* colours accounting for attenuation by dust in a 30 kpc cubic region. The galaxy below the one in the main panel has a smaller mass ($M_{200c} \approx 10^{11.4} M_{\odot}$, $M_{*} \approx 10^{9.9} M_{\odot}$, SFR $\approx 0.5 M_{\odot} \text{ yr}^{-1}$, and $m_{\text{BH}} \approx 10^{6.0} M_{\odot}$). In 3D, the two galaxies are separated by 2.8 Mpc. Shock fronts driven by AGN jet activity are clearly visible in the main panel, while the jets themselves are also visible in the first zoom-in panel.

in the past, would have been rapidly destroyed when the gas was hot.

The nearly isothermal track just above $T = 10^4 \text{ K}$, region H in the left panel of Fig. 8 and visible in all temperature–density plots, which emerges above $n_{\text{H}} \sim 10^{-2} \text{ cm}^{-3}$ and extends to very high densities, consists of ionized gas, i.e. H II regions, as can be inferred from the two top panels of Fig. 9. The right panel of Fig. 8 shows that the pressures in H II regions extend to values higher than those of the cold ISM, suggesting that these regions may expand.

The hot phase of the ISM, region G in the left panel of Fig. 8, is produced by stellar and AGN feedback. This can be seen from the middle-right panel of Fig. 9, though one should keep in mind that gas that was recently shock-heated by winds, but in which thermal feedback energy was not directly injected, is not included in the fraction used for the colour coding. The horizontal features at $T \approx 10^{6.5}$ and $\approx 10^{6.75} \text{ K}$, which are most clearly visible in the left panel of Fig. 8, reflect the minimum heating temperatures for AGN and SN feedback, respectively, while the features at $T \approx 10^{10}$ and $\approx 10^8 \text{ K}$ reflect the corresponding maximum heating temperatures. There are additional but weaker features at integer multiples of the maximum SN heating temperature caused by gas particles being hit by multiple SN events. Gas that is (shock-

heated by feedback will either cool radiatively, in which case it will rapidly join the warm or even the cold ISM, or, if the temperature is sufficiently high, cool adiabatically to lower densities and temperatures until it reaches pressure equilibrium with the ambient gas or until radiative cooling takes over.

The bottom-left panel of Fig. 9 shows that the hot ISM can be dusty, though only for dense gas above the minimum feedback heating temperatures, which is an artefact caused by grain destruction lagging one time-step behind the feedback: because the injection of energy by feedback takes place at the end of the time-step, there has not yet been an opportunity for the dust to be destroyed. The dust in these particles will generally be destroyed by sputtering (or by hand for directly heated particles, see Section 3.2) in the next time-step. The fact that most of the particles that are both hot ($T \gtrsim 10^7 \text{ K}$) and dense ($n_{\text{H}} \gtrsim 1 \text{ cm}^{-3}$) are dusty indicates that these particles have only just been heated, as can also be inferred from the middle-right panel. This implies that this hot and dense phase is very short-lived, as expected given its very high pressures (see the right panel of Fig. 8).

The left panel of Fig. 8 shows that the highest concentration of gas mass (region A) follows a power-law temperature–density relation at low densities ($n_{\text{H}} < 10^{-5} \text{ cm}^{-3}$) and low temperatures ($T \lesssim 10^4 \text{ K}$). This is the diffuse IGM, which has low overdensities (for reference, the cosmic mean density $\bar{n}_{\text{H}} \approx 1.9 \times 10^{-7} \text{ cm}^{-3}$) and which follows a track corresponding to the quasi-equilibrium between photoheating by the metagalactic ionizing background radiation and adiabatic cooling due to the *Hubble* expansion (L. Hui & N. Y. Gnedin 1997). Fig. 9 shows that this gas is highly ionized (top row), nearly free of both metals (middle left) and dust (bottom left), and has not recently been injected with thermal feedback energy (middle right).

The concentration of gas in Fig. 8 at very high temperatures, $T \gtrsim 10^7 \text{ K}$, comprises the intragroup and intracluster media, where adiabatic compression and shock heating result in a positive correlation between temperature and density (region C). A large amount of gas also resides at intermediate temperatures ($10^5 \lesssim T/\text{K} \lesssim 10^7$) and densities ($10^{-6} \lesssim n_{\text{H}}/\text{cm}^{-3} \lesssim 10^{-4}$), i.e. region B, which corresponds to the shock-heated warm-hot intergalactic medium and CGM. The shock-heated gas is highly ionized, largely dust-free,²⁴ and substantially enriched with metals. At densities $10^{-4} \lesssim n_{\text{H}}/\text{cm}^{-3} \lesssim 10^{-1}$ a significant fraction of the gas is at temperatures $\sim 10^4 \text{ K}$, which corresponds to the cooler parts of the CGM, where photoheating is balanced by radiative cooling (region D). Fig. 9 shows that much of this gas has a very low metallicity and a high atomic hydrogen fraction.

6.2 The properties of the gas from which the stars are born and in which CCSNe explode

In this section, we explore some of the physical properties of the gas particles at the time they were converted into stellar particles. We combine the birth properties of all stars, regardless of the redshift at which they formed and of the mass of the galaxies in which they currently reside. We thus consider the (ZAMS) stellar mass weighted mean properties of star-forming gas averaged over all of space and time.

²⁴The preference for large dust grain sizes in the nearly dust-free hot gas (bottom-right panel of Fig. 9) reflects the stellar yields of small and large dust grains, since grain growth is inefficient in this hot gas in which the dust is quickly destroyed by thermal sputtering.

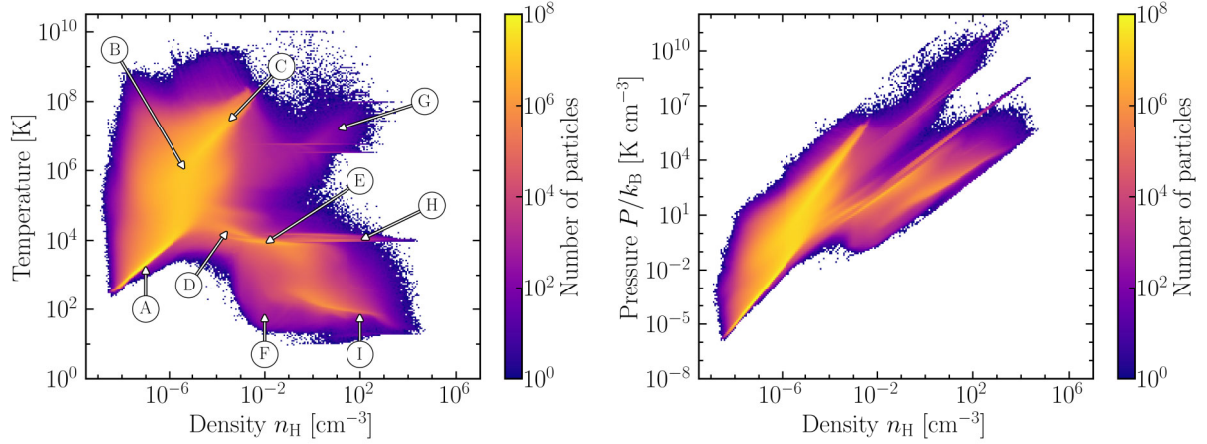


Figure 8. Distribution of gas particles in the L200m6 simulation at $z = 0$ in the temperature–density (left panel) and the thermal pressure–density (right panel) planes. The colour coding shows the fraction of the gas particles in the pixel on a logarithmic scale. The labels in the left panel are used to facilitate identification of particular regions discussed in the text. The gas in COLIBRE spans a wide range of physical conditions, with multiple gas phases overlapping in pressure.

Before showing results, we caution that ‘star-forming gas’ is a simulation concept. In COLIBRE, gas becomes star-forming if it is gravitationally unstable at the resolution limit (see Section 3.3). As discussed in F. S. J. Nobels et al. (2024), we do not expect stellar birth properties to converge with increasing resolution because the star formation criterion is applied at the resolution limit as opposed to at a fixed mass scale. If the resolution is higher, then the mass scale at which instability is tested is smaller and higher densities, lower temperatures and/or smaller turbulent velocity dispersions are required to form stars. In higher resolution simulations, there is generally less gas that is star-forming. However, because the star-forming gas has higher densities, it has a smaller gas consumption time-scale, which means that observables such as the total SFR or the SFR at a fixed resolved physical scale can still be converged. Checking for convergence for observables is sensible, but checking for convergence of the amount or physical properties of star-forming gas is not. This lack of convergence of the stellar birth properties is intentional because we wish to take advantage of the ability to follow the collapse of gas clouds to higher densities in higher resolution simulations.

We note that ‘star-forming gas’ is also an ambiguous concept when analysing observations. If we observe a galaxy at a resolution similar to the size of the galaxy, we may find the atomic gas content to correlate well with star formation, but this association may break down on molecular cloud scales. Similarly, observations that resolve molecular cloud cores will find tighter correlations between star formation and the densest molecular gas than with the total amount of molecular gas (e.g. R. C. Kennicutt & N. J. Evans 2012). Continuing this trend, observations resolving protostars would indicate that only gas with near-stellar densities is ‘star-forming’.

Because of the explicit resolution dependence, the properties of star-forming gas and of the environmental densities of CCSNe can also not be compared directly to those in much higher resolution simulations of individual $z = 0$ low-mass dwarf galaxies (e.g. C.-Y. Hu et al. 2017; M. C. Smith et al. 2021; T. A. Gutcke et al. 2021; J. M. Hislop et al. 2022; U. P. Steinwandel et al. 2023). Another reason is that we combine all stars that formed in any

galaxy during the history of the Universe. At higher redshift and in the higher mass galaxies that dominate the cosmic star formation history (for a decomposition, see fig. 6 of E. Chaikin et al. 2025b), stars will likely form (and CCSNe explode) in gas with higher densities.

The top and middle panels of Fig. 10 show the probability density distributions (PDFs) of, respectively, the density and temperature of the gas from which the stellar particles formed. The bottom panel shows the PDF of the ratio of turbulent and thermal velocity dispersions of the gas particles that were converted into stellar particles, where the former is the local 3D velocity dispersion between particles and the latter is estimated from the gas temperature. The PDFs include all stellar particles in the simulation, i.e. stars formed at any redshift. The red, orange, and blue curves correspond to simulations L025m5, L025m6, and L025m7, respectively.

The temperature PDF shows the most pronounced features. There are peaks at $T \approx 20$, $\sim 10^2$, and $\sim 10^4$ K. The first two peaks correspond to the cold interstellar gas phase, while the last corresponds to the warm phase. The spike at ≈ 20 K is due to the lower limit that we impose on the internal energy (corresponding to 10 K for atomic gas, but the gas is molecular at these temperatures). Most stars form in the cold phase ($T \lesssim 10^2$ K), but for the lower resolutions the contributions from the warm phase are significant. The density PDF is very wide, with the low-density tail corresponding to star formation in the warm phase. The bottom panel shows that turbulence is typically supersonic, with the secondary peak at subsonic/mildly supersonic values corresponding to birth temperatures $T \sim 10^4$ K. In Appendix C, we show that the physical conditions correlate strongly with galaxy stellar mass, where stars in more massive galaxies tend to have formed from gas with higher densities, lower temperatures, and more supersonic turbulence.

As resolution increases, the birth densities increase, the temperatures decrease, and the turbulence becomes less supersonic. This is expected because gas is only allowed to form stars if the kernel mass is gravitationally unstable, i.e. if the kernel mass exceeds the local Jeans mass. A higher resolution corresponds to a smaller kernel mass, which means that higher densities and

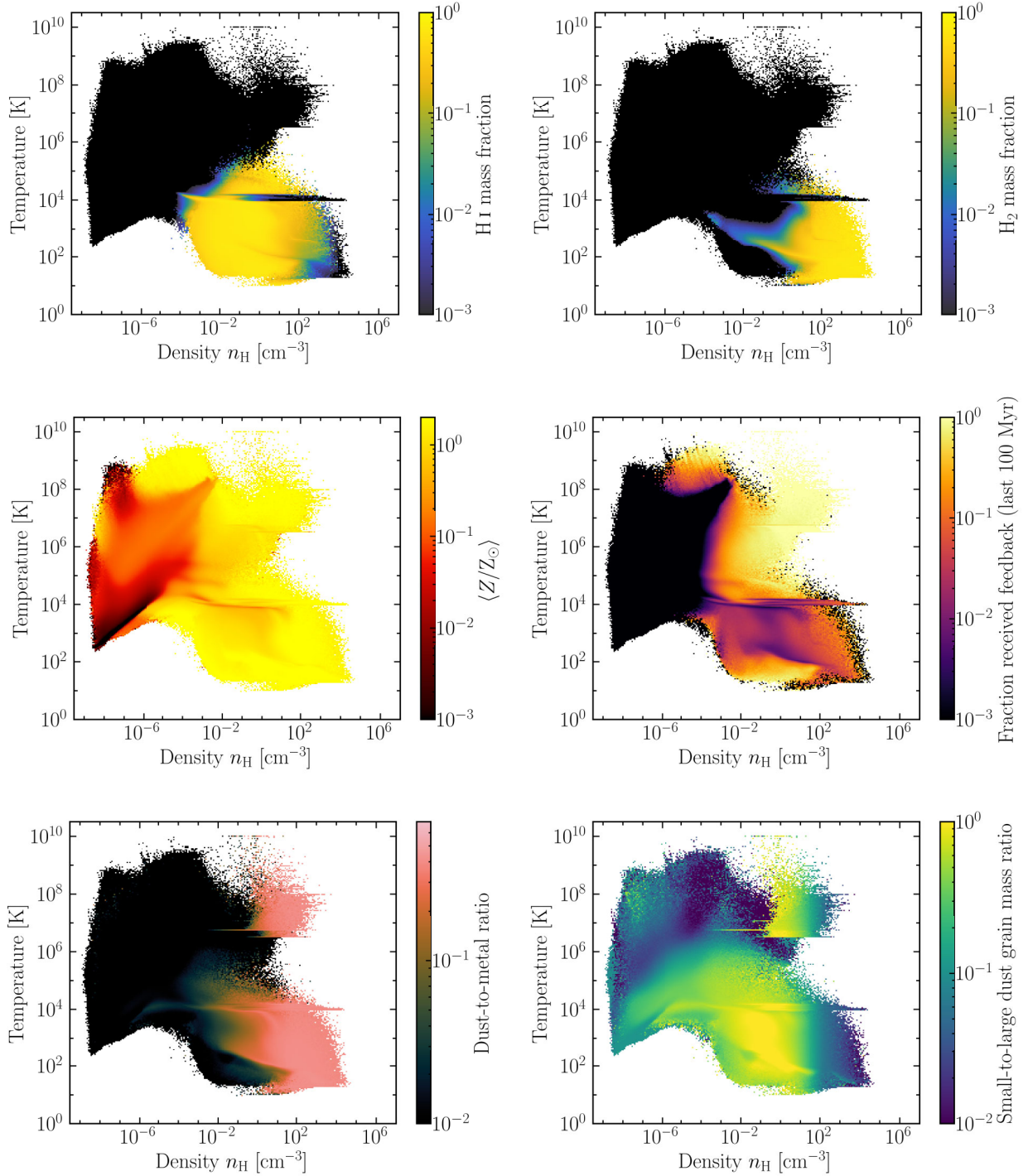


Figure 9. Various gas properties in temperature–density space for the L200m6 simulation at $z = 0$. The colour of each pixel encodes, on a logarithmic scale, the mean atomic hydrogen mass fraction (top left), the mean molecular hydrogen mass fraction (top right), the mean metallicity $Z/Z_\odot$ (using the solar abundance $Z_\odot = 0.0134$; M. Asplund et al. 2009) (middle left), the fraction of particles in which thermal CCSN or AGN feedback was injected during the last 100 Myr (middle right), the mean dust-to-metal mass ratio (bottom left), and the mean mass ratio of small-to-large dust grains (bottom right). All means are computed as the ratio of the sum of the quantity in the numerator and the sum of the quantity in the denominator, e.g. $\sum_i m_{\text{HI},i} / \sum_i m_{\text{H},i}$ for the top-left panel, where the sum is over all particles i with temperatures and densities that fall into the pixel. Note that the colour scales do not span the full ranges of values.

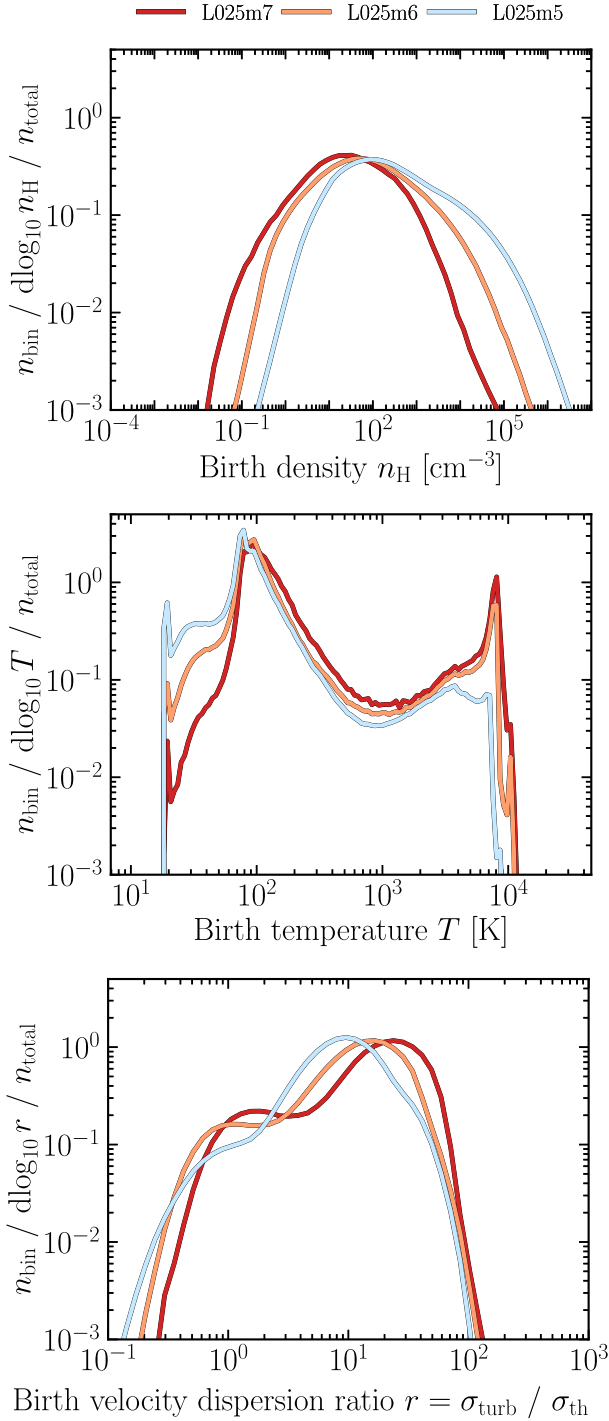


Figure 10. PDF of the density (top panel), temperature (middle panel), and the ratio ($r = \sigma_{\text{turb}} / \sigma_{\text{th}}$) of the turbulent-to-thermal velocity dispersion (bottom panel) of the gas at the time it was converted into stellar particles (at any $z \geq 0$) in simulations using the same box size (L025), but different resolutions (different colours). The turbulent dispersion, σ_{turb} , is the local 3D velocity dispersion between gas particles. The thermal dispersion is estimated from the birth temperature using $\sigma_{\text{th}} = 13.8 \text{ km s}^{-1} (T/10^4 \text{ K})^{1/2}$, which assumes a mean molecular weight of 1.3 (which is appropriate for atomic gas but leads to an overestimate of σ_{th} in molecular gas). The spike at $T \approx 20 \text{ K}$ reflects the minimum allowed internal energy. The temperature PDF is multimodal, with most stars forming in the cold ($T \ll 10^4 \text{ K}$) interstellar gas phase. Turbulence in star-forming gas is typically supersonic ($r > 1$). As the resolution increases, star formation shifts to higher gas densities and lower temperatures.

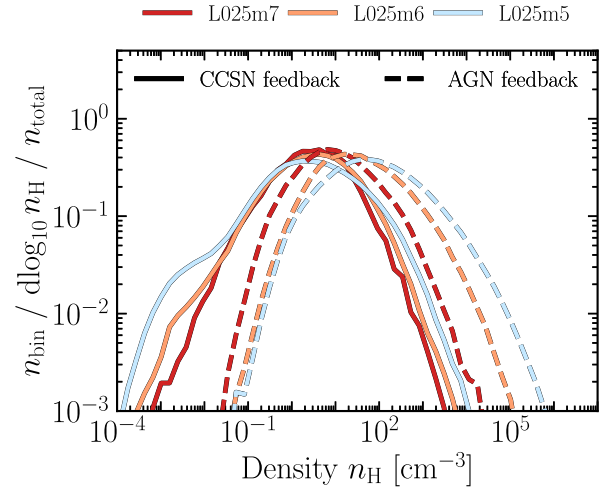


Figure 11. PDF of the density of the gas in which CCSN (solid lines) and AGN (dashed lines) thermal feedback energy is injected in simulations using the same box size (L025) but different resolutions (different colours). This is based on the last, if any, CCSN and AGN feedback events that resulted in the direct injection of thermal energy in the gas particle, which can be at any redshift $z \geq 0$. While the typical feedback gas densities for CCSNe are nearly converged with the resolution, the density of the gas in which AGN feedback is injected increases with the resolution and tends to be much higher than for CCSNe. The density of the gas in which CCSNe explode is typically much lower than the density of the gas from which the stars form, particularly for higher resolutions (compare to the top panel of Fig. 10; the median birth densities are factors of ~ 10 – 10^2 higher).

lower total velocity dispersions are required for star formation (see equation 6). A higher density typically corresponds to a lower temperature (see Fig. 9). Because the turbulence in the low-temperature gas is typically supersonic, stars tend to form in regions where the turbulent velocity is relatively low.

We note that in simulations such as EAGLE, which have a baryonic resolution similar to COLIBRE m5 and m6, but impose an equation of state on the ISM, stellar birth properties are similar to those in simulations using the COLIBRE model, but with a resolution much lower than m7. In such simulations, birth temperatures are $\sim 10^4 \text{ K}$, median birth densities are orders of magnitude lower than for similar resolution COLIBRE models, and turbulence tends to be at most mildly supersonic. Direct simulation of the cold ISM therefore results in dramatic changes to the physical conditions of the gas from which stars are born for a fixed resolution.

For all COLIBRE resolutions, most of the stars form in the cold ISM. Although not shown, we note that a larger fraction of stars form in the warm ISM if the metallicity is lower because in that case the transition from the warm to the cold ISM occurs at higher densities. Lower metallicities tend to correspond to higher formation redshifts and lower galaxy masses.

Fig. 11 shows the PDFs of the density of the gas in which thermal energy was injected by CCSNe (solid lines) and AGN feedback (dashed lines). Comparing the solid lines with the PDF of the stellar birth densities shown in the top panel of Fig. 10, we note two things. First, compared to the stellar birth densities, the gas densities in which CCSNe explode are typically much lower than the stellar birth densities. Second, the density of the gas in which CCSNe explode converges much better with the resolution. The difference between the median birth and CCSN

feedback densities increases from 1.0 dex at m7 to 1.9 dex at m5 resolution. Because dense cold clouds contain a large fraction of the ISM mass, the density of the intercloud medium is relatively low. While stars form in dense clouds, CCSNe explode mostly in the intercloud medium because stellar winds, photoheating, and earlier CCSNe destroy birth clouds and because the relative motion of stars and gas during the pre-CCSN phase allows many stars to move out of their birth clouds. The relatively low density of the gas in which the feedback events occur helps to reduce numerical overcooling and hence facilitates better convergence. In simulations with a multiphase ISM, such as COLIBRE, we therefore expect numerical overcooling to be reduced, and we can achieve better convergence for observables, compared to simulations that prevent the formation of a cold interstellar phase, such as EAGLE.

The density of the gas in which AGN feedback energy is injected tends to be much higher than that for CCSN feedback, and the difference increases with numerical resolution. This is expected because simulations with higher resolution can sample the gas closer to the BH, which tends to have higher densities, and the AGN feedback energy is given to the gas particle closest to the BH. Numerical convergence is therefore harder to achieve for AGN feedback than for CCSN feedback, although the higher maximum heating temperatures used for AGN feedback help suppress numerical overcooling.

7 COMPARISON WITH OBSERVATIONS

In this section, we compare the fiducial COLIBRE models, i.e. the simulations that use thermally driven AGN feedback, with observations of a variety of fundamental galaxy properties. We will separately consider the observables used to calibrate the subgrid models for CCSN and AGN feedback (Section 7.1) and other observables (Section 7.2). Apart from the cosmic star formation history, all the observations with which we compare are at $z \approx 0$. Comparisons with higher redshift observations and additional observables of galaxies, including spatially resolved properties, and the IGM will be presented in future work. The simulations using the hybrid AGN feedback model are compared with the data (and the fiducial model) in Section 8. Where necessary, observational data have been converted to the G. Chabrier (2003) IMF and to the cosmology assumed in COLIBRE.

As discussed in Section 2.3, galaxy properties are measured using SOAP from gravitationally bound particles within a spherical aperture of radius 50 pkpc centred on the most bound particle in the HBT-HERONS subhalo. The choice of aperture size was motivated by the results of A. de Graaff et al. (2022), but as discussed by E. Chaikin et al. (2025b), using a larger aperture only has a non-negligible effect for very massive galaxies, $M_* \gg 10^{11} M_\odot$, where it gives higher masses. Because of the somewhat arbitrary and analysis-specific distinction between the brightest cluster galaxy (BCG) and the intracluster light, a dedicated analysis is required for a quantitative comparison of such massive galaxies to observations.

Nearly all plots that we show in this section have galaxy stellar mass along the x -axis. Observational measurements of stellar masses include both random and systematic errors. Stellar masses are typically measured by fitting stellar population synthesis models to the spatially integrated spectral energy distribution (SED). Even if the IMF is fixed, the results depend on the assumed star formation history, chemical composition and its evolution, dust obscuration, stellar rotation, stellar multiplicity,

and the model/library for stellar spectra. To account for the Edington bias resulting from the random component of these and other uncertainties, i.e. the fact that, for declining mass functions, random errors in the stellar mass cause more objects to scatter up to higher mass bins than to scatter down to lower mass bins, we add random errors to the predicted stellar masses. Based on a review of the literature, E. Chaikin et al. (2025b) adopt a lognormal distribution with mean zero and standard deviation of $\log_{10} M_*$ of

$$\sigma_{\text{random}}(z) = \min(0.1 + 0.1 z, 0.3) \text{ dex.} \quad (43)$$

In Appendix D, we show that at $z = 0$ this amount of scatter results in only a slight smoothing of trends with mass, but that larger errors can have a strong impact for $M_* \gtrsim 10^{11} M_\odot$. Even though they are likely significant, we do not attempt to simulate the effects of random errors in galaxy properties other than mass because such errors are less well understood than errors in stellar mass and because they do not affect the binning for the plots presented here.

We expect the uncertainties to result not only in random errors but also in systematic errors. Accounting for systematic errors, which we do not do, would shift the entire curves along the mass axis and would change the shape of the curves if the systematic error depends on mass. E. Chaikin et al. (2025b) estimate the systematic error on the $z = 0$ stellar mass for a fixed IMF to be ≈ 0.15 dex in the mass range used for calibration.

We focus on the largest simulations currently available at $z = 0$ for each of the three resolutions: L400m7, L200m6, and L025m5, for which we will use red, orange, and blue line colours, respectively. Comparing these simulations tests not just for convergence with resolution, but also with volume. The L400m7 and L200m6 volumes are sufficiently large that the differences in the median trends with mass due to box size are limited to the most massive objects. However, the small volume of the L025m5 can play a significant role, mainly because the satellite fraction and the host masses of satellites are lower than for the other simulations due to the limited range of scales on which primordial fluctuations can be sampled in a volume of this size. Although we will not show them here, L025m6 and L025m7, as well as L200m7, are also available to facilitate comparing different resolutions at fixed box size.

In all figures with stellar mass on the x -axis, we will switch from a solid to a dashed line style where the mass becomes smaller than 100 times the mean initial baryonic particle mass²⁵ and where there are fewer than 10 objects per stellar mass bin of width 0.2 dex. We stress that the 100 stellar particle limit is just for reference and does not necessarily correspond to the resolution limit, which can and should be estimated separately for each question of interest by comparing the different resolutions, preferably using simulations of the same volume. However, it is important to note that the physical processes captured in hydrodynamical simulations using subgrid models, such as COLIBRE, effectively change with resolution. Hence, even if they are recalibrated to match a limited set of observations, some differences between different resolution simulations may only be indirectly caused by the numerical resolution. Instead, the differences may

²⁵ Because of stellar mass loss, the number of stellar particles corresponding to this limit depends somewhat on the stellar age and metallicity. For m7 resolution, the change of line style occurs at ≈ 140 particles, while for m6 and m5, it corresponds to ≈ 160 particles.

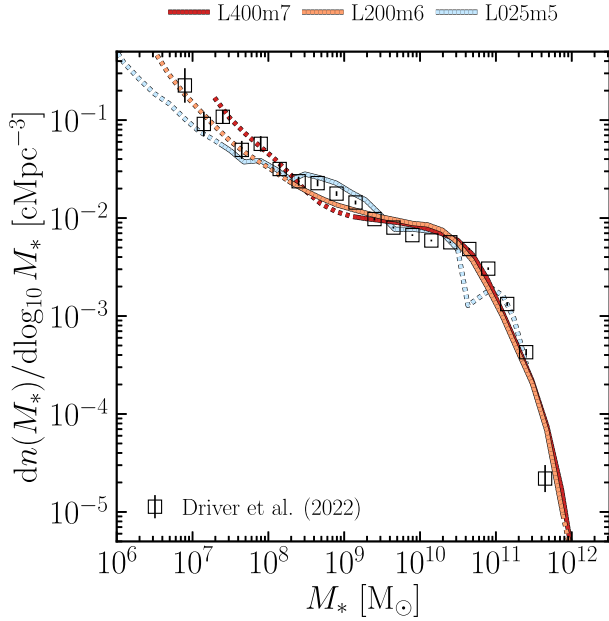


Figure 12. The $z = 0$ galaxy SMF in the L400m7 (red), L200m6 (orange), and L025m5 (blue) simulations. The line style switches from solid to dotted where the mass becomes smaller than 100 times the mean initial baryonic particle mass and at high mass where there are fewer than 10 objects per 0.2 dex mass bin. The data points with error bars show the observed mass function from GAMA DR4 (S. P. Driver et al. 2022). The different resolution simulations are in very good agreement for galaxy masses corresponding to at least a few tens of stellar particles. The agreement between the simulations and the data are excellent.

largely reflect the implicit freedom in the subgrid physics that remains after calibration.

7.1 Observations used for calibration of CCSN and AGN feedback

In this section, we will compare the fiducial COLIBRE simulations with the observations used to calibrate the CCSN and AGN feedback, namely the $z = 0$ galaxy SMF (Section 7.1.1) and the $z \approx 0$ galaxy size–SMR (Section 7.1.2), as well as the $z = 0$ relation between BH and galaxy stellar mass (Section 7.1.3).

As summarized in Section 3.10 and discussed in detail in E. Chaikin et al. (2025a), the calibration was carried out by first using machine learning to emulate the effect of varying four subgrid parameters, three for CCSN feedback plus the BH seed mass, at m7 resolution using volumes of 50 cMpc on a side (L050m7). The emulator predictions were then used to find the parameter values that provide the best fit to the observed SMF and SMR for the galaxy stellar mass range $9.0 < \log_{10} M_*/M_\odot < 11.3$. We also consider the high-mass end of the $z = 0$ relation between BH and galaxy stellar mass to belong to the calibration target observations because it motivated the choice of the AGN feedback efficiency, ϵ_f . While the parameter values were kept fixed when increasing the box size, they were slightly adjusted when increasing the resolution, guided by smaller volume test simulations (L025m6 and L012.5m5).

7.1.1 The galaxy stellar mass function

Fig. 12 compares the COLIBRE SMFs to the observations of S. P. Driver et al. (2022) from the GAMA DR4 survey. We applied the corrections suggested by the authors for evolution from the actual redshift range of $z < 0.1$ to 0, based on the evolution measured by the Deep Extragalactic Visible Legacy Survey (DEVILS) (J. E. Thorne et al. 2021), and for cosmic variance, obtained by normalizing to the larger-area SDSS. Together, these corrections increase the observed number densities by 0.0807 dex.

The agreement with the data is excellent. Taking into account the estimated systematic error on the observed stellar masses of 0.15 dex, the simulations are fully consistent with the observations from the highest masses that are sampled with at least 10 objects per mass bin down to masses corresponding to 100 particles (the solid parts of the curves) and for m7 and m6 even down to < 10 particles. When combined, the three simulations reproduce the observations over 5 orders of magnitude in stellar mass. Notably, for m6 and m5, the agreement with the data extends to far lower masses than the lowest mass used in the calibration ($10^9 M_\odot$). The agreement at each resolution with the same data implies excellent convergence with the numerical resolution and volume of the simulations.

In Appendix D, we show that the predictions for $M_* > 10^{11} M_\odot$ become somewhat sensitive to the assumed random errors on the stellar masses (because of Eddington bias). E. Chaikin et al. (2025b) show that the same holds for the chosen aperture. Due to the steepness of the mass function, the expected systematic error on the observed stellar masses, 0.15 dex for a fixed IMF, becomes also important. This suggests that comparisons with the high-mass data would particularly benefit from applying observational techniques to virtual observations, which is, however, beyond the scope of this work. Finally, we note that E. Chaikin et al. (2025b) find that the agreement with the observed SMF is not limited to the redshift used for calibration (i.e. $z = 0$), but extends up to the highest redshifts for which constraints are available, $z \lesssim 12$.

Fig. 13 shows the median stellar-to-halo mass ratio for central galaxies as a function of halo mass, which is closely related to the SMF. The simulations are compared with results inferred by P. Behroozi et al. (2019) by fitting the semi-empirical model UniverseMachine to a compilation of diverse observations. For consistency with P. Behroozi et al. (2019), we define the halo mass as the mass within the radius for which the mean internal density is equal to the critical overdensity provided by G. L. Bryan & M. L. Norman (1998), which is based on a fit to the simulations of V. R. Eke, S. Cole & C. S. Frenk (1996) and at $z = 0$ has the value ≈ 101.8 times the critical density of the universe. However, because UniverseMachine is built on DMO simulations, it does not account for the reduction in halo mass resulting from baryonic effects, which is expected to be small at cluster masses, but ≈ 0.1 dex for $M_{\text{halo}} \lesssim 10^{12} M_\odot$ (e.g. M. Schaller et al. 2015a). While P. Behroozi et al. (2019) used the peak halo mass, i.e. the maximum mass the halo has ever reached, we use the current halo mass, but for central galaxies, the difference is negligible for our purposes. To facilitate easy conversion between the stellar and halo masses of central galaxies, Fig. E1 in Appendix E shows the median stellar mass as a function of halo mass and vice versa. As can be clearly seen from Fig. 13, the agreement with the data inferred using the semi-empirical model is very good, as expected given that both COLIBRE and UniverseMachine agree with the observed SMF. The small difference at $M_{\text{halo}} \sim 10^{11.5} M_\odot$ is simi-

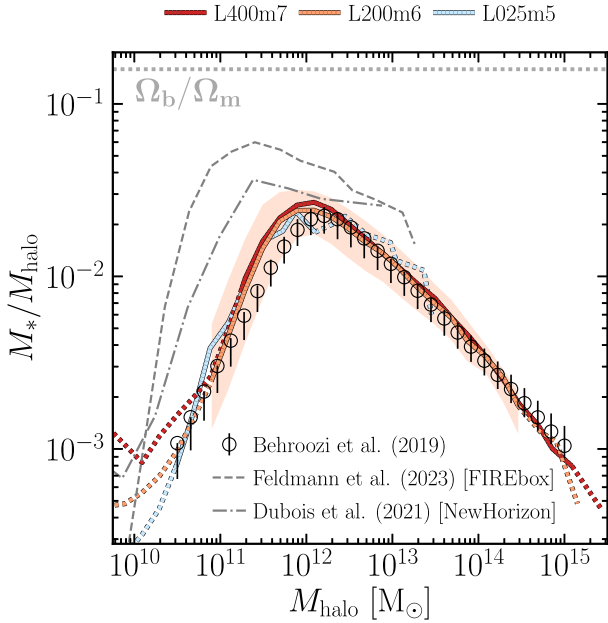


Figure 13. Median $z = 0$ stellar-to-halo mass ratio for central galaxies as a function of halo mass for the L400m7 (red), L200m6 (orange), and L025m5 (blue) simulations. The orange shaded region indicates the 16th–84th percentile range for L200m6. The line style switches from solid to dotted where the stellar mass becomes smaller than 100 times the mean initial baryonic particle mass and at high mass where there are fewer than 10 objects per 0.2 dex mass bin. The horizontal grey, dotted line indicates the universal baryon fraction. Data points with error bars show the best-fitting median and the 16th–84th percentile confidence level on the median inferred from observations using the semi-empirical model UniverseMachine (P. Behroozi et al. 2019). The thin black lines show the results from two other cosmological simulations that, like COLIBRE, allow cooling to cold ISM temperatures: FIREbox (R. Feldmann et al. 2023) and NewHorizon (Y. Dubois et al. 2021). Different resolution COLIBRE simulations are in excellent agreement with each other and with the relation inferred from observations using UniverseMachine, but there are large discrepancies with the other simulations.

lar in magnitude to the effect expected from UniverseMachine’s inability to capture baryonic effects on the halo mass.

In Fig. 13, we also compare to two cosmological simulations that have been run to low redshift and that, like COLIBRE, allow cooling to cold ISM temperatures.²⁶ NewHorizon (Y. Dubois et al. 2021) is a resimulation of a $\approx (16 \text{ cMpc})^3$ volume taken from the HORIZONAGN simulation (Y. Dubois et al. 2014) that was run to $z = 0.25$. Because NewHorizon does not simulate a volume with a mean density equal to the cosmic mean, we cannot directly compare its SMF to observations, but the zoom-in ICs should not affect the relation between the stellar and halo masses of central galaxies, which is expected to be insensitive to the large-scale environment (e.g. R. A. Crain et al. 2009). NewHorizon uses the

²⁶We do not compare to the Romulus25 simulation (M. Tremmel et al. 2017) because that simulation uses a much more basic model for radiative cooling that ignores molecules, dust grains and, for $T > 10^4 \text{ K}$, metal-line cooling. Fig. 3 of M. Tremmel et al. (2017), in which the simulated stellar masses were reduced by 40 per cent to correct for observational biases, shows that the stellar masses in Romulus25 are similar to those in NewHorizon for $M_{\text{halo}} \gtrsim 10^{12} M_{\odot}$, but closer to the semi-empirical data for lower halo masses.

adaptive mesh refinement code RAMSES (R. Teyssier 2002). The CDM particle mass is $1.2 \times 10^6 M_{\odot}$, which is close to our intermediate resolution m6. The smallest gas cell size is 34 pc, which is larger than the minimum distance between baryonic particles in COLIBRE, but much smaller than our maximum physical gravitational softening length. NewHorizon includes CCSN and AGN feedback, but not feedback from stellar winds, H II regions, and type Ia SNe. Radiative cooling rates assume ionization and chemical equilibrium and there is no model for dust grains. FIREbox (R. Feldmann et al. 2023) simulates a $(22.1 \text{ cMpc})^3$ volume with the meshless finite-mass GIZMO code (P. F. Hopkins 2015) down to $z = 0$. The CDM and gas particle masses are 3.3×10^5 and $6.3 \times 10^4 M_{\odot}$, respectively, which are comparable to our m5 resolution. Radiative cooling rates, ion and molecular fractions assume ionization and chemical equilibrium, and dust grain evolution is not modelled. Stellar feedback processes are included, but AGN feedback is not. Like COLIBRE, both NewHorizon and FIREbox use the current rather than the peak halo mass and define halo mass using the virial overdensity of G. L. Bryan & M. L. Norman (1998) that was also used by P. Behroozi et al. (2019). However, each of the four models shown in Fig. 13 (COLIBRE, UniverseMachine, NewHorizon, and FIREbox) uses a different structure finder and a somewhat different cosmology. The results for NewHorizon are for $z = 0.25$ instead of $z = 0$, but for COLIBRE, the relation hardly evolves from that redshift to the present (see E. Chaikin et al. 2025b).

While COLIBRE is in close agreement with the stellar mass–halo mass relation inferred from observations, there are large discrepancies with the other two simulations, which also do not agree with each other. NewHorizon and FIREbox both predict much higher galaxy masses than inferred from observations by P. Behroozi et al. (2019). For NewHorizon, the difference is small near the peak of galaxy formation efficiency at $M_{\text{halo}} \approx 1 \times 10^{12} M_{\odot}$, but increases to a factor of 2 at the highest halo mass sampled by that simulation, $M_{\text{halo}} \sim 10^{13} M_{\odot}$. The difference increases sharply toward lower masses, reaching almost an order of magnitude for $M_{\text{halo}} < 10^{11.5} M_{\odot}$. The difference with FIREbox is large at all masses. The factor of 2 difference for $M_{\text{halo}} \gtrsim 10^{12} M_{\odot}$ could be due to the lack of AGN feedback in FIREbox, but the difference increases rapidly toward lower masses, even exceeding a factor of 10 for $M_{\text{halo}} \sim 10^{11} M_{\odot}$.

Finally, we refer to Appendix F for a comparison to EAGLE and IllustrisTNG, which, like COLIBRE, span multiple resolutions, but, unlike COLIBRE, impose an effective equation of state onto the ISM. The TNG subgrid physics was calibrated for the resolution of the TNG100 simulation, which is similar to our m6 resolution. Unlike EAGLE and COLIBRE, for TNG, the subgrid parameters are held fixed when the resolution is changed, such as for the higher (m5) resolution TNG50 or the lower (m7) resolution TNG300 simulations. For EAGLE, the convergence is similar to that of COLIBRE, but it is worse for TNG. The level of agreement with the data is slightly worse for EAGLE, similarly good for TNG100, and substantially worse for TNG50 and TNG300.

7.1.2 The galaxy size–mass relation

The other key observable, besides the SMF, used to calibrate the subgrid feedback is the galaxy SMR, specifically the relation between galaxy stellar mass and the projected galaxy half-stellar mass radius at $z = 0$. Fitting both observables simultaneously is challenging because there is tension between the two require-

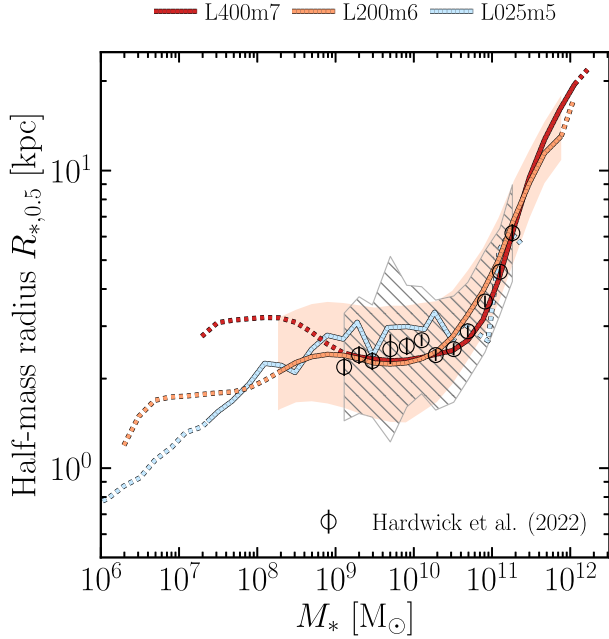


Figure 14. Median $z = 0$ projected galaxy half-stellar-mass radius as a function of galaxy stellar mass for the L400m7 (red), L200m6 (orange), and L025m5 (blue) simulations. The orange shaded region indicates the 16th–84th percentile range for L200m6. The line style switches from solid to dotted where the mass becomes smaller than 100 times the mean initial baryonic particle mass and at high mass where there are fewer than 10 objects per 0.2 dex mass bin. The data points show the observed median half-mass radii of J. A. Hardwick et al. (2022) and the error bars, the 1σ uncertainty on the median. The grey hatched region shows the 1σ scatter in the observed relation. For galaxies resolved with $\gtrsim 10^2$ stellar particles, the agreement of the m6 and m7 resolutions with each other and with the data is excellent. At intermediate mass, the sizes in m5 are slightly larger than for the lower resolutions, but the difference is small compared with the scatter and the deviation from the data is $\lesssim 0.1$ dex.

ments (E. Chaikin et al. 2025a). Feedback not only suppresses star formation, it also tends to increase galaxy sizes at fixed mass (e.g. L. V. Sales et al. 2010). The strong feedback needed to suppress the efficiency of galaxy formation to levels consistent with observation tends to result in galaxies becoming too large. The degree of complexity of the fiducial COLIBRE model for CCSN feedback, specifically the inclusion of a dependence of the injected energy on the pressure of the stellar birth cloud, is driven by the requirement to reproduce both the SMF and the SMR (E. Chaikin et al. 2025a, see R. A. Crain et al. 2015 for a similar demonstration for the EAGLE model).

Fig. 14 shows the median SMR for the three fiducial COLIBRE simulations L400m7 (red), L200m6 (orange), and L025m5 (blue). The orange shaded region indicates the scatter, i.e. the 16th–84th percentile range, for L200m6. For clarity, we do not show the scatter for the other simulations, which is similar to that in L200m6. Galaxy sizes increase gradually with mass up to $M_* \sim 10^9 M_\odot$, but remain nearly constant between 10^9 and $10^{10.5} M_\odot$. The SMR steepens dramatically at $M_* \gtrsim 10^{11} M_\odot$, i.e. when star formation is quenched (see Section 7.2.2) and mass growth is therefore expected to transition from being driven by *in situ* star formation to mergers.

The convergence between the different resolutions is very good. While the m5 simulation predicts larger sizes for inter-

mediate masses, the difference is $\lesssim 0.1$ dex. This discrepancy is probably smaller than the systematic effects due to differences in the way the sizes and masses are measured (possibly implying that we have overfitted the data) and is also small compared to the scatter.

The simulations are compared with the observations from J. A. Hardwick et al. (2022), which were used in the calibration of the model. The J. A. Hardwick et al. (2022) data points are based on Sérsic fits in the g , r , and i bands, converted to mass profiles based on the r -band magnitude and $g - i$ colours, for ≈ 1200 stellar mass-selected galaxies at $0.01 < z < 0.05$ from the extended GALEX Arecibo SDSS Survey (xGASS) (B. Catinella et al. 2018). The volumes of the L200m6 and L400m7 simulations are large enough to cover both the relatively flat part of the SMR below $M_* \approx 10^{10.5} M_\odot$ and the relatively steep part at higher masses. The resolutions of L025m5 and L200m6 are sufficiently high to probe the relation below $10^9 M_\odot$, where they predict the SMR to steepen. The agreement of the median trends with the data is excellent over the full mass range spanned by the observations. The excellent agreement also applies to the scatter, which is indicated by the grey hatched region for the J. A. Hardwick et al. (2022) data. A. D. Ludlow et al. (in preparation) will show that the agreement with the observed sizes extends to other data sets, different redshifts, individual morphological classes, and different definitions of galaxy size.

Finally, we note that the agreement with the data is slightly better than for EAGLE, and much better than for IllustrisTNG (see Appendix F).

7.1.3 The BH mass–stellar mass relation

The BH mass scales inversely with the assumed subgrid efficiency of AGN feedback, ϵ_f , i.e. the fraction of the AGN luminosity that is coupled to the ISM (e.g. C. M. Booth & J. Schaye 2009, 2010). Because cosmological simulations lack the resolution and physics to predict that efficiency, they cannot predict BH masses from first principles. The normalization of the observed relation between stellar mass and the mass of the supermassive BH in the galaxy nucleus, which is only determined robustly at high galaxy masses, $M_* \gg 10^{10} M_\odot$, is therefore used to calibrate ϵ_f . Because previous simulations using efficiencies of order 10 per cent gave reasonable agreement with the data (e.g. C. M. Booth & J. Schaye 2009; J. Schaye et al. 2015), we assumed $\epsilon_f = 0.1$ during the calibration of the m7 resolution using emulators described in E. Chaikin et al. (2025a). Although we did not fit to the observed BH masses, we would have changed the value of ϵ_f if there had been a strong disagreement with the data, and we halved the value to $\epsilon_f = 0.05$ for the m6 and m5 resolutions to improve convergence.

We note that in COLIBRE the subgrid AGN feedback efficiency can have a significant, indirect effect on galaxy properties. This is because the AGN feedback heating temperature, which is the main parameter influencing the actual efficiency of the AGN feedback on resolved scales, is proportional to the BH mass (see Section 3.9). This situation differs from simulations that use a fixed AGN heating temperature, such as OWLS, EAGLE, and FLAMINGO, in which case galaxy properties other than the BH mass are nearly independent of ϵ_f (C. M. Booth & J. Schaye 2010).

Fig. 15 shows the median BH mass as a function of galaxy stellar mass at $z = 0$. The median BH mass jumps from zero to the BH seed mass, which is indicated by a horizontal dashed line, at the stellar mass corresponding to the halo mass at which BHs

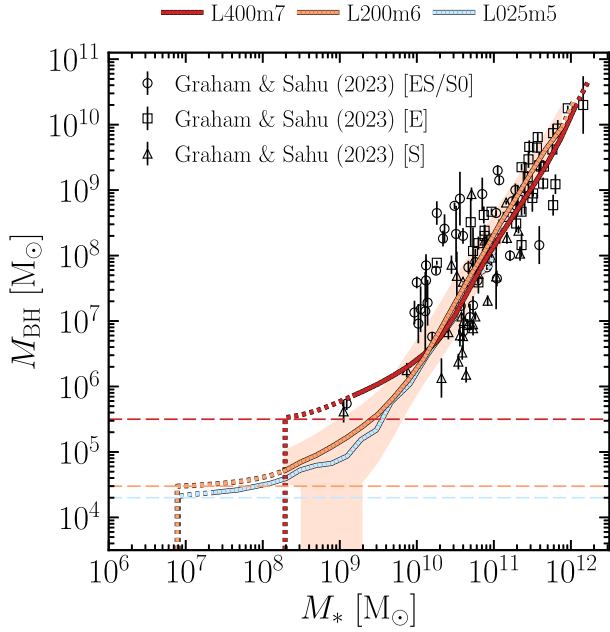


Figure 15. Median $z = 0$ BH mass (i.e. the median mass of the most massive BH in the subhalo at $z = 0$) as a function of galaxy stellar mass. The horizontal dashed lines indicate the BH seed mass, $m_{\text{BH,seed}}$, for the three simulation resolutions. The meanings of the line colours, line styles, and shading are as in Fig. 14. Data points show the compilation of direct BH mass measurements and stellar masses of local galaxies from A. W. Graham & N. Sahu (2023, 2024), with different symbols indicating different morphologies. For BH masses $\gg m_{\text{BH,seed}}$, the convergence is very good, as is the agreement between the median trend and the data.

are seeded. The BH seed mass depends on the resolution (see Table 1) and is one of the four subgrid parameters included in the calibration of the m7 resolution using emulators. As the galaxy mass increases, the BHs lose memory of their seed mass, and the different resolutions converge. The BH seed masses hardly differ between m5 and m6, and the BH mass–galaxy mass relations for these two resolutions are very similar over the full galaxy mass range. The low-resolution m7 simulation uses significantly larger values for the halo mass for BH seeding and the BH seed mass, but by $M_* = 10^{10} M_\odot$, its BH masses agree very well with those in the higher resolution simulations.

The simulations are compared with the compilation of dynamical BH mass measurements of nearby galaxies from²⁷ A. W. Graham & N. Sahu (2023), where data points with different symbols indicate different galaxy morphologies. To facilitate a fair comparison, the stellar masses measured by A. W. Graham & N. Sahu (2023) under the assumption of a P. Kroupa (2002) IMF have been converted to our G. Chabrier (2003) IMF by increasing them by 0.05 dex (M. Bernardi et al. 2010). The agreement between the median trends and the data is excellent over the full mass range covered by the observations. However, because the scatter in the data is large and the number of measurements modest, the observations are not yet very constraining. At high mass, the scatter in the simulations is smaller than for the data, but measuring BH masses is notoriously difficult, and the spread in the data

²⁷We corrected the measurements of A. W. Graham & N. Sahu (2023) following the erratum (A. W. Graham & N. Sahu 2024).

may be affected by measurement errors including unrecognized systematics.

In Appendix F, we show that the BH masses are in good agreement with those in the EAGLE simulations. For $M_* > 10^{11} M_\odot$, this is also the case for TNG, but at lower galaxy masses TNG predicts much higher BH masses.

7.2 Comparison with selected other observations

We now proceed to compare the simulation predictions to observations that were not used to calibrate the subgrid models for CCSN and AGN feedback. Except for the evolution of the cosmic SFR density (Section 7.2.1), we will restrict the discussion to the dependence of a range of fundamental, spatially integrated properties of $z \approx 0$ galaxies on stellar mass, leaving their evolution and spatial trends for future work. The galaxy properties considered here are specific SFRs (sSFRs) and quenched fractions (Section 7.2.2), atomic hydrogen and molecular gas masses (Section 7.2.3), gas and stellar metallicities (Section 7.2.4), dust masses and grain sizes (Section 7.2.5), and CGM X-ray luminosities (Section 7.2.6).

7.2.1 The cosmic star formation history

Fig. 16 shows the evolution of the cosmic SFR density, calculated by summing the instantaneous SFRs of all gas particles. Starting from high redshift, the SFR density initially increases rapidly, reaches a broad peak at $2 \lesssim z \lesssim 4$, and then declines by nearly an order of magnitude.

At redshift $z > 11$, the SFR density increases systematically with resolution because it is dominated by poorly resolved galaxies. At $z < 8$, L200m6 and L400m7 agree very well, but down to $z \approx 2$, the m5 resolution simulations are ≈ 0.1 dex lower. For $z < 2$ all resolutions converge to nearly identical results. The good agreement of L025m5 with L050m5 down to the redshift currently reached by L050m5, $z = 1.1$, implies that the L025 volume is sufficiently large to obtain a converged cosmic SFR density at these redshifts. This is in agreement with the results of J. Schaye et al. (2010) for the OWLS simulations. Looking more closely, we see that while L100m5 agrees very well with the smaller volumes simulated at m5 resolution, for $6 < z < 9$ its SFR density is slightly higher, agreeing nearly perfectly with L200m6, and causing its curve to be hidden by the L200m6 one. However, as can be seen from fig. A1 of E. Chaikin et al. (2025b), this is due to the L100 ICs leading to a slightly higher SFR density at these redshifts than predicted for both smaller and larger volumes.

The simulations are generally in good agreement with the compilation of SFR densities inferred from observations shown in Fig. 16. For $7 < z < 10$, the simulations appear to overpredict the data, but E. Chaikin et al. (2025b) show that this is because of the UV magnitude limit down to which the observations were integrated. If we restrict the cosmic SFR density predicted by the simulations to include only the SFRs of galaxies above this limit, then the agreement with the data is very good also at these redshifts. For $1 < z < 3$, several observational studies prefer a slightly higher SFR density, by ≈ 0.2 dex, though they are mostly consistent with the simulations within the error bars and some observations (e.g. M. Novak et al. 2017) prefer lower values. This tendency of the data at these redshifts to yield slightly higher SFRs than predicted also holds for other simulations and semi-analytic models (e.g. J. Leja et al. 2022, and references therein).

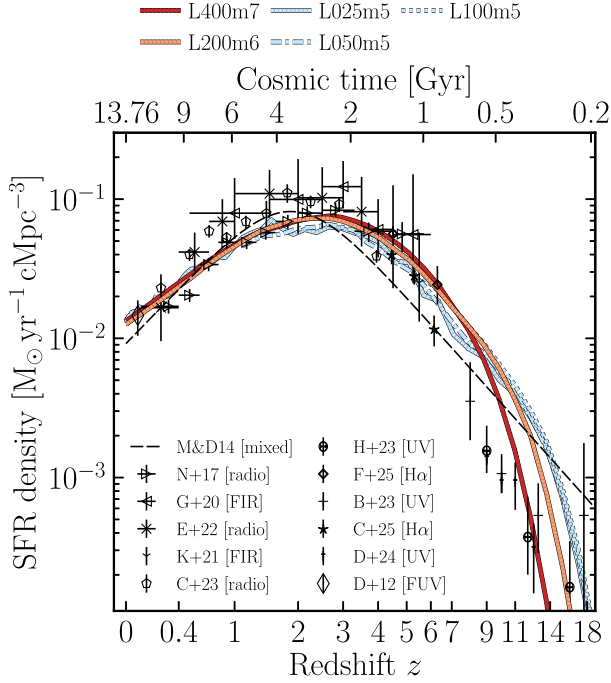


Figure 16. The evolution of the comoving cosmic SFR density. Besides L400m7 (solid red), L200m6 (solid orange), and L025m5 (solid blue), we show L050m5 and L100m5 down to their current redshifts ($z = 1.1$ and 6 , respectively). Data points show recent measurements based on rest-frame ultraviolet (S. P. Driver et al. 2012; R. Bouwens et al. 2023; Y. Harikane et al. 2023; C. T. Donnan et al. 2024), $H\alpha$ (S. Fu et al. 2025; A. Covello-Paz et al. 2025), far-infrared (Gruppioni et al. 2020; Y. Khusanova et al. 2021), and radio (M. Novak et al. 2017; A. Enia et al. 2022; R. K. Cochrane et al. 2023) data, as well as the fit to a compilation of older observations from Madau & Dickinson (2014). Except at very high redshift ($z > 10$), where poorly resolved galaxies dominate and higher resolution thus gives higher SFR densities, the convergence between the m6 and m7 resolutions is excellent. The m5 resolution yields slightly lower SFR densities for $2 \lesssim z \lesssim 10$, but only by $\lesssim 0.1$ dex. The agreement with the data is very good. For $z \gtrsim 8$, the simulations overshoot most of the data, but E. Chaikin et al. (2025b) show that the agreement is excellent if we account for the UV magnitude limit to which those observational measurements have been normalized.

This discrepancy was also present for the EAGLE simulations (e.g. M. Furlong et al. 2015, and Fig. F1) and TNG100 (see Fig. F2), but A. Katsianis et al. (2020) showed that for EAGLE it disappears if virtual observations of EAGLE galaxies are analysed using observational techniques. A similar reduction in the observed SFRs is also required to obtain consistency between the time integral of the SFRs and the evolution of the cosmic stellar mass density obtained by integrating the observed SMF over mass (e.g. M. Shuntov et al. 2025). We conclude that the cosmic star formation history predicted by COLIBRE is consistent with the data.

7.2.2 The main sequence of star-forming galaxies and quenched fractions

Fig. 17 shows the $z = 0$ relation between the sSFR and stellar mass. Only galaxies that are actively star-forming are included, which we define as $sSFR > 10^{-2} \text{ Gyr}^{-1}$. For $M_* \lesssim 10^{10} M_\odot$, the relation remains relatively flat at $sSFR \sim 10^{-1} \text{ Gyr}^{-1}$, which implies $SFR \sim 1 M_\odot \text{ yr}^{-1} (M_*/10^{10} M_\odot)$. The track followed by star-

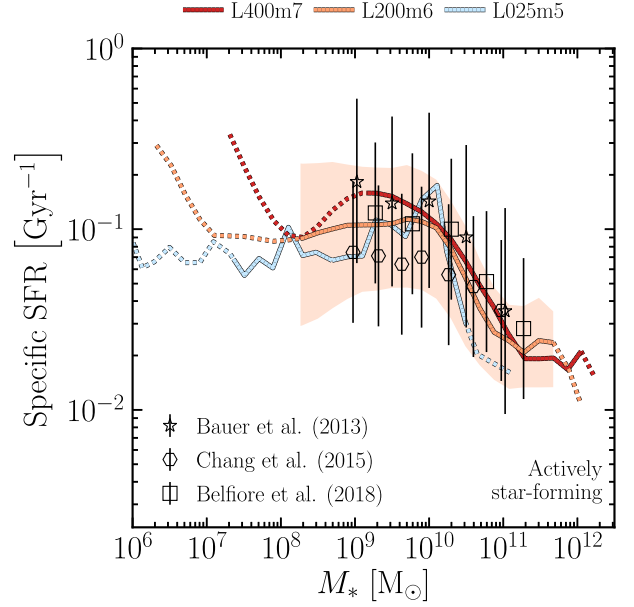


Figure 17. Median $z = 0$ sSFR as a function of galaxy stellar mass for star-forming galaxies, i.e. those with $sSFR > 10^{-2} \text{ Gyr}^{-1}$. The meanings of the line colours, line styles, and shading are as in Fig. 14. The data points show observations of star-forming galaxies from GAMA (A. E. Bauer et al. 2013), SDSS+WISE (Y.-Y. Chang et al. 2015), and MaNGA (F. Belfiore et al. 2018), with the error bars indicating the 1σ scatter. The convergence is excellent for galaxies containing $\gtrsim 10^3$ stellar particles, but even for $\sim 10^2$ particles the differences are $\lesssim 0.2$ dex and smaller than the scatter. The agreement with the data is excellent.

forming galaxies over the mass range for which the SFR is roughly proportional to mass is often referred to as the ‘main sequence’. At very low mass, the curves turn up, but comparison of the different simulations shows that this upturn is a resolution effect. The curves bend down above $10^{10} M_\odot$ and by $10^{11} M_\odot$ they approach the selection limit of $sSFR > 10^{-2} \text{ Gyr}^{-1}$. The convergence is excellent when galaxies are resolved with more than 10^3 particles. For 10^2 particles, which is more than sufficient to obtain converged stellar masses (see Fig. 13), differences of ≈ 0.2 dex are seen, but these differences are small compared to the scatter, which is indicated for L200m6 by the shaded region. The simulations are compared with observations of star-forming galaxies at $z < 0.15$ from MaNGA (F. Belfiore et al. 2018), at $z < 0.2$ from SDSS+WISE (Y.-Y. Chang et al. 2015) and at $0.05 < z < 0.32$ from GAMA (A. E. Bauer et al. 2013), selected as main-sequence galaxies in the SFR – M_* diagram, selected in colour–colour space, and selected based on $H\alpha$ equivalent width, respectively. The predicted relations are consistent with the data.

Fig. 18 shows the mass dependence of the fraction of galaxies that are excluded from Fig. 17, i.e. the fraction of $z = 0$ galaxies in which star formation is quenched, $sSFR < 10^{-2} \text{ Gyr}^{-1}$, and are thus evolving passively. The passive fraction increases rapidly between $M_* \sim 10^{9.5}$ and $10^{11} M_\odot$. The convergence between L200m6 and L400m7 is excellent in this regime, but L025m5 appears to predict lower quenched fractions below $10^{10} M_\odot$. E. Chaikin et al. (2025b) show that for central galaxies the quenched fractions in L025m5 and L025m6 are converged, but that for satellites with $M_* \lesssim 10^9 M_\odot$ the quenched fraction is lower for L025m5. For $M_* > 10^9 M_\odot$, the difference visible in Fig. 18 is mostly due to the small simulation volume of L025m5, which

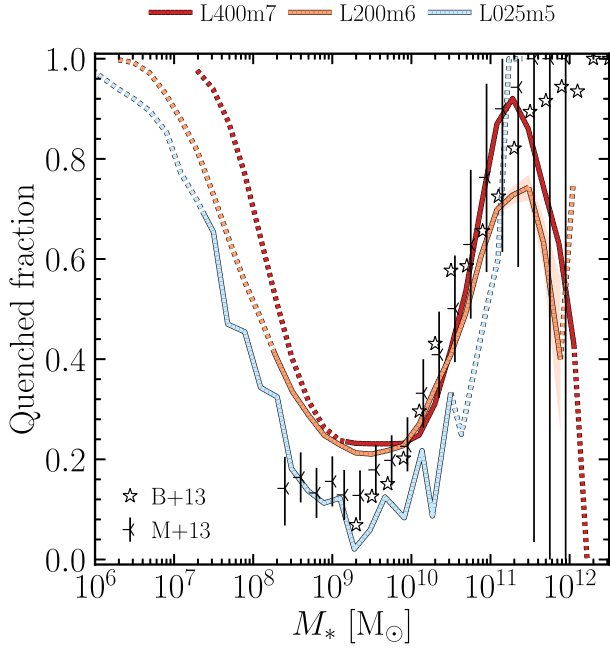


Figure 18. Fraction of $z = 0$ galaxies that are quenched, i.e. $\text{sSFR} < 10^{-2} \text{ Gyr}^{-1}$, as a function of galaxy stellar mass. The meanings of the line colours and line styles are as in Fig. 14. The orange shaded region indicates the Clopper–Pearson 68 per cent confidence interval for L200m6. The data points show observations from GAMA (A. E. Bauer et al. 2013), which use the same definition of quenched as applied to the simulations, and from the COSMOS/UVISTA Survey (A. Muzzin et al. 2013) at $0.2 < z < 0.5$, for which quenched galaxies are defined based on *UVJ* colours. The convergence is good in the mass range over which the quenched fraction increases rapidly, particularly between m6 and m7. It is less good for L025m5, but this is mostly a box size effect, which leads to smaller fractions and lower host masses of satellites in L025m5. E. Chaikin et al. (2025b) demonstrate that the convergence is excellent for centrals, but that the quenched fraction decreases with resolution for satellites. The agreement with the data is excellent at intermediate masses, but there are hints that the quenched fractions may be too large for $M_* \lesssim 10^9 M_\odot$ and too small for $M_* \sim 10^{12} M_\odot$. However, the sSFRs of very high-mass ‘star-forming’ galaxies are only a factor of ≈ 2 above the critical value (see Fig. 17) and the predicted high-mass downturn/dip of the quenched fraction disappears if we measure the SFR in an aperture of 10 kpc or if we add 0.3 dex lognormal scatter to the stellar masses.

biases the satellite fraction and the host masses of satellite galaxies to be low. Because the passive fraction of galaxies with $M_* \ll 10^{11} M_\odot$ depends mostly on environment (e.g. Y.-j. Peng et al. 2010), this leads to the quenched fraction being biased low.

Quenched fractions increase with decreasing mass for $M_* < 10^9 M_\odot$. At the lowest masses, where the sSFR of actively star-forming galaxies turns up in Fig. 17, the quenched fractions are boosted by a resolution effect. At such low masses even a single star-forming gas particle can push the galaxies above the main sequence. Because the stellar masses are calibrated to fit the data, these excessively high sSFRs must be compensated for, which implies that most galaxies must have an SFR of zero. The low-mass upturn in the quenched fractions is partly caused by an increase in the satellite fraction towards lower masses, although E. Chaikin et al. (2025b) show that there is also an upturn for central galaxies.

The passive fractions decline towards $M_* \gtrsim 10^{11.5} M_\odot$, indicating that AGN feedback does not keep the most massive galaxies fully quenched. In Section 8, we will show that this is also true for the simulations using hybrid AGN feedback. The difference between the m6 and m7 resolutions is greater than at lower masses, with the higher resolution simulation yielding smaller quenched fractions. This reflects the greater sensitivity to resolution of AGN feedback compared to stellar feedback caused by the increase of the typical density of the gas in which AGN feedback energy is injected with resolution (see Section 6.2).

The quenched fractions are generally in good agreement with the $z \approx 0.1$ observations from the GAMA survey (A. E. Bauer et al. 2013), which were converted to our $\text{sSFR} < 10^{-2} \text{ Gyr}^{-1}$ criterion by P. Behroozi et al. (2019), and with the measurements of A. Muzzin et al. (2013), who apply *UVJ* colour selection to $z \approx 0.3$ galaxies from the COSMOS/UVISTA survey. There is no sign of an upturn in the data at low masses, $M_* < 10^9 M_\odot$, although this regime is only probed by A. Muzzin et al. (2013) and selection effects may lead to an underestimate of the passive fraction (e.g. S. Kaviraj et al. 2025). In contrast to COLIBRE, at high mass, the observed quenched fraction does not turn over. This may suggest that the AGN feedback in the fiducial model is insufficiently effective, though we note that at these masses, $M_* \gtrsim 10^{11.5} M_\odot$, the star-forming galaxies in the simulation have sSFRs only slightly above the threshold of $\text{sSFR} = 10^{-2} \text{ Gyr}^{-1}$ (see Fig. 17) and that the quenched fraction increases again for $M_* > 10^{12} M_\odot$. Furthermore, the downturn/dip in the sSFR disappears if we measure the SFR in an aperture of $\lesssim 10 \text{ kpc}$, which may be appropriate for fibre-based measurements, or if we add ≥ 0.3 dex lognormal scatter to the mass measurements (see Fig. D1), which may be more realistic for BCGs than our fiducial scatter of 0.1 dex. We conclude that while the agreement with the data is excellent for $10^9 \lesssim M_*/M_\odot \lesssim 10^{11.5}$, there may be discrepancies outside of this mass range. A more careful comparison with the data, using virtual observations, is necessary to establish whether the quenched fractions are too high for $M_* < 10^9 M_\odot$ and too low for $M_* > 10^{11.5} M_\odot$.

In Appendix F, we show that the observed main sequence is similarly well reproduced by the EAGLE and TNG simulations. At m5 resolution, quenched fractions are similar in EAGLE and COLIBRE, but at m6 resolution, EAGLE fits the data somewhat less well. For $M_* \lesssim 10^{10} M_\odot$, both the numerical convergence and the level of agreement with the data are similar for TNG and COLIBRE, but at higher masses both are generally worse for TNG.

7.2.3 Atomic and molecular gas

The fractions of the baryonic mass that are atomic and molecular provide a probe of the warm and cold phases of the ISM. Unlike previous large-volume simulations, such as EAGLE, which do not directly model the cold phase, these fractions are directly predicted by COLIBRE. They therefore do not need to be estimated by post-processing the simulation output using prescriptions whose applicability is doubtful if the gas distribution is affected by the use of a temperature/pressure floor. Furthermore, COLIBRE also predicts the dust content, which is critical given that the formation of molecules occurs predominantly on the surfaces of dust grains and that dust has a strong effect on the thermal balance of the ISM. The direct modelling of dust grains, the coupling between dust physics and chemistry, and the non-equilibrium treatment of hydrogen and helium set COLIBRE

apart also from the few cosmological simulations that allow cooling to cold ISM temperatures, such as FIREbox and NewHorizon.

However, it is important to note that while the resolution of COLIBRE suffices to distinguish warm ($T \sim 10^4$ K) and cold ($T \ll 10^4$ K) ISM phases, it is insufficient to model the internal structure of molecular clouds. Because dust grain growth occurs predominantly in molecular cloud cores, the predicted dust content would be significantly reduced if the simulation did not boost the rates of gas accretion on to grains and grain coagulation using a subgrid clumping factor (see Section 3.2 and J. W. Trayford et al. 2025). Note that clumping factors are also required by the high-resolution simulations of individual galaxies on which the prescriptions for estimating molecular fractions that are employed by many cosmological simulations are based. For example, N. Y. Gnedin & A. V. Kravtsov (2011) boost the growth rate of H_2 on dust grains by a factor of 30. Although we do not boost the H_2 growth rate, for low-mass galaxies ($M_* < 10^{10} M_\odot$ at m6 resolution), the galaxy-scale molecular fraction would be much reduced without the use of a clumping factor for dust grain growth. The atomic mass fraction, on the other hand, is insensitive to the use of this clumping factor.

Fig. 19 shows the median atomic (left panel) and molecular hydrogen (right panel) mass relative to the stellar mass as a function of stellar mass. The sharp drops at low mass to median fractions of (nearly) zero are clearly a resolution effect because they shift to lower masses for higher resolution. For molecular hydrogen, the different resolutions begin to diverge when the line styles are still solid, indicating that (many) more than 100 stellar particles are required for convergence of the medians. However, given that the scatter (indicated by the shaded region for m6 resolution) is extremely large, which reflects the bimodality of the distribution, the median is more sensitive to resolution than the mean (not shown). At intermediate masses, $10^9 \lesssim M_*/M_\odot \lesssim 10^{10.5}$, convergence between m6 and m7 is excellent. The high-resolution simulation predicts slightly higher gas fractions, but the differences are small compared to the 1σ scatter. At higher masses, where the quenched fraction becomes large (see Fig. 18), the differences increase, though they remain smaller than the scatter. In this regime, the convergence is much better, and the dips in the gas fractions are much smaller, for the means (not shown) than for the medians, which again reflects the bimodal distributions of the gas fractions. Indeed, for $M_* \sim 10^{11} M_\odot$, the increase in the gas fractions with increasing resolution mirrors the decrease in the quenched fractions with resolution (see Fig. 18).

In the mass range for which the convergence is good, the gas fractions decrease with the stellar mass. Comparing the two panels, we see that at low mass, $M_* < 10^9 M_\odot$, the molecular fraction is very small compared to the atomic fraction, but that they become comparable at higher masses. The increase in the H_2 fraction with stellar mass for $M_* \gtrsim 10^{11.5} M_\odot$ mirrors the decrease in the quenched fraction (see Fig. 18).

The simulation predictions are compared with observations of H I from the xGASS (B. Catinella et al. 2018) and ALFALFA (D. Scholte et al. 2024) surveys, and of CO from the xCOLD GASS survey (A. Saintonge et al. 2017). Each of these surveys uses a sample of stellar-mass-selected galaxies. A. Saintonge et al. (2017) convert CO luminosity to H_2 mass using the multivariate conversion factor of G. Accurso et al. (2017). A. Saintonge et al. (2017) correct for the presence of helium by multiplying the H_2 masses by a factor of 1.36, but we have undone that correction. We note that the conversion from CO to H_2 introduces a large

systematic uncertainty, a factor ≈ 2 for galaxies such as the Milky Way and a much larger factor for low-metallicity galaxies (e.g. A. D. Bolatto, M. Wolfire & A. K. Leroy 2013). For B. Catinella et al. (2018) and A. Saintonge et al. (2017), who published the data for individual galaxies, we have computed the medians (and associated errors using bootstrap resampling), and the 16th–84th percentile scatter ourselves. Because the fraction of upper limits is substantial (between 16 and 50 per cent for all mass bins), we computed the 16th percentile in two ways: once treating upper limits as detections (16th–84th percentile scatter indicated by the top grey hatched regions) and once setting them to zero (indicated by the combination of the two hatched regions), in which case the 16th percentiles are zero for all mass bins. The true 16th percentile must be in between these two limiting cases.

For m6 resolution, the predicted median atomic masses are fully consistent with the data from $M_* \sim 10^9 M_\odot$ all the way to the most massive observed galaxies ($M_* \approx 10^{11.8} M_\odot$). The agreement is similarly good for m7 resolution, except at $M_* \sim 10^{11} M_\odot$, where the median H I mass is consistent with B. Catinella et al. (2018) but lower than that measured by D. Scholte et al. (2024). At m5 resolution, the median atomic masses are ≈ 0.1 dex too large, but this discrepancy is much smaller than the scatter. For both m6 and m7 resolutions, the median molecular masses are in excellent agreement with the data over the full observed mass range of $10^9 < M_*/M_\odot < 10^{11.3}$. At m5 resolution, the H_2 masses are slightly higher, typically by ≈ 0.1 dex, which is again a small difference compared to the scatter. The excellent agreement extends to the scatter, which is fully consistent with both the atomic and molecular gas data over the full observed stellar mass range.

7.2.4 The mass–metallicity relation

In the publication presenting the COLIBRE model for chemical enrichment, C. Correa et al. (in preparation) already compare results from small-volume simulations, L025m6 and L025m7, with a number of observables probing metal abundances at low redshift. Here, we will present the $z = 0$ relations between stellar mass and both gas and stellar metallicities, finding results consistent with C. Correa et al. (in preparation). We leave evolution, spatially resolved trends, and relative abundances for future work, though we note that C. Correa et al. (in preparation) already showed that the small-volume simulations reproduce the observed trend of alpha enhancement with stellar mass for low- z galaxies as well as abundance ratios of individual elements as a function of metallicity observed for stars in the Milky Way Galaxy.

Fig. 20 shows the median oxygen abundance in the ISM as a function of galaxy stellar mass, where the ISM is defined as gas with density $n_H > 0.1 \text{ cm}^{-3}$ and temperature $T < 10^{4.5}$ K. The oxygen abundance, O/H , is computed as the ratio of the total number of gas-phase oxygen nuclei, i.e. excluding oxygen depleted onto dust grains, to the total number of hydrogen nuclei. Because measurements of gas metallicities are based on observations of H II regions, only star-forming galaxies, i.e. $sSFR > 10^{-2} \text{ Gyr}^{-1}$, are included.

The gas metallicity increases with mass, but the relation becomes less steep with increasing mass. The convergence with resolution does not appear to be as good as for the observables considered so far. While the resolution-dependent sharp drop below stellar masses corresponding to $\sim 10^2$ particles is similar to the resolution effects seen for other observables, differences similar in magnitude to the scatter between objects (in-

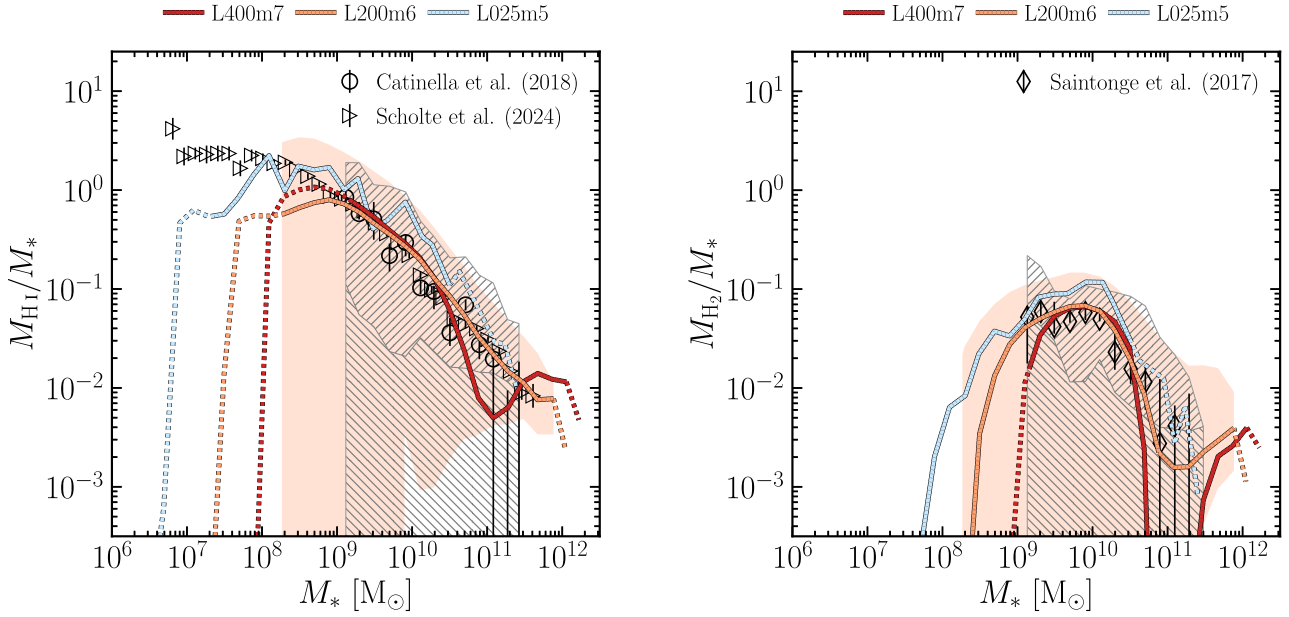


Figure 19. Median ratios of H I to stellar mass (left panel) and molecular hydrogen to stellar mass (right panel) as a function of galaxy stellar mass at $z = 0$. The meanings of the line colours, line styles, and shading are as in Fig. 14. The data points show $z \approx 0$ observations of H I from xGASS (B. Catinella et al. 2018) and from ALFALFA (D. Scholte et al. 2024), and of CO from xCOLD GASS, converted to molecular hydrogen masses (A. Saintonge et al. 2017). These are all stellar-mass-selected surveys. The upper grey hatched regions (striking towards the top right) indicate the 16th–84th percentile scatter in the xGASS and xCOLD GASS surveys with non-detections set to their upper limits. The lower grey hatched regions (striking towards the bottom left) extend the 16th percentiles to their lower limit, obtained by setting non-detections to zero; they therefore indicate the uncertainty on the downward scatter. The convergence is good for intermediate masses, but at the low-mass end, the medians turn over at lower masses for higher resolutions. Convergence requires more particles for H_2 than for H I. For $M_* \sim 10^{11} M_\odot$, the differences in the medians between the different resolutions are larger, particularly for H_2 , but the differences remain small compared to the scatter. The agreement with the data is excellent. The predicted downward scatter (shown only for L200m6) is very large, with the 16th percentile corresponding to negligible gas masses for all but the highest stellar mass bins. This large scatter is fully consistent with the observations.

indicated by the shaded region for L200m6) remain up to $M_* \sim 10^{11} M_\odot$, with higher resolution simulations tending to predict lower metallicities. However, the differences are actually quite small. For $10^9 \lesssim M_*/M_\odot \lesssim 10^{11} M_\odot$, the difference between m7 and m5 is ≈ 0.2 dex and the difference between m6 and m5 is only ≈ 0.1 dex. For higher masses, which are not sampled at m5 resolution, convergence between m7 and m6 is very good. We note that C. Correa et al. (in preparation) show that the slope of the relation depends somewhat on the assumed value of the turbulent diffusion coefficient, with stronger diffusion resulting in lower gas metallicities in massive galaxies.

The simulation predictions are compared with three sets of $z \approx 0.1$ observations. A. Fraser-McKelvie et al. (2022) used strong lines calibrated using the electron-temperature (T_e) method to measure the oxygen abundance within one effective radius for 472 star-forming galaxies from the SAMI survey. C. A. Tremonti et al. (2004) used comparisons of strong lines with photoionization models to measure the metallicity within the 3 arcsec fibre for $\approx 5 \times 10^4$ galaxies from the SDSS survey. M. Curti et al. (2020) used strong-line diagnostics calibrated on T_e measurements of individual low-metallicity and stacks of high-metallicity galaxies to measure fibre metallicities for $\approx 1.5 \times 10^5$ SDSS galaxies. The spread between the different observations reflects well-known systematic uncertainties, with the T_e -calibrated methods giving systematically lower metallicities than the theoretically calibrated strong-line diagnostics. Given these uncertainties, all simulations are consistent with the data. The m7, m6, and m5 resolutions agree particularly well with C. A. Tremonti et al.

(2004), M. Curti et al. (2020), and A. Fraser-McKelvie et al. (2022), respectively, but each simulation is inconsistent with at least one of these data sets.

The relation between stellar metallicity and galaxy stellar mass is shown in Fig. 21. In contrast to the gas metallicities, these do not depend directly on the dust content. Because of that, and because they are effectively time averages of the metallicity of star-forming gas, predictions for stellar metallicities are more robust than predictions for gas metallicities. As observations typically use diagnostics that are sensitive to iron and magnesium, we show $[\text{Fe}/\text{H}]$ in the left panel and $[\text{Mg}/\text{H}]$ in the right panel, where square brackets indicate $\log_{10}$ of the abundance in solar units ($12 + \log_{10}(\text{Fe}/\text{H})_\odot = 7.5$ and $12 + \log_{10}(\text{Mg}/\text{H})_\odot = 7.6$; M. Asplund et al. 2009). The Mg metallicity typically exceeds that of Fe, particularly for high-mass galaxies, $M_* \gtrsim 10^{11} M_\odot$. Such high-mass α -enhancement is observed (D. Thomas et al. 2010) and is consistent with the prediction from the EAGLE simulations, where it is caused by more massive galaxies forming their stars earlier and more rapidly as a result of AGN feedback (M. C. Segers et al. 2016b). The stellar mass–metallicity relation resembles that of the gas metallicity (see Fig. 20), though the decrease in the slope towards higher masses is more pronounced, even becoming negative at $M_* \sim 10^{11} M_\odot$.

The convergence with numerical resolution is excellent, better than for the gas metallicity, with differences between m7 and m5 $\lesssim 0.1$ dex for galaxies resolved with $\gtrsim 10$ stellar particles.

We compare with observations from A. Gallazzi et al. (2005) for $\sim 10^5$ redshift $z \approx 0.1$ SDSS galaxies, measured within the

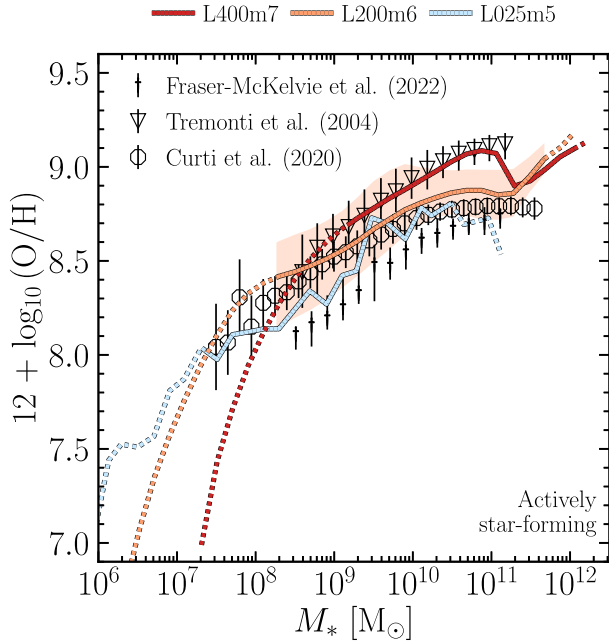


Figure 20. Median $z = 0$ gas-phase oxygen abundance as a function of galaxy stellar mass. Only star-forming galaxies, i.e. $\text{sSFR} > 10^{-2} \text{ Gyr}^{-1}$, are included. The metallicity is computed as the ratio of the total number of gas-phase oxygen nuclei, i.e. excluding oxygen depleted onto dust grains, to the total number of hydrogen nuclei, including only gas with density $n_{\text{H}} > 0.1 \text{ cm}^{-3}$ and temperature $T < 10^{4.5} \text{ K}$. The meanings of the line colours, line styles, and shading are as in Fig. 14. The data points show $z \approx 0.1$ observations from A. Fraser-McKelvie et al. (2022) (based on strong-line diagnostics calibrated on T_e measurements), C. A. Tremonti et al. (2004, based on strong lines and photoionization models), and M. Curti et al. (2020, based on strong-line diagnostics calibrated on T_e measurements). The sharp downturns at the low-mass end are clearly a resolution effect. For better-sampled galaxies, higher resolutions give systematically lower metallicities, with the possible exception of the most massive galaxies. The difference between m7 and m5 is $\lesssim 0.2$ dex, with m6 falling in between. Given the systematic differences between different data sets, all simulations are consistent with the observations.

3 arcsec SDSS fibre, of H. J. Zahid et al. (2017) for 2×10^5 redshift $z < 0.25$ star-forming SDSS galaxies, also measured within the 3 arcsec SDSS fibre, and from E. N. Kirby et al. (2013) for individual dwarf irregular and spheroidal satellite galaxies of the Milky Way and M31. Following R. M. Yates et al. (2021), we assume 0.2 dex metallicity errors for H. J. Zahid et al. (2017). The observed metallicities in solar units are taken to be identical for Fe and Mg. For stellar masses corresponding to $\gtrsim 10$ stellar particles up to $M_* \sim 10^{11} M_{\odot}$, the simulations are fully consistent with the data. Higher masses are only probed by A. Gallazzi et al. (2005). For $M_* \sim 10^{12} M_{\odot}$, they find a typical metallicity that is 0.4 dex higher than the predicted $[\text{Fe}/\text{H}]$, but the predicted $[\text{Mg}/\text{H}]$ is consistent with the data within the error bars. The strong dependence on the chosen element, and the fact that aperture effects and Eddington bias (see Fig. D1) can be important at these very high masses, warrant a like-for-like analysis using virtual observations, which we leave for future work.

In Appendix F, we show that both EAGLE and IllustrisTNG predict much shallower mass–metallicity relations than COLIBRE, resulting in too high metallicities for $M_* < 10^{10} M_{\odot}$, particularly at m5 and m6 resolutions for TNG and EAGLE, re-

spectively. However, for $M_* \gtrsim 10^{11.5} M_{\odot}$, TNG agrees better with the observations of A. Gallazzi et al. (2005) than COLIBRE does.

7.2.5 Dust

The inclusion of a model for the evolution of dust grains opens up a new class of observables that can be directly predicted. J. W. Trayford et al. (2025) already presented predictions from L025m6 simulations with the COLIBRE model, finding good agreement with observations of the evolution of the cosmic mass density in dust, the $z = 0$ dust mass function and various dust scaling relations. Here, we present the first results for dust from the flagship COLIBRE simulations, which include both lower and higher resolutions than the runs analysed by J. W. Trayford et al. (2025) and, for the m6 resolution, have a volume 512 times larger. Future work will present a wider range of dust-related observables as well as their evolution and spatially resolved trends.

Fig. 22 shows the median dust mass as a function of stellar mass. The convergence with resolution is excellent for stellar masses corresponding to $\sim 10^3$ stellar particles up to $M_* \approx 10^{10.7} M_{\odot}$. Towards lower masses systematic differences appear, with higher resolution simulations yielding higher dust masses. At $M_* \sim 10^{11} M_{\odot}$, the convergence is poor, though the differences are still within the scatter (indicated by the shaded region for m6). While the m6 and m7 resolutions both show a dip in the median dust content, the dip is an order of magnitude deeper for m7. For higher masses, the two resolutions converge, yielding very good agreement for $M_* \gtrsim 10^{11.5} M_{\odot}$. The resolution dependence parallels that of the molecular mass fractions (compare with the right panel of Fig. 19). In the mass range that is well converged and for which the quenched fraction is still small (i.e. $M_* \lesssim 10^{10} M_{\odot}$, see Fig. 18), the dust mass is roughly proportional to the stellar mass, which parallels the constant molecular-to-stellar mass ratio in the same mass range (see Fig. 19). As is the case for the molecular mass, there is a very large downward scatter in the dust masses. The close similarity between the dust and molecular mass trends suggests that a large fraction of the dust resides in the cold interstellar gas phase, as can indeed be seen in the bottom-left panel of Fig. 9.

As discussed in J. W. Trayford et al. (2025) and in Section 3.2, because COLIBRE does not resolve the molecular cloud cores where most grain growth is expected to occur, the growth rates due to gas accretion and coagulation have to be boosted by a clumping factor. In order to limit the boost to densities typical of the cold ISM, the clumping factor increases monotonically from 1 (i.e. no boost) to 100 over the density range $n_{\text{H}} = 10^{-1}$ to 10^2 cm^{-3} (see equation 2). The dust content becomes insensitive to the value of the clumping factor when grain growth is limited by the availability of bottleneck elements. J. W. Trayford et al. (2025) show that for m6 resolution this occurs in high-density gas ($n_{\text{H}} \gg 10 \text{ cm}^{-3}$) provided the clumping factor is $\gtrsim 10$. An L025m6 simulation that does not use a clumping factor (not shown) results in median $z = 0$ dust masses that are a factor of ≈ 3 lower for $M_* > 10^{10} M_{\odot}$. The effect of the clumping factor is maximum for $M_* \sim 10^9 M_{\odot}$, where it increases the median dust masses by about an order of magnitude.

The small open circles and crosses in Fig. 22 indicate observational measurements presented by S. Bianchi et al. (2018) and I. De Looze et al. (2020), respectively. These are based on SED fitting of nearby galaxies from a variety of surveys. Because none of the surveys are mass-selected, it is unclear how representa-

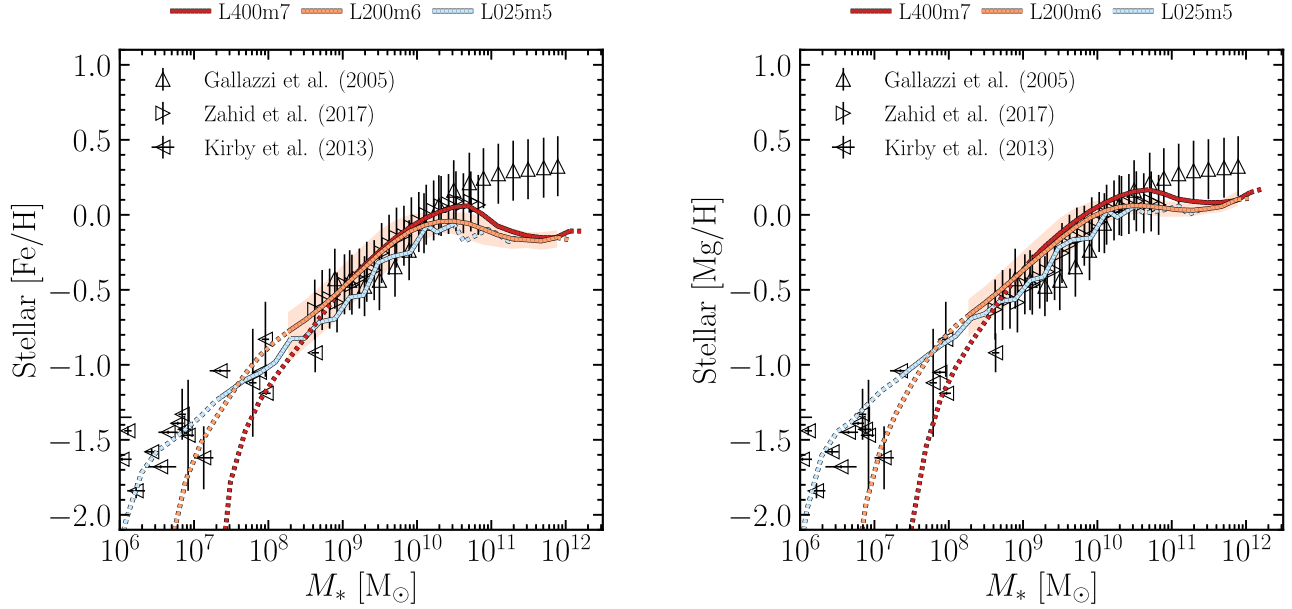


Figure 21. Median $z = 0$ stellar iron (left panel) and magnesium (right panel) abundance as a function of galaxy stellar mass. The meanings of the line colours, line styles, and shading are as in Fig. 14. The data points show observations of A. Gallazzi et al. (2005) and H. J. Zahid et al. (2017), both for SDSS galaxies measured within the 3 arcsec SDSS fibres, and from E. N. Kirby et al. (2013) for individual dwarf satellite galaxies of the Milky Way and M31. The observed metallicities in solar units are assumed to be identical for iron and magnesium. All abundances are relative to the solar abundances from M. Asplund et al. (2009). The downturn below masses corresponding to ~ 10 stellar particles is clearly a resolution effect. For larger particle numbers, the differences between m7 and m5 are $\lesssim 0.1$ dex. The agreement with the data is generally very good. For massive galaxies, $M_* \gtrsim 10^{11} M_\odot$, the observed metallicity is higher than the predicted iron abundance, but for magnesium the difference is only $\lesssim 0.2$ dex and consistent within the errors.

tive the trends with mass are. Comparisons with multivariate trends could reduce this concern, which we leave for future work. The simulation predictions are in good agreement with the data, reproducing both the trend followed by the majority of galaxies and the presence of a population with much smaller dust masses.

The COLIBRE dust model includes two grain sizes, with radii of 0.01 and 0.1 μm . As discussed in Section 3.2, the coagulation rate is boosted by a density-dependent clumping factor. In addition, to improve the convergence with resolution, in the low-resolution m7 simulation, the e -folding time for grain coagulation is further reduced by a factor $10^{0.5}$. This decreases the small-to-large grain dust mass ratio.

Fig. 23 shows the median ratio of the dust mass in small grains to that in large grains as a function of stellar mass. For stellar masses corresponding to $> 10^2$ particles, the convergence with resolution is very good, particularly between m5 and m6. For stellar masses corresponding to $< 10^2$ particles the small-to-large ratio drops rapidly towards the value of 10^{-1} used for the grain seeds. Above this resolution limit, grain sizes tend to increase with galaxy mass, except for a small excursion towards smaller grain sizes at the stellar mass for which the molecular mass fractions and the dust masses are minimum (see Figs 19 and 22, respectively). However, the trend of a decreasing median small-to-large grain ratio with mass is much smaller than the scatter.

The bottom-right panel of Fig. 9, and fig. 20 of J. W. Trayford et al. (2025), show that the size distribution depends on the gas density, with higher densities favouring larger grain sizes, as expected due to the stronger density dependence for coagulation than for shattering (see Section 3.2). The density dependence reported by J. W. Trayford et al. (2025) is strong, with the size ratio decreasing from $\sim 10^{-0.5}$ for $n_{\text{H}} \sim 10 \text{ cm}^{-3}$ to $\sim 10^{-1.2}$ for $n_{\text{H}} \sim 10^2 \text{ cm}^{-3}$.

This suggests that the trend of grain size with mass reflects an increase in the typical density of the dusty ISM with mass. The strong density dependence also implies that grain sizes inferred from virtual observations could be larger, and hence that the size ratio could be smaller, in star-forming regions than the galaxy-wide averages presented here. As the grain sizes of observed galaxies are inferred from SED fitting, the observed grain size ratios may thus be biased low.

The simulation results are compared with a compilation of observations from M. Relaño et al. (2020, 2022). Their measurements are for nearby galaxies from the same surveys as used for the dust mass measurements shown in Fig. 22. The surveys are not mass-limited and the size ratio, which is determined by SED fitting, is sensitive to the assumed dust model. Despite these limitations, the predictions are in broad agreement with the observations, with most of the low-mass ($M_* < 10^{10} M_\odot$) data points falling within the 1σ scatter (indicated by the shaded region for L200m6). A subset of the observed high-mass ($M_* > 10^{10} M_\odot$) galaxies are inferred to have relatively large grain sizes, with size ratios similar to what COLIBRE predicts for densities of $n_{\text{H}} \sim 10^2 \text{ cm}^{-3}$ (see fig. 10 of J. W. Trayford et al. 2025).

7.2.6 X-ray luminosity of the circumgalactic medium

The previous figures compared the predictions for galaxy stellar and ISM properties with observations. Fig. 24 instead shows a property of the CGM that is constrained by observations over a wide range of halo mass. The predicted mean $z = 0$ X-ray luminosity of the CGM, $L_{\text{X,CGM}}$, is plotted as a function of halo mass, M_{500c} . Here, the CGM X-ray luminosity of an individual object is computed by summing the luminosities of its gas particles in the

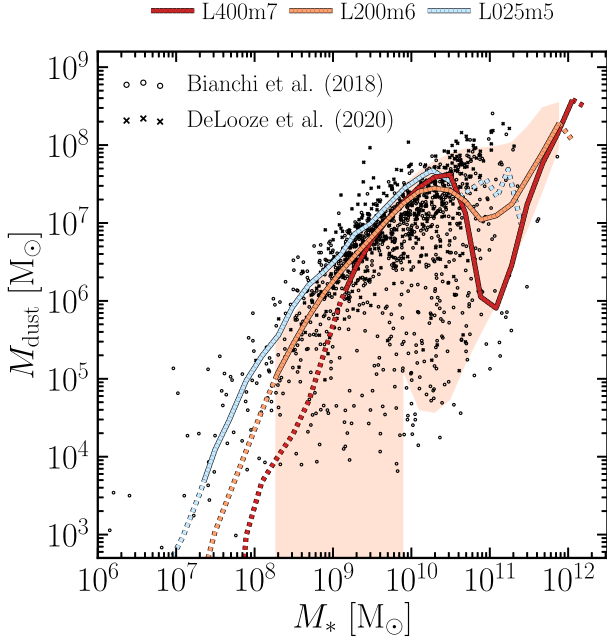


Figure 22. Median $z = 0$ dust mass as a function of galaxy stellar mass. The meanings of the line colours, line styles, and shading are as in Fig. 14. The small open circles indicate observations from *DustPedia* taken from S. Bianchi et al. (2018) and the small crosses indicate the compilation of survey data from I. De Looze et al. (2020). None of the surveys are mass-selected. At low masses, corresponding to $\ll 10^3$ particles, and at $M_* \sim 10^{11} M_\odot$, higher resolution simulations yield higher dust masses, while the convergence is excellent at other masses. Comparison to Fig. 19 shows that the differences between the resolutions reflect differences in the cold gas content. The agreement with the data, including the large downward scatter, is very good.

radial range $(0.15 - 1) \times R_{500c}$, where R_{500c} is the radius within which the mean total matter density is 500 times the critical density of the universe, and M_{500c} is the mass inside R_{500c} . Gas particles bound to satellites are excluded, but this has a negligible effect. The particle luminosities are computed in the 0.5–2 keV band using CLOUDY (version 17.02, G. J. Ferland et al. 2017) with the method described in J. Braspennig et al. (2024), but accounting for the non-equilibrium free electron densities from hydrogen and helium. We plot the mean instead of the median luminosity to facilitate comparison with stacks of observed objects, which is how emission is detected statistically in the observations of the CGM of low-mass galaxies from Y. Zhang et al. (2024) that we compare with.

The radial cut $(0.15 - 1) \times R_{500c}$, sometimes referred to as ‘core excision’, is commonly adopted in comparisons of simulations and observations, including by Y. Zhang et al. (2024). Core excision results in the exclusion of emission from the central galaxy’s ISM and the inner halo. Such a cut is motivated by the fact that observational analyses attempt to subtract emission from galaxies. Another motivation is that simulations generally do not model several important sources of X-ray emission within galaxies, such as X-ray binaries, AGN, and SN remnants. Furthermore, the X-ray emission from the ISM and inner CGM is sensitive to the precise subgrid implementation of feedback. In COLIBRE, particles with high gas densities that have recently been directly heated by the large temperature increments associated with SN and AGN feedback, as well as the immediate neighbours of these particles,

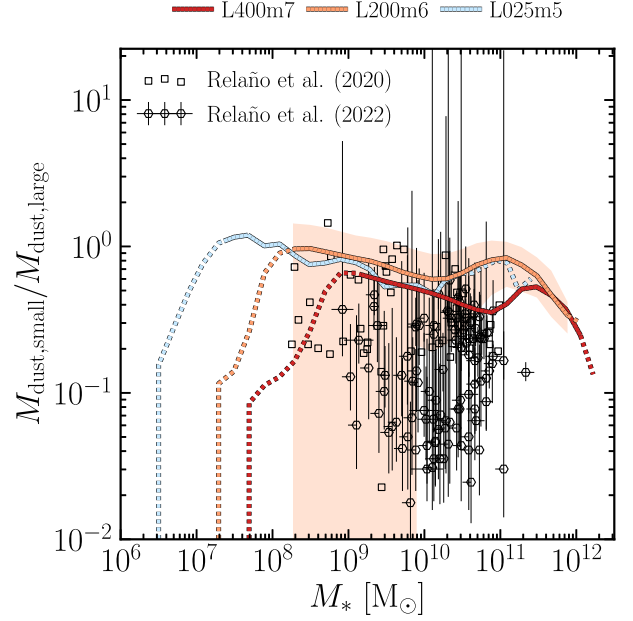


Figure 23. Median $z = 0$ ratio of the dust mass in small ($0.01 \mu\text{m}$) grains and the dust mass in large ($0.1 \mu\text{m}$) grains as a function of galaxy stellar mass. The meanings of the line colours, line styles, and shading are as in Fig. 14. The convergence with resolution is good for galaxy masses corresponding to $> 10^2$ stellar particles. The downward scatter is large, particularly for low-mass galaxies. The data points show observations of nearby galaxies from M. Relaño et al. (2020, 2022). These measurements are model-dependent and for galaxies from surveys that were not mass-selected. As the observed grain sizes are inferred from SED fitting, they are likely biased to dense gas in star-forming regions. In COLIBRE, such regions have lower dust grain size ratios than the medians shown here, which would further improve the agreement with the data.

can temporarily have very high X-ray luminosities. Such high luminosities of individual particles reflect the coarse-grained nature of subgrid models for feedback and are therefore probably artificial. In addition, they would likely be subtracted as point sources in observational analyses. For COLIBRE, core excision has only a small effect for $M_{500c} \gtrsim 10^{13} M_\odot$, but is very important for $M_{500c} \lesssim 10^{12.5} M_\odot$ ($M_* \lesssim 10^{11} M_\odot$). Without core excision, the mean X-ray luminosity would increase by $\approx 0.1 - 0.2$ dex for $M_{500c} \gtrsim 10^{13} M_\odot$, but by ≈ 1 dex for $M_{500c} \sim 10^{12} M_\odot$. For low masses, the spread in L_X also increases dramatically if the core region is included.

Fig. 24 shows that the CGM X-ray luminosity increases super-linearly with halo mass. The convergence between the different resolutions is very good. Comparison of the orange line, which shows the mean for L200m6, and the orange shading, which shows the 16th–84th percentile spread, reveals that for $M_{500c} \gtrsim 10^{12} M_\odot$ the mean is higher than the 84th percentile, which is caused by the large and non-Gaussian scatter in $L_{X,\text{CGM}}$ at low masses.

For $11.3 < \log_{10} M_{500c}/M_\odot < 13.7$, the predictions are compared with CGM observations of Y. Zhang et al. (2024), who stacked eROSITA data of $\sim 10^5$ central galaxies with spectroscopic redshifts $z < 0.2$ from SDSS and subtracted X-ray emission from the galaxies. The observed luminosities were scaled by $E^{-2}(z) = [H(z)/H_0]^{-2}$ to correct for evolution under the assumption of self-similarity. At higher masses, we compare with

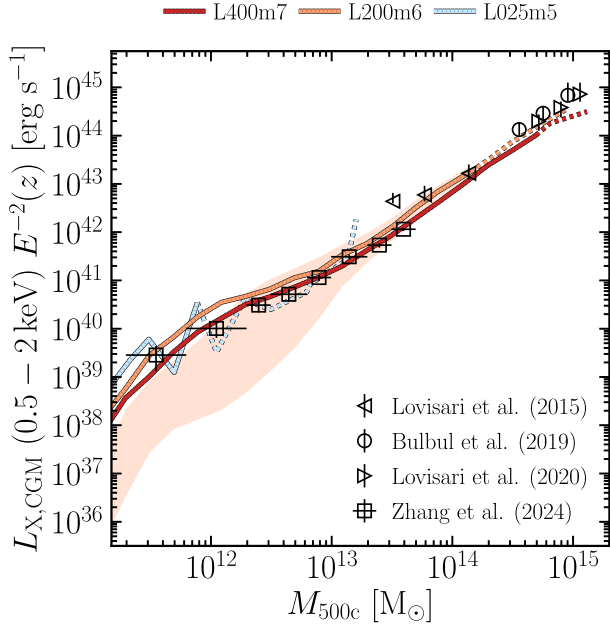


Figure 24. Mean X-ray luminosity from the CGM, i.e. integrated over the radial range $(0.15 - 1) \times R_{500c}$, as a function of halo mass, M_{500c} , at $z = 0$. The meanings of the line colours, line styles, and shading are as in Fig. 14, except that the lines indicate means rather than medians. For masses corresponding to galaxy clusters, $M_{500c} > 10^{13} M_{\odot}$, the medians (not plotted) are very close to the means, but for $M_{500c} \lesssim 10^{12} M_{\odot}$, the mean exceeds even the 84th percentile, which for L200m6 corresponds to the top of the shaded region. The data points show observations of Y. Zhang et al. (2024), who computed mean CGM luminosities by stacking eROSITA observations of $\sim 10^5$ central $z < 0.2$ galaxies, and with the median luminosities for the samples of individually detected clusters (without core excision) from L. Lovisari, T. H. Reiprich & G. Schellenberger (2015, $z < 0.035$), E. Bulbul et al. (2019, $0.2 < z < 0.66$), and L. Lovisari et al. (2020, $0.059 < z < 0.55$) compiled by J. Braspennig et al. (2024). All observed luminosities have been corrected for the expected self-similar evolution by dividing by $E^2(z) = [H(z)/H_0]^2$. The numerical convergence and the agreement with the data are both very good. Note that at the highest masses, the agreement would be further improved if we had matched the radial range (i.e. no core excision) and redshifts of the observed clusters.

the median luminosities for the samples of individually detected clusters from L. Lovisari et al. (2015, $z < 0.035$), E. Bulbul et al. (2019, $0.2 < z < 0.66$), and L. Lovisari et al. (2020, $0.059 < z < 0.55$) compiled by J. Braspennig et al. (2024). As described in J. Braspennig et al. (2024), the observed cluster X-ray luminosities have been shifted to the 0.5–2.0 keV band using PIMMS (K. Mukai 1993) and the observed cluster masses have been divided by 0.743 to correct for the hydrostatic mass bias that affects halo mass measurements based on X-ray observations. The cluster luminosities are not core-excised, but for cluster-sized objects core-excision has little impact on the COLIBRE predictions.

The simulations are in good agreement with the data. At the highest masses, $M_{500c} > 10^{14.5} M_{\odot}$, the luminosity is slightly underpredicted, but the discrepancy is only 0.3 dex and would be slightly smaller if we had not applied core-excision to the simulations. Matching the redshifts of the observed clusters instead of correcting the observed luminosities to $z = 0$ would likely further reduce the small discrepancy because the luminosity evolution is not precisely self-similar (J. Braspennig et al. 2024). However,

we caution that the very good agreement with the low-mass data may be somewhat fortuitous given the large predicted scatter and the importance of subtracting emission from the galaxies, which is done differently for the observations and the simulations.

8 DEPENDENCE ON THE IMPLEMENTATION OF AGN FEEDBACK

The comparison between the simulations and observations presented in the previous section was limited to the fiducial COLIBRE simulations, in which the AGN feedback is driven by the injection of thermal energy. Because the nature of AGN feedback is uncertain, and because AGN-powered jets are observed, the COLIBRE suite also includes simulations that use ‘hybrid AGN feedback’, i.e. both thermal isotropic and kinetic jet feedback. The model is detailed in F. Huško et al. (2026) and summarized in Sections 3.8 and 3.9. At each resolution, the largest simulation with hybrid AGN feedback was run in a box with a $2^3 = 8$ times smaller volume than the flagship fiducial simulation (see Table 2).

The hybrid AGN feedback simulations include a sophisticated subgrid model that tracks the evolution of the spin of every BH, which is then used to determine subgrid AGN feedback efficiencies and jet directions. Depending on the Eddington ratio, which is based on the rate of gas accretion onto the outer subgrid BH accretion disc, the accretion disc can be thick ($f_{\text{Edd}} < 0.01$), thin ($0.01 < f_{\text{Edd}} < 1$), or slim ($f_{\text{Edd}} > 1$). Jet feedback occurs in all three disc states, accretion disc winds are present in the thick and slim modes, but radiation pressure-driven winds are limited to the thin disc regime. Injection of thermal energy is used for both types of winds, employing the same implementation as the fiducial simulations. Some of the parameters of the CCSN and AGN feedback subgrid models were adjusted from their values in the fiducial models to maintain agreement with the observables used for the calibration, i.e. the $z \approx 0$ galaxy SMF, SMR, and BH masses at high stellar masses (see Section 3.10). The most important feedback mechanisms are thermal injections from the thin disc (at all z), jets from the thick disc (at $z \lesssim 1$), and jets from the thin disc (at $z \gtrsim 1$).

Figs 25 and 26 compare the fiducial (solid) and hybrid (dashed) AGN feedback simulations at m5 (blue), m6 (orange), and m7 (red) resolution in volumes of 25, 100, and 200 cMpc on a side, respectively. The panels show the same observables as were presented for the fiducial model in Section 7 in the order in which they appeared in that section. The data points show the same observations as were plotted in the figures in Section 7. To isolate the effects of the modelling of BHs and AGN feedback, the hybrid and fiducial AGN feedback simulations shown in these figures were started from the same ICs, which means that the latter use volumes smaller than presented in Section 7 for the same resolution.

The differences in the predictions of the two AGN models are typically very small for the observables investigated here. Consequently, the level of agreement with the data is very similar for the two models. For L025m5, the differences are least significant, so we will focus on the larger volumes available at the lower resolutions. In the hybrid model, BCG masses and half-mass radii are up to ≈ 0.15 dex larger and smaller, respectively, but the differences are significantly smaller for m7 than for m6 resolution. The hybrid model predicts up to ≈ 0.2 dex lower sSFRs for star-forming galaxies, and quenched fractions that are ≈ 20 percentage points higher for both $M_* \sim 10^{9.5} M_{\odot}$ at m7 resolution and $M_* \sim 10^{11.5} M_{\odot}$ at m6 resolution, where the differences are

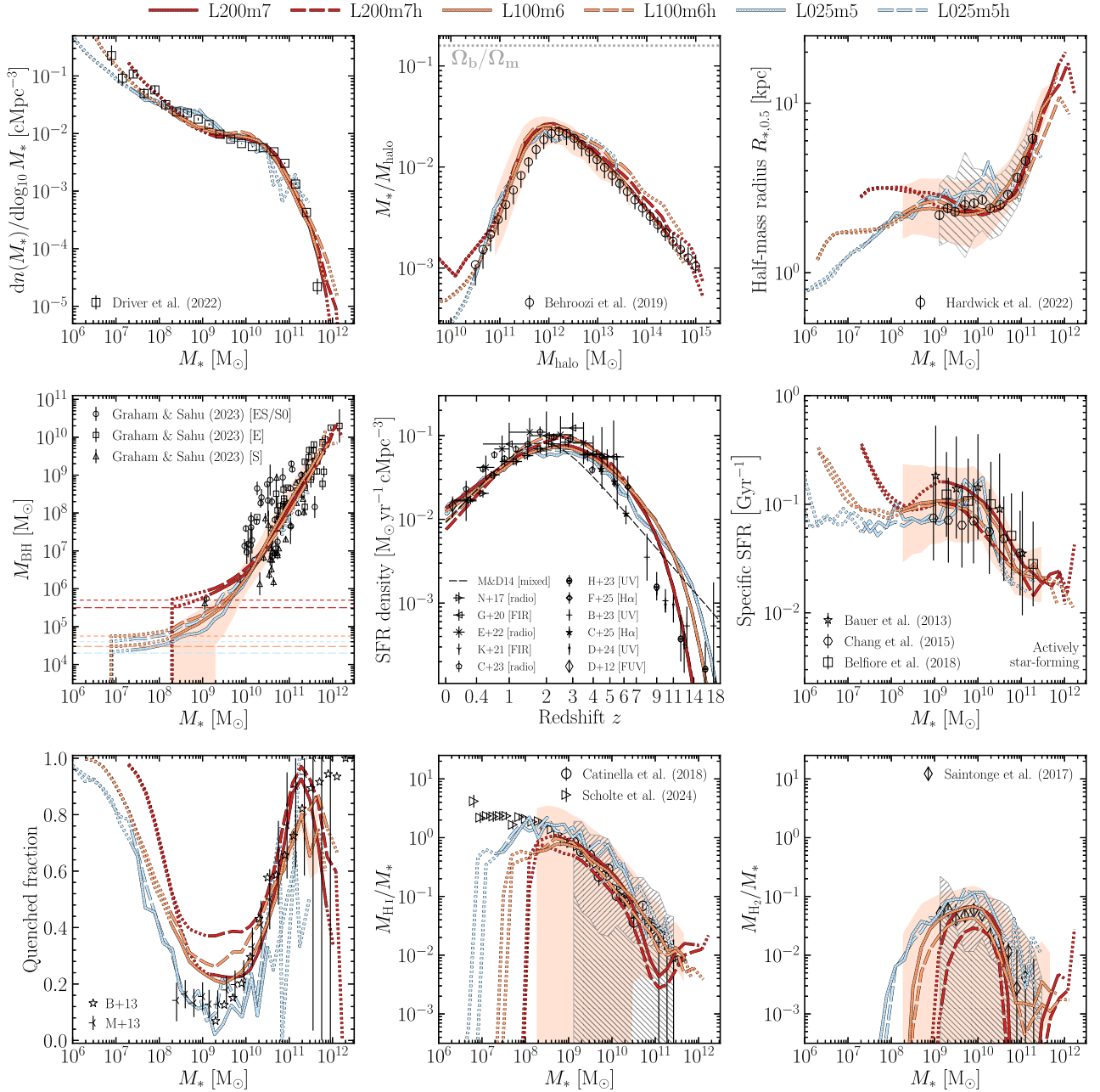


Figure 25. Comparison of the hybrid (dashed lines) and fiducial (solid lines) AGN feedback simulations. The panels show the same observables in the same order as presented for the fiducial simulations in Section 7. The data points show the same observations as in the figures shown there. Red, orange, and blue lines indicate fiducial and hybrid models for L200m7, L100m6, and L025m5, respectively. The volumes of the runs using fiducial AGN feedback are identical to those of the hybrid runs of the same resolution, but up to eight times smaller than for the fiducial simulations shown in Section 7. The lines become dotted where the stellar mass becomes smaller than 100 times the mean initial baryonic particle mass and at high mass where there are fewer than 10 objects per 0.2 dex mass bin. Where present, the orange shading indicates the 16th–84th percentile scatter for L100m6, except for the panel showing the quenched fraction, where it indicates the 1σ confidence interval for L100m6. The differences between the predictions of the two types of AGN feedback models are mostly small and hence their levels of agreement with the data are similarly good.

largest. Consistent with these differences, the H I, molecular and dust masses are somewhat smaller for hybrid AGN feedback. For the other $z = 0$ observables, the differences between the feedback models are smaller than the differences between the simulations of different resolution. The differences between hybrid and fiducial feedback generally have the same sign for both resolutions.

The $z = 0$ cosmic SFR density is ≈ 0.25 dex lower for the hybrid model, consistent with its lower sSFRs and higher quenched fractions. Interestingly, the peak of the cosmic SFR density, which occurs at $z \approx 2 - 3$, is higher for hybrid AGN feedback, though only by $\approx 0.1 - 0.2$ dex. This suggests that for other observables the sign of the differences may also change going from low to high redshift. At $z < 2$, the cosmic SFR density declines more steeply

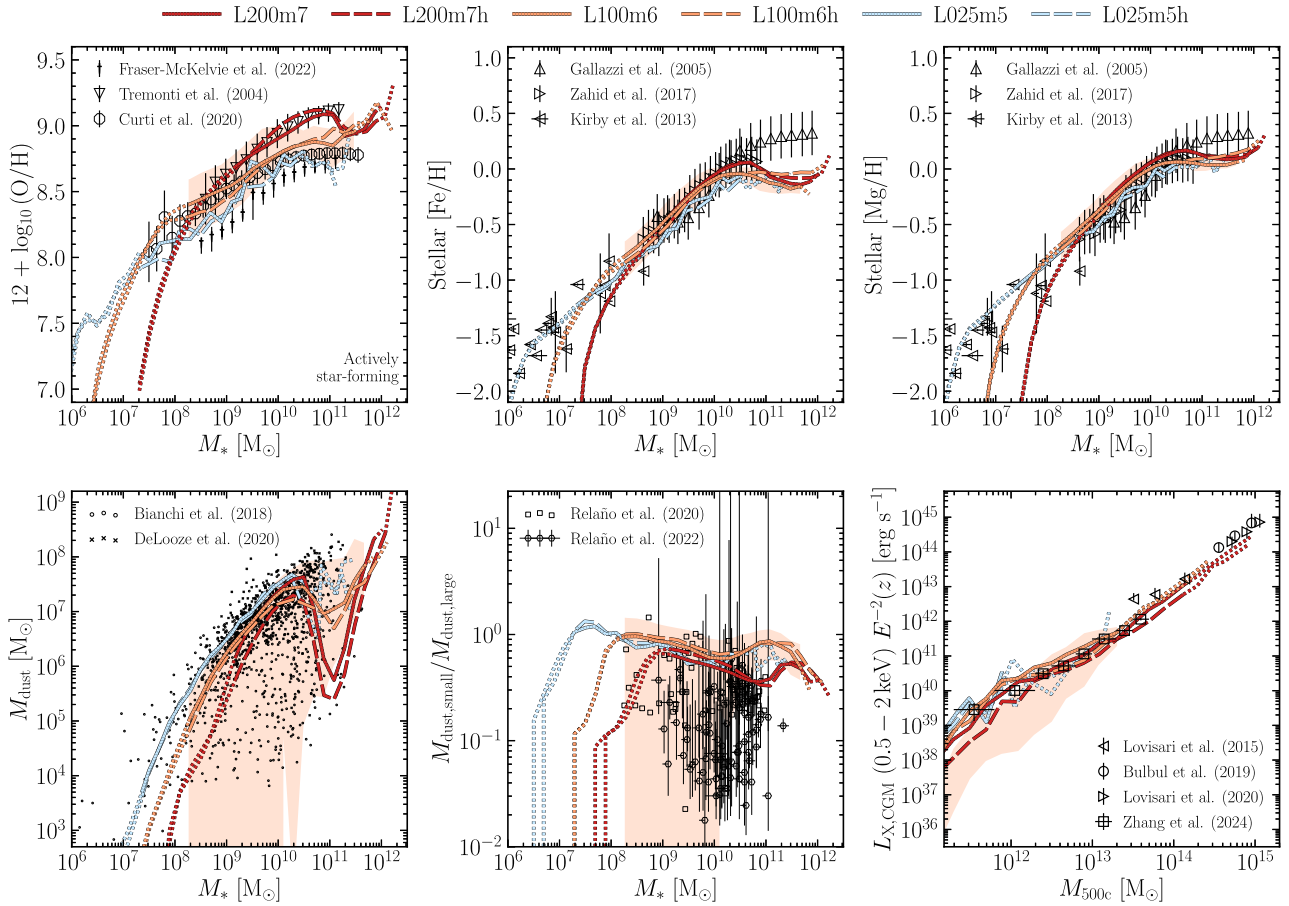


Figure 26. Continuation of Fig. 25.

in the simulation using hybrid AGN feedback. The SFR density drops by a full order of magnitude between its peak and $z = 0$ compared to a drop of ≈ 0.75 dex for the fiducial simulations.

Although the differences in $z \approx 0$ galaxy properties are relatively small, at least for the spatially integrated quantities investigated here, preliminary work shows that larger differences occur in the IGM and for BHs at high redshift. We conclude that, in terms of the wide variety of observables presented here, the hybrid AGN feedback simulations are similarly realistic as the fiducial models. The good agreement between the two types of simulations suggests that, provided the subgrid models for the feedback are calibrated to reproduce observables that constrain the efficiencies of the feedback on galaxy scales, in our case the $z \approx 0$ stellar and BH masses, self-regulated galaxy and BH growth ensure that differences in other basic observables probing similar scales will also tend to be small. However, this does not preclude the possibility of larger differences on sub- and intergalactic scales and at higher redshifts. F. Huško et al. (2026) and future work will provide more in-depth comparisons for a wider range of observables and redshifts.

9 SUMMARY AND DISCUSSION

We have introduced the COLIBRE galaxy formation model and a suite of cosmological hydrodynamical simulations that use it.

Compared to previous simulations run to $z = 0$ using similar baryonic resolution, and also compared to most higher-resolution

cosmological simulations, COLIBRE breaks new ground in several ways. This includes direct simulation of the multiphase ISM (rather than imposing an equation of state), direct simulation of the evolution of dust grains and their coupling to the chemistry and cooling, using four times more DM than baryonic particles in order to suppress spurious transfer of energy from DM to stars, repeating simulations using different models for BHs and AGN feedback that have been calibrated to the same data, and the sheer size of the runs ($5 \times 3008^3 \approx 1.4 \times 10^{11}$ particles; e.g. 20 times more than used in the EAGLE simulation).

The COLIBRE simulations span a factor of 64 in mass resolution, with particle masses (for both baryons and DM) of $\sim 10^5 M_\odot$ (m5), $10^6 M_\odot$ (m6), and $10^7 M_\odot$ (m7) in cubic volumes with comoving side lengths, L , of up to 100, 200, and 400 cMpc, respectively (see Table 2). The effective resolution is higher than for previous cosmological simulations with similar baryonic particle masses because the equation of state imposed on the ISM in those simulations artificially smooths the gas on sub-kpc scales (i.e. scales smaller than the Jeans scale in the warm ISM) even if the quoted force resolution or cell size is much smaller. The effective resolution is also higher because COLIBRE uses more DM particles per baryonic particle, by a factor of four compared to most previous simulations. The availability of different simulation resolutions and volumes, which were each calibrated to the $z = 0$ galaxy SMF, galaxy sizes, and BH masses in massive galaxies, allows testing for numerical convergence. The large number of resolution elements translates into relatively large volumes for

a given resolution, enabling the creation of statistically significant samples and the study of environmental effects, and yields a relatively large range of galaxy masses probed for a given volume. The mass range can be further expanded by combining simulations using different resolutions.

The simulations were run with the SWIFT code using the SPHENIX implementation of SPH. All subgrid prescriptions used by COLIBRE were newly developed for the project. COLIBRE includes models for radiative cooling, dust grains, star formation, stellar mass loss and chemical enrichment, turbulent diffusion, early stellar feedback, SN feedback, supermassive BHs, and AGN feedback.

In this work, we limit the comparison with observations to $z \approx 0$, except for the cosmic star formation history, which we find to be consistent with the data (Fig. 16; see also E. Chaikin et al. 2025b). We focus on the stellar, gas, metal, and dust content of galaxies, although we do compare with the observed relation between the X-ray luminosity of the CGM and halo mass, finding very good agreement (Fig. 24).

Radiative cooling rates are computed element-by-element using HYBRID-CHIMES (Section 3.1 and S. Ploekinger et al. 2025), which combines non-equilibrium abundances of hydrogen and helium species with tabulated equilibrium abundances for heavy elements. For self-consistency, the tables are computed using the same CHIMES reaction network (A. J. Richings et al. 2014a, b) used to compute hydrogen and helium rates on-the-fly. A novel aspect is that the radiative cooling rates for heavy elements are computed in ‘quasi-equilibrium’: they assume equilibrium ion/molecular fractions but account for the non-equilibrium free electrons from hydrogen and helium. The rates are computed in the presence of a redshift-dependent metagalactic radiation field and a pressure-dependent local ISRF (including cosmic rays) and account for local shielding by gas and dust. Another feature that is rare for cosmological simulations is that the reaction network is coupled to the dust grain module. The depletion of heavy elements, dust cooling and photoelectric heating, and the formation of molecules on dust grains all account for the evolving abundances of the different types of dust grains. The predicted $z = 0$ H I and H₂ masses as a function of stellar mass agree very well with the data, both for the medians and the scatter (Fig. 19).

The abundances of dust grains of three different chemical compositions (carbon and two types of silicates) and two different grain sizes (0.01 and 0.1 μ m) are tracked (Section 3.2 and J. W. Trayford et al. 2025). Dust masses can increase through seeding by AGB stars and CCSN, and depletion of heavy elements by gas accretion in the ISM. Collisions with hot gas and astration in stars decrease dust masses. Grain sizes evolve as a result of coagulation and shattering. The dust-to-metal ratio and the ratio of large-to-small grains are predicted to increase toward higher gas densities (Fig. 9). Visually, within the disc of a typical galaxy, the dust is concentrated towards the centre and in the spiral arms, with a spatial distribution intermediate between that of the atomic and molecular gas (Fig. 4). The relations between $z = 0$ dust mass and stellar mass (Fig. 22), and between grain size and stellar mass (Fig. 23) trace the observed trends.

Gravitationally unstable gas is stochastically converted into collisionless stellar particles according to a Schmidt law with an efficiency per free-fall time of 1 per cent (Section 3.3 and F. S. J. Nobels et al. 2024). Because stability is evaluated at the resolution limit, star formation shifts towards higher densities and lower temperatures as the resolution increases. At all resolutions, most stars form in the cold interstellar gas phase ($T \sim 10^2$ K) (Fig. 10),

but the warm phase ($T \sim 10^4$ K) contributes significantly to star formation in poorly resolved galaxies (Fig. C1).

Time- and metallicity-dependent stellar mass loss from AGB stars, massive stars, Type Ia SNe, and CCSNe is implemented using up-to-date nucleosynthetic yields for the 11 elements that dominate the cooling rates, as well from AGB stars for the s-process elements Ba and Sr, and from neutron star mergers, common-envelope jets SNe and collapsars for the r-process element Eu (Section 3.4 and C. Correa et al., in preparation). Diffusion driven by local turbulence is modelled for both gaseous elements and dust grains (Section 3.5 and C. Correa et al., in preparation). The predicted $z = 0$ relations between stellar mass and both gas and stellar metallicity agree well with the data, with the exception of a predicted turnover of the relations at $M_* \sim 10^{11} M_\odot$.

Before the first CCSNe explode, stellar particles start to exert feedback on the surrounding gas through stellar winds and radiation pressure, where the latter depends on the local shielding length and dust content. Furthermore, young stellar particles can heat and ionize gas by creating H II regions (Section 3.6 and A. Benítez-Llambay et al. 2025). However, the most important stellar feedback channel is CCSNe. Most (90 per cent) of the CCSN energy is used to stochastically heat nearby gas to high temperatures (C. Dalla Vecchia & J. Schaye 2012), but a fraction (10 per cent) is injected kinetically in the form of low-velocity (50 km s⁻¹) kicks (E. Chaikin et al. 2023). The heating temperature increases with the local gas density in order to meet the competing demands of good sampling, which prefers low temperatures, and overcoming numerical overcooling, which demands higher temperatures at higher densities (Section 3.7). While galactic outflows are driven mainly by the thermal channel, the kinetic channel helps drive turbulence (E. Chaikin et al. 2023). The probability of a heating event increases with the thermal gas pressure, which compensates for residual numerical overcooling and may also represent the physical effect of the reduction in thermal losses expected for more strongly clustered CCSNe (Section 3.10 and E. Chaikin et al. 2025a).

BHs are seeded in haloes identified by an FoF algorithm and then grow by accreting gas, assuming Bondi–Hoyle accretion limited by turbulence and vorticity, and by merging with other BHs (Section 3.8). AGN feedback is implemented as thermal energy injections (C. M. Booth & J. Schaye 2009) with a heating temperature that scales with the BH mass to maintain equal sampling of feedback events for BHs with different masses accreting at the same Eddington ratio (Section 3.9). To enable investigations of the sensitivity of predictions to the implementation and type of AGN feedback, a subset of the simulations use ‘hybrid AGN feedback’. These models track the BH spin. The hybrid model combines thermal AGN feedback, which is implemented as in the fiducial model, with kinetic jets directed along the BH spin axis, where the feedback mode depends on the BH accretion rate, mass, and spin (Section 3.8, Section 3.9 and F. Huško et al. 2026).

The subgrid models for CCSN and AGN feedback are calibrated to the $z = 0$ galaxy SMF and the galaxy SMR over the range $9.0 < \log_{10} M_*/M_\odot < 11.3$ (Section 3.10 and E. Chaikin et al. 2025a). For the fiducial, low-resolution simulations, this is accomplished by optimizing the seed BH mass and the values of three parameters of the subgrid model for CCSN feedback using Gaussian process emulators trained on modest-volume simulations (50 cMpc on a side; L050m7). In addition, the subgrid AGN

feedback efficiency is chosen to yield BH masses consistent with observations of massive $z = 0$ galaxies. For higher resolutions and for hybrid AGN feedback, small adjustments are made by hand. The simulations are in excellent agreement with the calibration data (Figs 12, 14, and 15).

The simulations reproduce the observed $z = 0$ relation between SFR and stellar mass for actively star-forming galaxies (Fig. 17). The fraction of passively evolving galaxies as a function of stellar mass matches the observed relation at intermediate masses (Fig. 18), but there are signs that the quenched fractions may be too high for galaxies with $M_* \lesssim 10^9 M_\odot$ and too low for the most massive galaxies, $M_* \sim 10^{12} M_\odot$. These disagreements either point to subtle shortcomings in the simulations, or they might reflect systematic biases in the comparison to the observations (e.g. for BCGs aperture choices or random errors in the stellar masses may be plausible explanations).

To first order, the two AGN feedback implementations yield very similar predictions for all the observables investigated here, and thus yield very similar levels of agreement with the observations. However, there are some small but significant differences. Depending on the resolution, hybrid AGN feedback gives a 0.1 – 0.2 dex higher peak cosmic SFR density, while the $z = 0$ SFRs are ≈ 0.25 dex lower. Hence, for hybrid AGN feedback, the cosmic SFR density declines more rapidly between $z \approx 2$ and 0. Consistent with the lower present-day SFR density, the $z = 0$ sSFRs and molecular masses are slightly lower, while the quenched fractions are somewhat higher than for the fiducial model. For hybrid AGN feedback, BCG masses and half-mass radii are up to ≈ 0.15 dex larger and smaller, respectively.

Comparison of the predictions of the simulations using different resolutions shows that convergence is generally excellent for objects sampled with more than $\sim 10 - 10^3$ stellar particles, where the limiting particle number depends on the observable. For galaxies of mass $M_* \sim 10^{11} M_\odot$, convergence is less good for quenched fractions and for median atomic, molecular, and dust masses. The relatively good convergence is notable, especially compared to simulations such as EAGLE (Fig. F1) and IllustrisTNG (Fig. F2), whose predictions generally also do not agree as well with the data for the properties investigated here. Like COLIBRE, EAGLE, but not IllustrisTNG, makes small adjustments to some of the subgrid parameters that control the feedback when the resolution is changed. However, unlike COLIBRE, EAGLE and IllustrisTNG impose an equation of state on the ISM. This makes the convergence seen in COLIBRE particularly remarkable, since COLIBRE does not resolve the internal structure of molecular clouds. We attribute this to the reduced importance of numerical overcooling in simulations that allow cooling to cold ISM temperatures, and hence enable more of the gas mass to concentrate in spiral arms and high-density clouds. The more inhomogeneous gas distribution results in a lower density for the intercloud medium, where most of the CCSN feedback events occur (Fig. 11), and hence smaller radiative losses. AGN feedback, on the other hand, is concentrated in galactic nuclei, where the gas density reaches very high values (Fig. 11). For that reason, the efficiency of AGN feedback is more sensitive to resolution, which is consistent with the poorer convergence for $M_* \sim 10^{11} M_\odot$.

Although we consider COLIBRE to be a major step forward compared to previous simulations of representative volumes, it has many known weaknesses (which are not unique to COLIBRE). Although the resolution suffices to distinguish the cold and warm interstellar gas phases, except in low-mass galaxies sampled with few particles, the simulations do not resolve the structure of molecular clouds. We therefore need to boost the

rates of gas accretion onto dust grains and of dust grain coagulation using a subgrid clumping factor. Because of resolution limitations, some star formation occurs in warm gas, particularly in low-mass galaxies. To compensate for numerical overcooling, we need to inject more CCSN energy in gas with a higher thermal pressure, use higher CCSN feedback temperatures in denser gas, and use higher AGN feedback temperatures for more massive BHs. A higher energy per feedback event implies poorer sampling. The parameters controlling stellar and AGN feedback require calibration, which removes some predictive power and which is more difficult to accomplish at higher resolution because of the greater computational expense and longer runtimes. We cannot resolve dynamical friction for all but the most massive BHs, which necessitates important but ad-hoc corrections to the BH positions. Although we implement prescriptions for self-shielding and a spatially varying ISRF, radiative transfer is not simulated directly, and we assume flash reionization. We do not vary the gravitational softening with density. There are no magnetic fields, and although cosmic rays are considered when calculating cooling rates and species fractions, they do not exert any pressure.

To our knowledge, both the level of numerical convergence and the level of agreement with a diverse range of galaxy data that we find for COLIBRE are unprecedented for cosmological hydrodynamical simulations. However, except for the cosmic star formation history, the comparisons presented here are limited to observations of galaxy scaling relations in the low- z Universe, and are performed ‘in theory space’. In future work, we will compare the simulations with a wider variety of observations, including spatially resolved galaxy properties and environmental dependencies, and extending to high redshift. We also plan to create virtual observations to enable a fairer comparison with the data and to test model-dependent aspects of observational analyses. We plan to complete the L050m5 and L025m5h simulations, to extend the suite to higher resolution and more model variations, and to complement the representative volumes with simulations that zoom-in on regions of interest.

ACKNOWLEDGEMENTS

We gratefully acknowledge discussions with researchers working on the analysis of the COLIBRE simulations, as well as discussions with and help from members of the Virgo and FLAMINGO collaborations. We specifically thank Joey Braspenning for help with the X-ray luminosities, Oliver Hahn for help with MONOFONIC. We thank Rob Crain and Joel Pfeffer for the addition of a passive subgrid model for globular clusters that we hope they will report on in future work. We are grateful to the late Richard Bower, Tom Theuns, and Bert Vandenbroucke for their contributions and enthusiasm in the early stages of the project. We thank Dylan Nelson and Annalisa Pillepich for comments on Appendix F. We thank the anonymous reviewer for helpful comments.

We acknowledge the use of the SWIFTSIMIO open source tools (J. Borrow & A. Borisov 2020). 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. This project has received funding from the Nether-

lands Organization for Scientific Research (NWO) through research programme Athena 184.034.002. SP acknowledges support from the Austrian Science Fund (FWF) through grant no. V 982-N. JT acknowledges support from an STFC Early Stage Research and Development grant (ST/X004651/1). ABL acknowledges support by the Italian Ministry for Universities (MUR) program ‘Dipartimenti di Eccellenza 2023-2027’ within the Centro Bicocca di Cosmologia Quantitativa (BiCoQ), and support by UNIMIB’s Fonuk redo Di Ateneo Quota Competitiva (project 2024-ATEQC-0050). CSF acknowledges support from European Research Council (ERC) Advanced Grant DMI-DAS (GA 786910). YMB acknowledges support from UK Research and Innovation through a Future Leaders Fellowship (grant agreement MR/X035166/1). This work was supported by the Swiss National Science Foundation (SNSF) under funding reference 200021_213076. AD is supported by an STFC doctoral studentship. AG acknowledges support from : FWO-Vlaanderen (Fund for Scientific Research Flanders) grant no.: FWO.3F0.2021.0030.01. ARJ and CGL acknowledge support from UKRI grants ST/T000244/1 and ST/X001075/1.

DATA AVAILABILITY

The data supporting the plots within this article are available on reasonable request to the corresponding author. The COLIBRE simulation data will eventually be made publicly available, although we note that the data volume (several petabytes) may prevent us from simply placing the raw data on a server. In the meantime, people interested in using the simulations are encouraged to contact the corresponding author. A public version of the SWIFT code (M. Schaller et al. 2024) is available at <http://www.swiftsim.com>. The COLIBRE modules implemented in SWIFT will be made publicly available after the public release of the simulation data.

REFERENCES

- Abbott T. M. C. et al., 2022, *Phys. Rev. D*, 105, 023520
- Accurso G. et al., 2017, *MNRAS*, 470, 4750
- Agertz O. et al., 2021, *MNRAS*, 503, 5826
- Ali-Haïmoud Y., Bird S., 2013, *MNRAS*, 428, 3375
- Angulo R. E., Pontzen A., 2016, *MNRAS*, 462, L1
- Aoyama S., Hou K.-C., Shimizu I., Hirashita H., Todoroki K., Choi J.-H., Nagamine K., 2017, *MNRAS*, 466, 105
- Applebaum E., Brooks A. M., Christensen C. R., Munshi F., Quinn T. R., Shen S., Tremmel M., 2021, *ApJ*, 906, 96
- Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, 47, 481
- Aver E., Berg D. A., Olive K. A., Pogge R. W., Salzer J. J., Skillman E. D., 2021, *J. Cosmol. Astropart. Phys.*, 2021, 027
- Bagla J. S., Ray S., 2003, *New Astron.*, 8, 665
- Bahé Y. M. et al., 2016, *MNRAS*, 456, 1115
- Bahé Y. M. et al., 2022, *MNRAS*, 516, 167
- Baker W. M. et al., 2025, *MNRAS*, 539, 557
- Barausse E., Morozova V., Rezzolla L., 2012, *ApJ*, 758, 63
- Barris B. J., Tonry J. L., 2006, *ApJ*, 637, 427
- Bate M. R., Burkert A., 1997, *MNRAS*, 288, 1060
- Bauer A. E. et al., 2013, *MNRAS*, 434, 209
- Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, *MNRAS*, 488, 3143
- Belfiore F. et al., 2018, *MNRAS*, 477, 3014
- Benítez-Llambay A. et al., 2025, *MNRAS*, 546, 268
- Bernardi M., Shankar F., Hyde J. B., Mei S., Marulli F., Sheth R. K., 2010, *MNRAS*, 404, 2087
- Bianchi S. et al., 2018, *A&A*, 620, A112
- Bigwood L., Bourne M. A., Irsic V., Amon A., Sijacki D., 2025, *MNRAS*, 542, 3206
- Bird S., Ni Y., Di Matteo T., Croft R., Feng Y., Chen N., 2022, *MNRAS*, 512, 3703
- Black J. H., 1987, in Hollenbach D. J., Thronson H. A., Jr., eds, *Proc. IAU Symp. 134, Interstellar Processes*. Springer, the Netherlands, p. 731
- Blanc G. et al., 2004, *A&A*, 423, 881
- Blandford R. D., Znajek R. L., 1977, *MNRAS*, 179, 433
- Bolatto A. D., Wolfire M., Leroy A. K., 2013, *ARA&A*, 51, 207
- Bondi H., 1952, *MNRAS*, 112, 195
- Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53
- Booth C. M., Schaye J., 2010, *MNRAS*, 405, L1
- Borrow J., Borrisov A., 2020, *J. Open Source Softw.*, 5, 2430
- Borrow J., Schaller M., Bower R. G., Schaye J., 2022, *MNRAS*, 511, 2367
- Borrow J., Schaller M., Bahé Y. M., Schaye J., Ludlow A. D., Ploekinger S., Nobels F. S. J., Altamura E., 2023, *MNRAS*, 526, 2441
- Botticella M. T. et al., 2008, *A&A*, 479, 49
- Bouwens R., Illingworth G., Oesch P., Stefanon M., Naidu R., van Leeuwen I., Magee D., 2023, *MNRAS*, 523, 1009
- Braspenning J., Schaye J., Borrow J., Schaller M., 2023, *MNRAS*, 523, 1280
- Braspenning J. et al., 2024, *MNRAS*, 533, 2656
- Bregman J. N., Harrington J. P., 1986, *ApJ*, 309, 833
- Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
- Buck T., Obreja A., Macciò A. V., Minchev I., Dutton A. A., Ostriker J. P., 2020, *MNRAS*, 491, 3461
- Buck T., Rybizki J., Buder S., Obreja A., Macciò A. V., Pfrommer C., Steinmetz M., Ness M., 2021, *MNRAS*, 508, 3365
- Buehlmann M., Winkler L., Hahn O., Helly J. C., Jenkins A., 2025, *Open J. Astrophys.*, 8, 32
- Bulbul E. et al., 2019, *ApJ*, 871, 50
- Cappellaro E., Evans R., Turatto M., 1999, *A&A*, 351, 459
- Cappellaro E. et al., 2015, *A&A*, 584, A62
- Carnall A. C. et al., 2024, *MNRAS*, 534, 325
- Catinella B. et al., 2018, *MNRAS*, 476, 875
- Chabrier G., 2003, *PASP*, 115, 763
- Chaikin E., Schaye J., Schaller M., Bahé Y. M., Nobels F. S. J., Ploekinger S., 2022, *MNRAS*, 514, 249
- Chaikin E., Schaye J., Schaller M., Benítez-Llambay A., Nobels F. S. J., Ploekinger S., 2023, *MNRAS*, 523, 3709
- Chaikin E. et al., 2025a, preprint ([arXiv:2509.04067](https://arxiv.org/abs/2509.04067))
- Chaikin E. et al., 2025b, preprint ([arXiv:2509.07960](https://arxiv.org/abs/2509.07960))
- Chandro-Gómez Á. et al., 2025, *MNRAS*, 539, 776
- Chang Y.-Y., van der Wel A., da Cunha E., Rix H.-W., 2015, *ApJS*, 219, 8
- Cheng H., Greengard L., Rokhlin V., 1999, *J. Comput. Phys.*, 155, 468
- Chevance M. et al., 2022, *MNRAS*, 509, 272
- Cinquegrana G. C., Karakas A. I., 2022, *MNRAS*, 510, 1557
- Cochrane R. K. et al., 2023, *MNRAS*, 523, 6082
- Conroy C., Gunn J. E., 2010, *ApJ*, 712, 833
- Conroy C., Gunn J. E., White M., 2009, *ApJ*, 699, 486
- Courant R., Friedrichs K., Lewy H., 1928, *Math. Ann.*, 100, 32
- Covelo-Paz A. et al., 2025, *A&A*, 694, A178
- Crain R. A., van de Voort F., 2023, *ARA&A*, 61, 473
- Crain R. A. et al., 2009, *MNRAS*, 399, 1773
- Crain R. A. et al., 2015, *MNRAS*, 450, 1937
- Curti M., Mannucci F., Cresci G., Maiolino R., 2020, *MNRAS*, 491, 944
- Dahlen T. et al., 2004, *ApJ*, 613, 189
- Dahlen T., Strolger L.-G., Riess A. G., 2008, *ApJ*, 681, 462
- Dalla Vecchia C., Schaye J., 2008, *MNRAS*, 387, 1431
- Dalla Vecchia C., Schaye J., 2012, *MNRAS*, 426, 140
- Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieeferantsoa M. H., Appleby S., 2019, *MNRAS*, 486, 2827
- De Looze I. et al., 2020, *MNRAS*, 496, 3668
- de Graaff A., Trayford J., Franx M., Schaller M., Schaye J., van der Wel A., 2022, *MNRAS*, 511, 2544
- de Salas P. F., Forero D. V., Gariazzo S., Martínez-Miravé P., Mena O., Ternes C. A., Tórtola M., Valle J. W. F., 2021, *J. High Energy Phys.*, 2021, 71
- Dehnen W., 2014, *Comput. Astrophys. Cosmol.*, 1, 1

- Dell’Aglì F., García-Hernández D. A., Schneider R., Ventura P., La Franca F., Valiante R., Marini E., Di Criscienzo M., 2017, *MNRAS*, 467, 4431
- Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, *ApJ*, 676, 33
- Dilday B. et al., 2010, *ApJ*, 713, 1026
- Doherty C. L., Gil-Pons P., Lau H. H. B., Lattanzio J. C., Siess L., 2014, *MNRAS*, 437, 195
- Dolag K. et al., 2025, preprint (arXiv:2504.01061)
- Donnan C. T. et al., 2024, *MNRAS*, 533, 3222
- Donnari M., Pillepich A., Nelson D., Marinacci F., Vogelsberger M., Hernquist L., 2021, *MNRAS*, 506, 4760
- Driver S. P. et al., 2012, *MNRAS*, 427, 3244
- Driver S. P. et al., 2022, *MNRAS*, 513, 439
- Dubois Y. et al., 2014, *MNRAS*, 444, 1453
- Dubois Y. et al., 2021, *A&A*, 651, A109
- Durier F., Dalla Vecchia C., 2012, *MNRAS*, 419, 465
- Dwek E., 1998, *ApJ*, 501, 643
- Eke V. R., Cole S., Frenk C. S., 1996, *MNRAS*, 282, 263
- Elbers W., Frenk C. S., Jenkins A., Li B., Pascoli S., 2021, *MNRAS*, 507, 2614
- Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, *Publ. Astron. Soc. Aust.*, 34, e058
- Engler C. et al., 2021, *MNRAS*, 500, 3957
- Enia A. et al., 2022, *ApJ*, 927, 204
- Esteban I., Gonzalez-Garcia M. C., Maltoni M., Schwetz T., Zhou A., 2020, *J. High Energy Phys.*, 2020, 178
- Faucher-Giguère C.-A., 2020, *MNRAS*, 493, 1614
- Faucher-Giguère C.-A., Oh S. P., 2023, *ARA&A*, 61, 131
- Feldmann R., Bieri R., 2025, preprint (arXiv:2507.08925)
- Feldmann R. et al., 2023, *MNRAS*, 522, 3831
- Ferland G. J. et al., 2017, *Rev. Mex. Astron. Astrofis.*, 53, 385
- Fishlock C. K., Karakas A. I., Lugaro M., Yong D., 2014, *ApJ*, 797, 44
- Font A. S. et al., 2020, *MNRAS*, 498, 1765
- Forouhar Moreno V. J., Helly J., McGibbon R., Schaye J., Schaller M., Han J., Kugel R., Bahé Y. M., 2025, *MNRAS*, 543, 1339
- Fraser-McKelvie A. et al., 2022, *MNRAS*, 510, 320
- Frohmaier C. et al., 2019, *MNRAS*, 486, 2308
- Fu S. et al., 2025, *ApJ*, 987, 186
- Fuente A. et al., 2019, *A&A*, 624, A105
- Furlong M. et al., 2015, *MNRAS*, 450, 4486
- Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, *MNRAS*, 362, 41
- Genina A., Springel V., Rantala A., 2024, *MNRAS*, 534, 957
- Gnedin N. Y., Kravtsov A. V., 2011, *ApJ*, 728, 88
- Graham A. W., Sahu N., 2023, *MNRAS*, 518, 2177
- Graham A. W., Sahu N., 2024, *MNRAS*, 530, 3429
- Granato G. L. et al., 2021, *MNRAS*, 503, 511
- Grand R. J. J. et al., 2017, *MNRAS*, 467, 179
- Graur O. et al., 2011, *MNRAS*, 417, 916
- Greengard L., Rokhlin V., 1987, *J. Comput. Phys.*, 73, 325
- Greif T. H., Glover S. C. O., Bromm V., Klessen R. S., 2009, *MNRAS*, 392, 1381
- Gruppioni C. et al., 2020, *A&A*, 643, A8
- Gunawardhana M. L. P. et al., 2011, *MNRAS*, 415, 1647
- Gutcke T. A., Pakmor R., Naab T., Springel V., 2021, *MNRAS*, 501, 5597
- Gutcke T. A., Pakmor R., Naab T., Springel V., 2022, *MNRAS*, 513, 1372
- Hahn O., Abel T., 2011, *MNRAS*, 415, 2101
- Hahn O., Michaux M., Rampf C., Uhlemann C., Angulo R. E., 2020, *Astrophysics Source Code Library, record ascl:2008.024*
- Hahn O., Rampf C., Uhlemann C., 2021, *MNRAS*, 503, 426
- Han J., Jing Y. P., Wang H., Wang W., 2012, *MNRAS*, 427, 2437
- Han J., Cole S., Frenk C. S., Benítez-Llambay A., Helly J., 2018, *MNRAS*, 474, 604
- Hardin D. et al., 2000, *A&A*, 362, 419
- Hardwick J. A., Cortese L., Obreschkow D., Catinella B., Cook R. H. W., 2022, *MNRAS*, 509, 3751
- Harikane Y. et al., 2023, *ApJS*, 265, 5
- Hirashita H., Voshchinnikov N. V., 2014, *MNRAS*, 437, 1636
- Hirashita H., Yan H., 2009, *MNRAS*, 394, 1061
- Hislop J. M., Naab T., Steinwandel U. P., Lahén N., Irodotou D., Johansson P. H., Walch S., 2022, *MNRAS*, 509, 5938
- Hopkins P. F., 2015, *MNRAS*, 450, 53
- Hopkins P. F. et al., 2018, *MNRAS*, 480, 800
- Hopkins P. F. et al., 2023, *MNRAS*, 519, 3154
- Horesh A., Poznanski D., Ofek E. O., Maoz D., 2008, *MNRAS*, 389, 1871
- Hoyle F., Lyttleton R. A., 1939, *Proc. Camb. Philos. Soc.*, 35, 405
- Hu C.-Y., Naab T., Glover S. C. O., Walch S., Clark P. C., 2017, *MNRAS*, 471, 2151
- Hu Z., Krumholz M. R., Pokhrel R., Gutermuth R. A., 2022, *MNRAS*, 511, 1431
- Hui L., Gnedin N. Y., 1997, *MNRAS*, 292, 27
- Huško F., Lacey C. G., Schaye J., Schaller M., Nobels F. S. J., 2022, *MNRAS*, 516, 3750
- Huško F., Lacey C. G., Schaye J., Nobels F. S. J., Schaller M., 2024, *MNRAS*, 527, 5988
- Huško F., Lacey C. G., Roper W. J., Schaye J., Briggs J. M., Schaller M., 2025, *MNRAS*, 537, 2559
- Huško F. et al., 2026, *MNRAS*, 547, stag324
- Indriolo N. et al., 2015, *ApJ*, 800, 40
- Iwamoto K., Brachwitz F., Nomoto K., Kishimoto N., Umeda H., Hix W. R., Thielemann F.-K., 1999, *ApJS*, 125, 439
- Izquierdo-Villalba D., Lupi A., Regan J., Bonetti M., Franchini A., 2023, preprint (arXiv:2311.03152)
- Jenkins E. B., 2009, *ApJ*, 700, 1299
- Jenkins A., 2013, *MNRAS*, 434, 2094
- Jönsson H. et al., 2018, *AJ*, 156, 126
- Karakas A. I., 2010, *MNRAS*, 403, 1413
- Karakas A. I., Lugaro M., 2016, *ApJ*, 825, 26
- Katsianis A. et al., 2020, *MNRAS*, 492, 5592
- Katz N., Weinberg D. H., Hernquist L., 1996, *ApJS*, 105, 19
- Katz H. et al., 2022, preprint (arXiv:2211.04626)
- Kaviraj S., Lazar I., Watkins A. E., Laigle C., Martin G., Jackson R. A., 2025, *MNRAS*, 538, 153
- Keller B. W., Wadsley J., Benincasa S. M., Couchman H. M. P., 2014, *MNRAS*, 442, 3013
- Kennicutt R. C., Jr., 1998, *ApJ*, 498, 541
- Kennicutt R. C., Evans N. J., 2012, *ARA&A*, 50, 531
- Khusanova Y. et al., 2021, *A&A*, 649, A152
- Kim C.-G., Kim J.-G., Gong M., Ostriker E. C., 2023, *ApJ*, 946, 3
- Kirby E. N., Cohen J. G., Guhathakurta P., Cheng L., Bullock J. S., Gallazzi A., 2013, *ApJ*, 779, 102
- Kobayashi C., Umeda H., Nomoto K., Tominaga N., Ohkubo T., 2006, *ApJ*, 653, 1145
- Kobayashi C., Karakas A. I., Lugaro M., 2020, *ApJ*, 900, 179
- Kroupa P., 2002, *Science*, 295, 82
- Krumholz M. R., McKee C. F., Klein R. I., 2005, *ApJ*, 618, 757
- Krumholz M. R., McKee C. F., Klein R. I., 2006, *ApJ*, 638, 369
- Kugel R. et al., 2023, *MNRAS*, 526, 6103
- Kumar A. et al., 2025, *A&A*, 698, A236
- Kuznetsova N. et al., 2008, *ApJ*, 673, 981
- Lagos C. d. P. et al., 2025, preprint (arXiv:2512.11309)
- Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, *MNRAS*, 441, 1270
- Leja J. et al., 2022, *ApJ*, 936, 165
- Leroy A. K. et al., 2025, *ApJ*, 985, 14
- Lesgourgues J., 2011, preprint (arXiv:1104.2932)
- Lesgourgues J., Tram T., 2011, *J. Cosmol. Astropart. Phys.*, 09, 032
- Leung S.-C., Nomoto K., 2018, *ApJ*, 861, 143
- Li W., Chornock R., Leaman J., Filippenko A. V., Poznanski D., Wang X., Ganeshalingam M., Mannucci F., 2011, *MNRAS*, 412, 1473
- Liang J., Gjergo E., Fan X., 2023, *MNRAS*, 522, 863
- Libeskind N. I. et al., 2020, *MNRAS*, 498, 2968
- Lovisari L., Reiprich T. H., Schellenberger G., 2015, *A&A*, 573, A118
- Lovisari L. et al., 2020, *ApJ*, 892, 102
- Ludlow A. D., Schaye J., Schaller M., Richings J., 2019a, *MNRAS*, 488, L123

- Ludlow A. D., Schaye J., Bower R., 2019b, *MNRAS*, 488, 3663
- Ludlow A. D., Schaye J., Schaller M., Bower R., 2020, *MNRAS*, 493, 2926
- Ludlow A. D., Fall S. M., Schaye J., Obreschkow D., 2021, *MNRAS*, 508, 5114
- Ludlow A. D., Fall S. M., Wilkinson M. J., Schaye J., Obreschkow D., 2023, *MNRAS*, 525, 5614
- Lupton R., Blanton M. R., Fekete G., Hogg D. W., O'Mullane W., Szalay A., Wherry N., 2004, *PASP*, 116, 133
- Ma L., Hopkins P. F., Ma X., Anglés-Alcázar D., Faucher-Giguère C.-A., Kelley L. Z., 2021, *MNRAS*, 508, 1973
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Madgwick D. S., Hewett P. C., Mortlock D. J., Wang L., 2003, *ApJ*, 599, L33
- Mannucci F., Della Valle M., Panagia N., Cappellaro E., Cresci G., Maiolino R., Petrosian A., Turatto M., 2005, *A&A*, 433, 807
- Marri S., White S. D. M., 2003, *MNRAS*, 345, 561
- Martínez-Serrano F. J., Serna A., Domínguez-Tenreiro R., Mollá M., 2008, *MNRAS*, 388, 39
- Mathis J. S., Mezger P. G., Panagia N., 1983, *A&A*, 128, 212
- Mazzali P. A., McFadyen A. I., Woosley S. E., Pian E., Tanaka M., 2014, *MNRAS*, 443, 67
- McAlpine S. et al., 2016, *Astron. Comput.*, 15, 72
- McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, *MNRAS*, 465, 2936
- McCarthy I. G., Bird S., Schaye J., Harnois-Deraps J., Font A. S., van Waerbeke L., 2018, *MNRAS*, 476, 2999
- McGibbon R., Helly J., Schaye J., Schaller M., Vandenbroucke B., 2025, *J. Open Source Softw.*, 10, 8252
- McKinnon R., Torrey P., Vogelsberger M., 2016, *MNRAS*, 457, 3775
- Melinder J. et al., 2012, *A&A*, 545, A96
- Michaux M., Hahn O., Rampf C., Angulo R. E., 2021, *MNRAS*, 500, 663
- Mina M., Shen S., Keller B. W., Mayer L., Madau P., Wadsley J., 2021, *A&A*, 655, A22
- More S., Kravtsov A. V., Dalal N., Gottlöber S., 2011, *ApJS*, 195, 4
- Mukai K., 1993, *Legacy*, 3, 21
- Muzzin A. et al., 2013, *ApJ*, 777, 18
- Narayan R., Igumenshchev I. V., Abramowicz M. A., 2003, *PASJ*, 55, L69
- Neill J. D. et al., 2006, *AJ*, 132, 1126
- Nelson D. et al., 2019a, *Comput. Astrophys. Cosmol.*, 6, 2
- Nelson D. et al., 2019b, *MNRAS*, 490, 3234
- Neumayer N., Seth A., Böker T., 2020, *A&AR*, 28, 4
- Nobels F. S. J., Schaye J., Schaller M., Bahé Y. M., Chaikin E., 2022, *MNRAS*, 515, 4838
- Nobels F. S. J., Schaye J., Schaller M., Ploekinger S., Chaikin E., Richings A. J., 2024, *MNRAS*, 532, 3299
- Nomoto K., Kobayashi C., Tominaga N., 2013, *ARA&A*, 51, 457
- Novak M. et al., 2017, *A&A*, 602, A5
- Novikov I. D., Thorne K. S., 1973, in Dewitt C., Dewitt B. S., eds, *Black Holes (Les Astres Occlus)*. Gordon and Breach, New York, p. 343
- Ochsendorf B. B., Meixner M., Roman-Duval J., Rahman M., Evans II N. J., 2017, *ApJ*, 841, 109
- Oku Y., Nagamine K., 2024, *ApJ*, 975, 183
- Okumura J. E. et al., 2014, *PASJ*, 66, 49
- Pain R. et al., 2002, *ApJ*, 577, 120
- Pakmor R. et al., 2023, *MNRAS*, 524, 2539
- Peng Y.-j. et al., 2010, *ApJ*, 721, 193
- Perrett K. et al., 2012, *AJ*, 144, 59
- Pettini M., Bowen D. V., 2001, *ApJ*, 560, 41
- Pillepich A. et al., 2018a, *MNRAS*, 473, 4077
- Pillepich A. et al., 2018b, *MNRAS*, 475, 648
- Ploekinger S., Schaye J., 2020, *MNRAS*, 497, 4857
- Ploekinger S., Nobels F. S. J., Schaller M., Schaye J., 2024, *MNRAS*, 528, 2930
- Ploekinger S., Richings A. J., Schaye J., Trayford J. W., Schaller M., Chaikin E., 2025, *MNRAS*, 543, 891
- Pokhrel R. et al., 2021, *ApJ*, 912, L19
- Portinari L., Chiosi C., Bressan A., 1998, *A&A*, 334, 505
- Poznanski D. et al., 2007, *MNRAS*, 382, 1169
- Puchwein E., Bolton J. S., Haehnelt M. G., Madau P., Becker G. D., Haardt F., 2015, *MNRAS*, 450, 4081
- Quimby R. M., Yuan F., Akerlof C., Wheeler J. C., Warren M. S., 2012, *AJ*, 144, 177
- Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, *MNRAS*, 430, 2427
- Regan J., Volonteri M., 2024, *Open J. Astrophys.*, 7, 72
- Relaño M., Lisenfeld U., Hou K. C., De Looze I., Vilchez J. M., Kennicutt R. C., 2020, *A&A*, 636, A18
- Relaño M. et al., 2022, *MNRAS*, 515, 5306
- Rey M. P. et al., 2025, *MNRAS*, 541, 1195
- Richings A. J., Schaye J., Oppenheimer B. D., 2014a, *MNRAS*, 440, 3349
- Richings A. J., Schaye J., Oppenheimer B. D., 2014b, *MNRAS*, 442, 2780
- Richings A. J., Faucher-Giguère C.-A., Gurvich A. B., Schaye J., Hayward C. C., 2022, *MNRAS*, 517, 1557
- Robertson B. E., Kravtsov A. V., 2008, *ApJ*, 680, 1083
- Rodney S. A., Tonry J. L., 2010, *ApJ*, 723, 47
- Rodney S. A. et al., 2014, *AJ*, 148, 13
- Romano D., Karakas A. I., Tosi M., Matteucci F., 2010, *A&A*, 522, A32
- Rosas-Guevara Y. M. et al., 2015, *MNRAS*, 454, 1038
- Rosdahl J., Schaye J., Teyssier R., Agertz O., 2015, *MNRAS*, 451, 34
- Ruffert M., Arnett D., 1994, *ApJ*, 427, 351
- Saintonge A. et al., 2017, *ApJS*, 233, 22
- Sales L. V., Navarro J. F., Schaye J., Dalla Vecchia C., Springel V., Booth C. M., 2010, *MNRAS*, 409, 1541
- Sawala T. et al., 2016, *MNRAS*, 457, 1931
- Scannapieco C. et al., 2012, *MNRAS*, 423, 1726
- Schaller M. et al., 2015a, *MNRAS*, 451, 1247
- Schaller M., Dalla Vecchia C., Schaye J., Bower R. G., Theuns T., Crain R. A., Furlong M., McCarthy I. G., 2015b, *MNRAS*, 454, 2277
- Schaller M. et al., 2024, *MNRAS*, 530, 2378
- Schaye J., 2001, *ApJ*, 562, L95
- Schaye J., 2004, *ApJ*, 609, 667
- Schaye J., Dalla Vecchia C., 2008, *MNRAS*, 383, 1210
- Schaye J. et al., 2010, *MNRAS*, 402, 1536
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Schaye J. et al., 2023, *MNRAS*, 526, 4978
- Schmidt M., 1959, *ApJ*, 129, 243
- Scholte D. et al., 2024, *MNRAS*, 535, 2341
- Segers M. C., Crain R. A., Schaye J., Bower R. G., Furlong M., Schaller M., Theuns T., 2016a, *MNRAS*, 456, 1235
- Segers M. C., Schaye J., Bower R. G., Crain R. A., Schaller M., Theuns T., 2016b, *MNRAS*, 461, L102
- Sellwood J. A., 2013, *ApJ*, 769, L24
- Sellwood J. A., 2015, *MNRAS*, 453, 2919
- Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2017, *ApJ*, 845, 133
- Shakura N. I., Sunyaev R. A., 1973, *A&A*, 24, 337
- Sharma P., Roy A., Nath B. B., Shchekinov Y., 2014, *MNRAS*, 443, 3463
- Shen S., Wadsley J., Stinson G., 2010, *MNRAS*, 407, 1581
- Shuntov M. et al., 2025, *A&A*, 695, A20
- Smith M. C., Bryan G. L., Somerville R. S., Hu C.-Y., Teyssier R., Burkhard B., Hernquist L., 2021, *MNRAS*, 506, 3882
- Springel V., Hernquist L., 2003, *MNRAS*, 339, 289
- Springel V., Di Matteo T., Hernquist L., 2005, *MNRAS*, 361, 776
- Stanway E. R., Eldridge J. J., 2018, *MNRAS*, 479, 75
- Steinwandel U. P., Bryan G. L., Somerville R. S., Hayward C. C., Burkhard B., 2023, *MNRAS*, 526, 1408
- Strolger L.-G., 2003, PhD thesis, University of Michigan
- Teyssier R., 2002, *A&A*, 385, 337
- Thacker R. J., Couchman H. M. P., 2000, *ApJ*, 545, 728
- Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, *MNRAS*, 404, 1775
- Thorne J. E. et al., 2021, *MNRAS*, 505, 540
- Tonry J. L. et al., 2003, *ApJ*, 594, 1
- Trayford J. W. et al., 2026, *MNRAS*, 545, staf2040
- Tremmel M., Governato F., Volonteri M., Quinn T. R., 2015, *MNRAS*, 451, 1868
- Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, *MNRAS*, 470, 1121

- Tremonti C. A. et al., 2004, *ApJ*, 613, 898
 Tsai J. C., Mathews W. G., 1995, *ApJ*, 448, 84
 Utomo D. et al., 2018, *ApJ*, 861, L18
 Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, *Nat. Rev. Phys.*, 2, 42
 Wendland H., 1995, *Adv. Comput. Math.*, 4, 389
 Wiersma R. P. C., Schaye J., Smith B. D., 2009a, *MNRAS*, 393, 99
 Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, *MNRAS*, 399, 574
 Wilkinson M. J., Ludlow A. D., Lagos C. d. P., Fall S. M., Schaye J., Obreschkow D., 2023, *MNRAS*, 519, 5942
 Yates R. M., Henriques B. M. B., Fu J., Kauffmann G., Thomas P. A., Guo Q., White S. D. M., Schady P., 2021, *MNRAS*, 503, 4474
 Zahid H. J., Kudritzki R.-P., Conroy C., Andrews B., Ho I. T., 2017, *ApJ*, 847, 18
 Zhang Y. et al., 2024, *A&A*, 690, A268
 Zhukovska S., Gail H. P., Tieloff M., 2008, *A&A*, 479, 453
 Zolotov A. et al., 2015, *MNRAS*, 450, 2327

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

output_list.txt

output_properties.pdf

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: POSSIBLE SUPPRESSION OF GRAVITATIONAL INSTABILITY DUE TO GRAVITATIONAL SOFTENING

For all COLIBRE resolutions, gravity is softened for baryonic densities $n_{\text{H}} > n_{\text{H},\epsilon} \equiv m_{\text{g}}X/(m_{\text{H}}\epsilon^3) \approx 0.2 \text{ cm}^{-3}$ at $z < 1.57$ and at densities a factor $[(1+z)/2.57]^3$ higher than $n_{\text{H},\epsilon}$ at $z \geq 1.57$. As discussed and demonstrated by S. Ploekinger et al. (2024) and F. S. J. Nobels et al. (2024), for these high densities, softening does not necessarily eliminate gravitational instability on scales greater than the SPH kernel, because the gravitational force is not zero below the softening scale and because such dense gas tends to be cold ($T \ll 10^4 \text{ K}$), which facilitates gravitational instability.

Assuming that turbulence is subsonic, gas clouds of mass equal to the SPH kernel mass are physically gravitationally unstable (i.e. the kernel mass exceeds the thermal Jeans mass) below the dotted contours in the temperature–density plots shown in Fig. A1. If there were turbulence, then the contours would truncate below the density for which the thermal velocity dispersion drops below the turbulent value (e.g. fig. 1 of F. S. J. Nobels et al. 2024). If the turbulence is supersonic, which is commonly the case for the cold ISM ($T \ll 10^4 \text{ K}$), then there is thus a minimum density for gravitational instability at the mass resolution limit. The different line colours show different numerical resolutions, where orange corresponds to m6, which is the resolution of the simulation whose mass distribution is indicated by the colour image. The left and right panels show redshifts $z = 0$ and 4, respectively. Because the kernel mass decreases with increasing resolution, higher densities and/or lower temperatures are required for gravitational instability at higher resolution. Above the solid contours (kernel

mass equals the softened Jeans mass; see S. Ploekinger et al. 2024), which bend down above $n_{\text{H},\epsilon}$, gas clouds with mass equal to the kernel mass are gravitationally stable in the simulation. Gas clouds in the region in between the solid and dotted contours should be unstable but are stabilized by softening, at least for a cloud mass equal to the kernel mass. Higher mass clouds can still be unstable.

Fig. A1 shows that at both redshifts there is little gas in the potentially problematic region between the solid and dotted contours. Note that gas in this region is eligible to form stars because the star formation criterion does not account for softening. Hence, star formation in this region is only slowed down indirectly by the suppression of gravitational instability in low-mass clouds. Even if the gas is unstable in the simulation, i.e. below the solid contours, if $n_{\text{H}} > n_{\text{H},\epsilon}$, then softening may slow gravitational collapse in the regime where the dynamical time is small compared to the sound-crossing time. Here, the time-scale for gravitational collapse is the softened free-fall time, instead of the classical Newtonian free-fall time (S. Ploekinger et al. 2024). If the weight of overlying gas layers pushes the gas to sufficiently high densities, then gas clouds that are stabilized by softening may still become unstable, but such layers will only be resolved if the galaxy contains a sufficient number of particles. The problem of suppression of gravitational instability due to softening is thus more likely to occur in lower mass galaxies because they are resolved with fewer particles. Another reason is that lower mass galaxies also tend to have gas with lower metallicities, which tends to be hotter.

If we increase the resolution, then the contours in Fig. A1 shift down, and higher gas densities are required for star formation (see equation 6). Hence, for resolutions much higher than those considered here, the slowdown of gravitational collapse by softening could become a bottleneck. This may result in excessively long gas consumption time-scales and hence too much mass in the ISM. If softening were to cause problems, then we would expect to see signs first in the lowest mass galaxies in the highest resolution runs. Resolving such issues would require switching to adaptive (i.e. density-dependent) gravitational softening,²⁸ which, however, would require recalibration of subgrid models and would, if it results in smaller softening lengths for collisionless particles, exacerbate spurious energy transfer due to discreteness effects. Future work, particularly with simulations using higher resolution than used for the COLIBRE simulations presented here, could benefit from a switch to adaptive softening, at least for gas particles.

²⁸Note that adaptive softening can only help if the softening length is allowed to become sufficiently small.

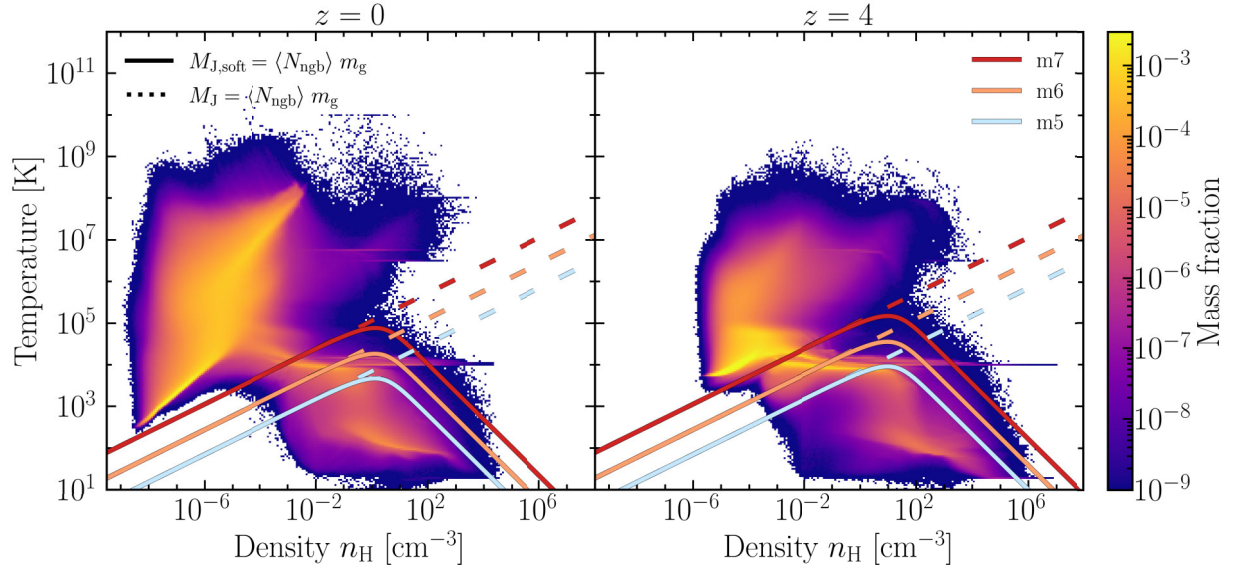


Figure A1. Temperature–density plots for the gas in the L200m6 simulation at $z = 0$ (left panel) and $z = 4$ (right panel). The colour encodes the fraction of the gas mass per pixel. Gas below the dotted (solid) contours is gravitationally unstable at the resolution limit in Newtonian (softened) gravity, if the turbulent velocity dispersion is negligible. The line colour indicates the numerical resolution of the simulation, with orange corresponding to the resolution of the simulation whose density–temperature distribution is plotted (m6). Gas clouds with mass equal to the kernel mass that fall in between the solid and dotted contours of the same colour are stabilized by softening, but this is not necessarily true for higher mass clouds.

APPENDIX B: THE LINEAR PHASES FOR THE INITIAL CONDITIONS

The Gaussian phases for the ICs of all simulation volumes use an updated version of the PANPHASIA hierarchical Gaussian white noise field introduced by A. Jenkins (2013). We used the same version as the FLAMINGO simulations (see appendix B of J. Schaye et al. 2023). The text descriptors in Table B1 specify the initial

Gaussian phases for the different simulation volumes presented in this work. The MONOFONIC code (O. Hahn et al. 2020; M. Michaux et al. 2021) can be used to create uniform mass resolution ICs using these text descriptors. The same descriptors can be read by the MUSIC2 code (O. Hahn & T. Abel 2011; M. Buehlmann et al. 2025) to generate zoom ICs for regions or objects within the COLIBRE volumes.

Table B1. PANPHASIA descriptors for different box sizes. Simulation variations in the same volume use the same descriptor.

L (cMpc)	Descriptor
25	[Panph6,L20,(967096,215599,663706),S1,KK1025,CH3222183807,COLIBRE025]
50	[Panph6,L20,(235287,445214,422255),S1,KK1025,CH2001508634,COLIBRE050]
100	[Panph6,L19,(266507,351694,328156),S1,KK1025,CH1356507509,COLIBRE100]
200	[Panph6,L19,(218575,323145,460872),S1,KK1025,CH2306922479,COLIBRE200]
400	[Panph6,L18,(200557,163876,161484),S1,KK1025,CH3518244376,COLIBRE400]

APPENDIX C: STELLAR BIRTH PROPERTIES IN GALAXY STELLAR MASS BINS

Fig. C1 is analogous to Fig. 10, but compares the stellar birth properties in three different bins of $z = 0$ galaxy stellar mass (different line styles). We see that the physical conditions of the gas in which

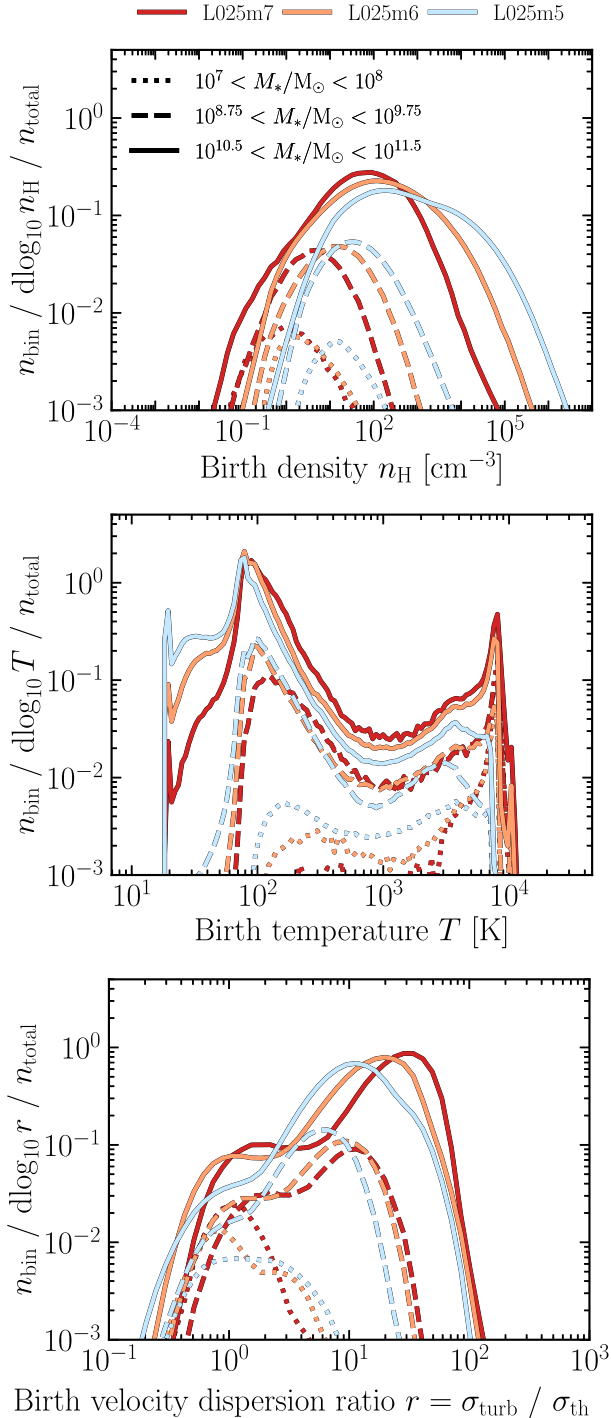


Figure C1. As Fig. 10, but for three different non-contiguous bins of $z = 0$ galaxy stellar mass (different line styles). In more massive galaxies, stars tend to form in gas that is denser, has a lower temperature, and in which turbulence is more supersonic.

the stellar particles were born depend strongly on galaxy mass. In more massive galaxies, stars tend to form from gas that is denser, colder and where turbulence is more supersonic. Because the masses are $z = 0$ values, the low-density, high-temperature, low-turbulence parts of the distributions for massive galaxies largely reflect stars born at higher redshifts in lower mass progenitors.

APPENDIX D: THE EFFECT OF RANDOM ERRORS IN STELLAR MASSES

As discussed in Section 7, when comparing to observations, we mimic the effects of random errors on the observed galaxy stellar mass by adding random lognormal scatter to the predicted masses with zero mean and standard deviation $\sigma_{\text{random}}(\log_{10} M_*)$, where $\sigma_{\text{random}} = 0.1$ dex at $z = 0$ (equation 43). Because it is unclear exactly how large the random errors should be, Fig. D1 shows the effect of varying σ_{random} between 0 and 0.5 dex for all plots shown in Section 7 that have stellar mass along the x-axis. For clarity, we only show the results for L400m7, which provides the best statistics at the massive end. We do not add scatter due to random errors on quantities other than galaxy mass, which may, however, also be important.

While the effect of our fiducial scatter of 0.1 dex is small, larger stellar mass errors can have a significant impact on galaxy scaling relations. They increase the scatter at fixed mass and can substantially affect trends with mass. Scatter in the stellar masses can be particularly important for galaxy scaling relations near masses where galaxy properties change rapidly with mass, as is the case for BH mass at the stellar mass corresponding to the halo mass used for BH seeding, and for the quenched fraction near the low-mass upturn. In the L400m7 simulation, these transitions are sharp due to numerical effects. The effect of mass scatter on the high-mass end of the scaling relations, $M_* \gtrsim 10^{11} M_\odot$, is physically more interesting. Here, random scatter results in Eddington bias, i.e. the fact that there is significantly more upscatter than downscatter if the mass function declines steeply, which smooths out both the SMF itself and trends of galaxy properties with mass.

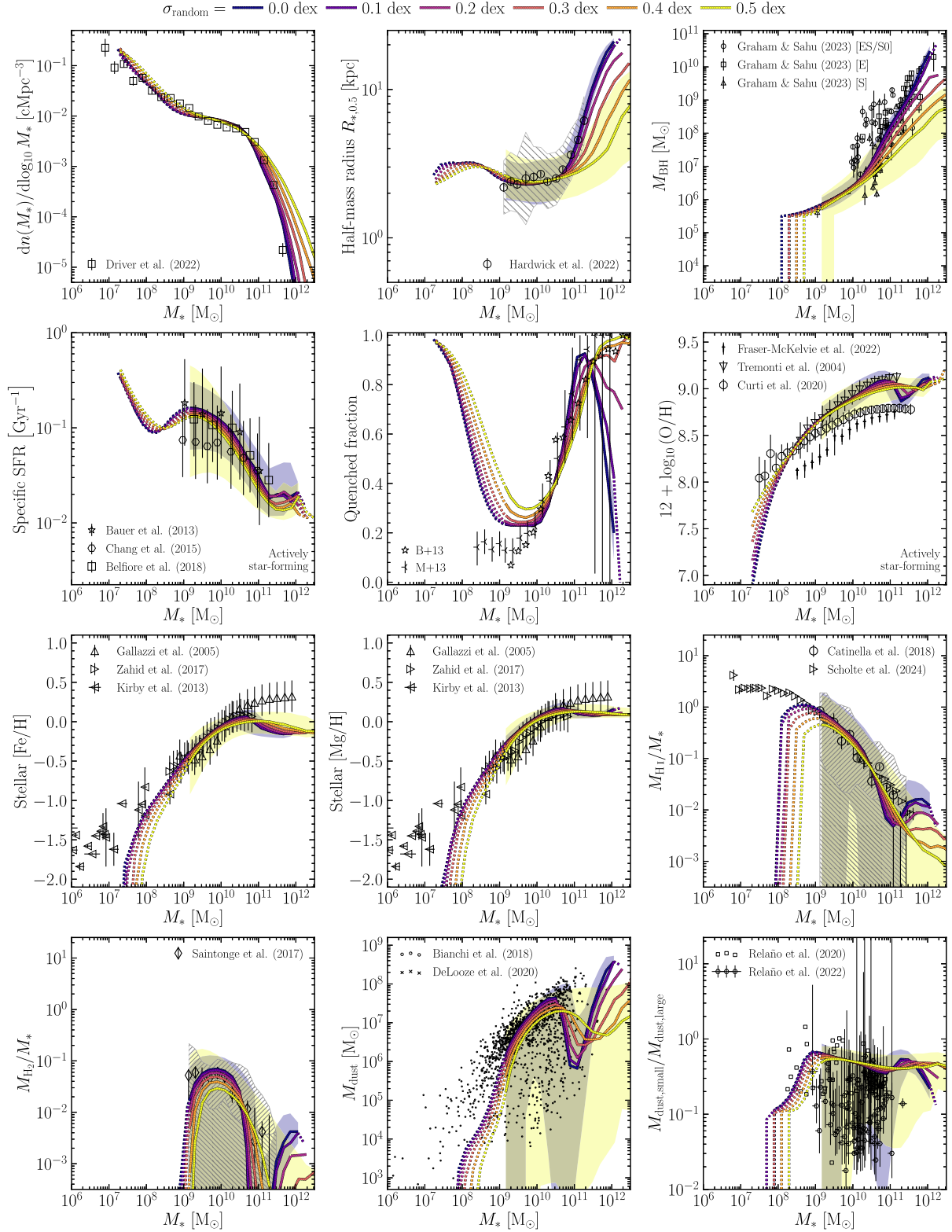


Figure D1. The effect of imposing random lognormal scatter on galaxy stellar mass, with zero mean and standard deviation σ_{random} on $\log_{10} M_*$, on $z = 0$ galaxy scaling relations in the L400m7 simulation. Different colour lines correspond to different amounts of mass scatter, with σ_{random} varying between 0 (i.e. no errors) to 0.5 dex. Dark blue and yellow shaded regions, which are present in all panels except for the one showing quenched fractions, show the 16th–84th percentile scatter for $\sigma_{\text{random}} = 0$ and 0.5 dex, respectively. The different panels repeat all the plots with stellar mass on the x-axis shown in Section 7, where we assumed $\sigma_{\text{random}} = 0.1$ dex. The Eddington bias resulting from random mass errors tends to smooth out mass trends and increase the scatter. It can be important near masses at which galaxy properties change rapidly with mass and for massive galaxies, $M_* \gtrsim 10^{11} M_\odot$, where the SMF drops exponentially.

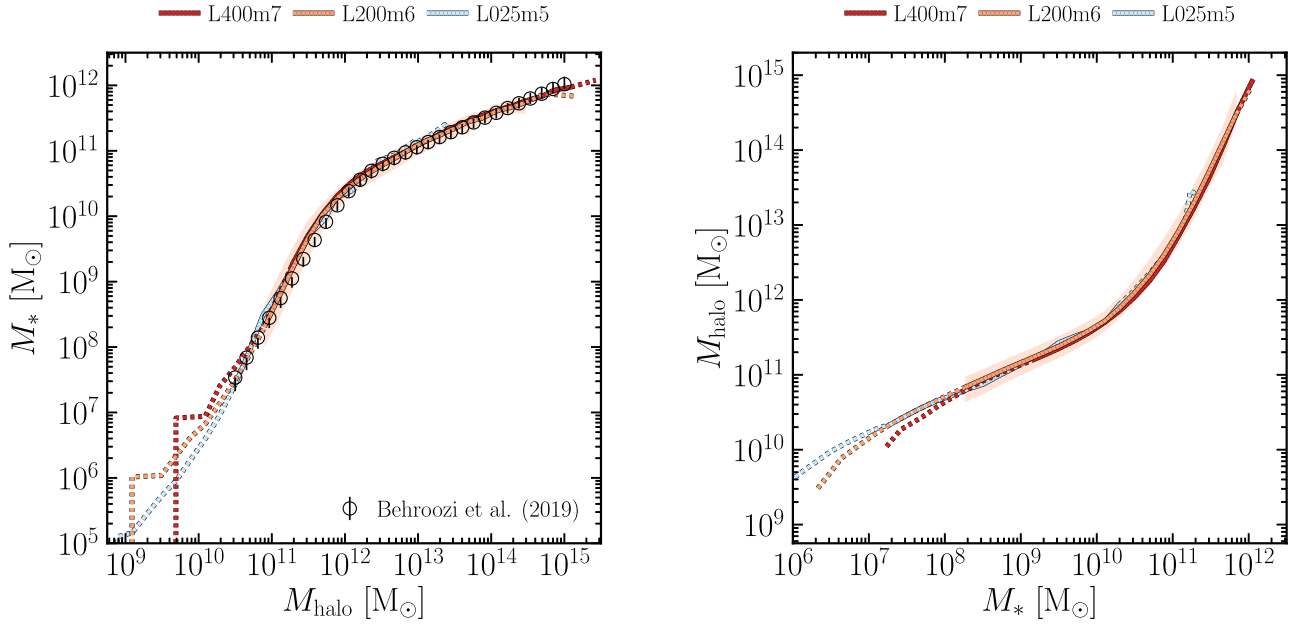


Figure E1. As Fig. 13, but showing stellar mass as a function of halo mass in the left panel and vice versa in the right panel. Halo mass is defined as the mass within the radius for which the mean internal density is equal to the critical overdensity provided by G. L. Bryan & M. L. Norman (1998).

APPENDIX E: THE RELATIONS BETWEEN STELLAR AND HALO MASS

To facilitate easy conversion between the stellar and halo masses of central galaxies, Fig. E1 shows the median stellar mass as a function of halo mass (left panel) and the median halo mass as a function of stellar mass (right panel).

APPENDIX F: COMPARISON WITH THE EAGLE AND ILLUSTRITNG SIMULATIONS

In this appendix, we compare the COLIBRE predictions with those from the two suites of cosmological hydrodynamical simulations that are currently the most widely used, EAGLE (J. Schaye et al. 2015; R. A. Crain et al. 2015; S. McAlpine et al. 2016) and IllustrisTNG (A. Pillepich et al. 2018a, b; D. Nelson et al. 2019a). Like COLIBRE, both simulations span multiple resolutions. EAGLE Recal_L025N752 and Ref_L100N1504 have baryonic particle masses of 2.3×10^5 and $1.8 \times 10^6 M_\odot$ in box sizes of 25 and 100 Mpc, respectively. These baryonic particle masses are similar to COLIBRE’s m5 and m6 resolutions, although for DM, the EAGLE resolutions are instead similar to COLIBRE’s m6 and m7 resolutions. EAGLE does not include a model with a baryonic mass resolution similar to m7. TNG50, TNG100, and TNG300 use box sizes of 51.7, 110.7, and 302.6 Mpc and have gas cell/stellar particle masses of 8.5×10^4 , 1.4×10^6 , and $1.1 \times 10^7 M_\odot$, respectively. These baryonic mass resolutions are similar to COLIBRE’s m5, m6, and m7 resolutions, although for DM, they are instead similar to m6, m7, and m8 using our notation.

Like COLIBRE, both sets of simulations include radiative cooling, star formation, chemical evolution, CCSN and SNIa feedback, BHs and AGN feedback. In contrast to COLIBRE, neither EAGLE nor TNG allows the ISM to cool to temperatures below $\approx 10^4$ K, and neither models the evolution of dust grains. As for COLIBRE, the stellar and AGN feedback in EAGLE are calibrated to the present-day SMF, galaxy sizes, and BH masses

in massive galaxies. TNG100 is calibrated to these observables and in addition to the stellar-to-halo-mass relation, the mass-metallicity relation, the mass scale above which star formation is quenched, the halo gas fraction in galaxy groups, and the cosmic star formation history. Unlike COLIBRE and EAGLE, the TNG subgrid feedback parameters are kept fixed when the resolution is changed.

Fig. F1 compares the SMF, stellar-to-halo mass ratio, galaxy sizes, BH masses, cosmic star formation history, sSFRs of star-forming galaxies, quenched fractions (i.e. fraction of galaxies with $\text{sSFR} < 10^{-11} \text{ yr}^{-1}$), and stellar metallicities for COLIBRE (solid) to EAGLE (dashed) and to the same observational data as shown in the corresponding figures from Section 7. Except for the star formation history, all predictions are for $z = 0$ and were taken from the public catalogues rather than computed from snapshot data. Fig. F2 shows a similar comparison with IllustrisTNG. We do not show H I, H₂, and dust masses (and dust grain sizes) because these properties are not directly predicted by EAGLE and IllustrisTNG. We show total stellar metallicities instead of iron and magnesium abundances. We do not compare gas metallicities because only COLIBRE accounts for the depletion of metals onto dust grains. We also do not compare the X-ray luminosity of the CGM, which is not part of the public catalogues. While we again use apertures with radii of 50 kpc for COLIBRE, we use 30 kpc for EAGLE and TNG, which are their fiducial values.²⁹ However, the two different apertures give nearly identical results for $M_* \gtrsim$

²⁹For IllustrisTNG, A. Pillepich et al. (2018a) use apertures of twice the half-mass radius. We use 30 kpc where possible for consistency, noting also that twice the half-mass radius can be very small, e.g. only ≈ 5 kpc for galaxies with masses similar to that of the Milky Way. For the stellar mass–halo mass relation, which for TNG300, we take from R. Pakmor et al. (2023) because they compute it down to lower masses than available in the public catalogue, we follow those authors and use all bound stellar mass.

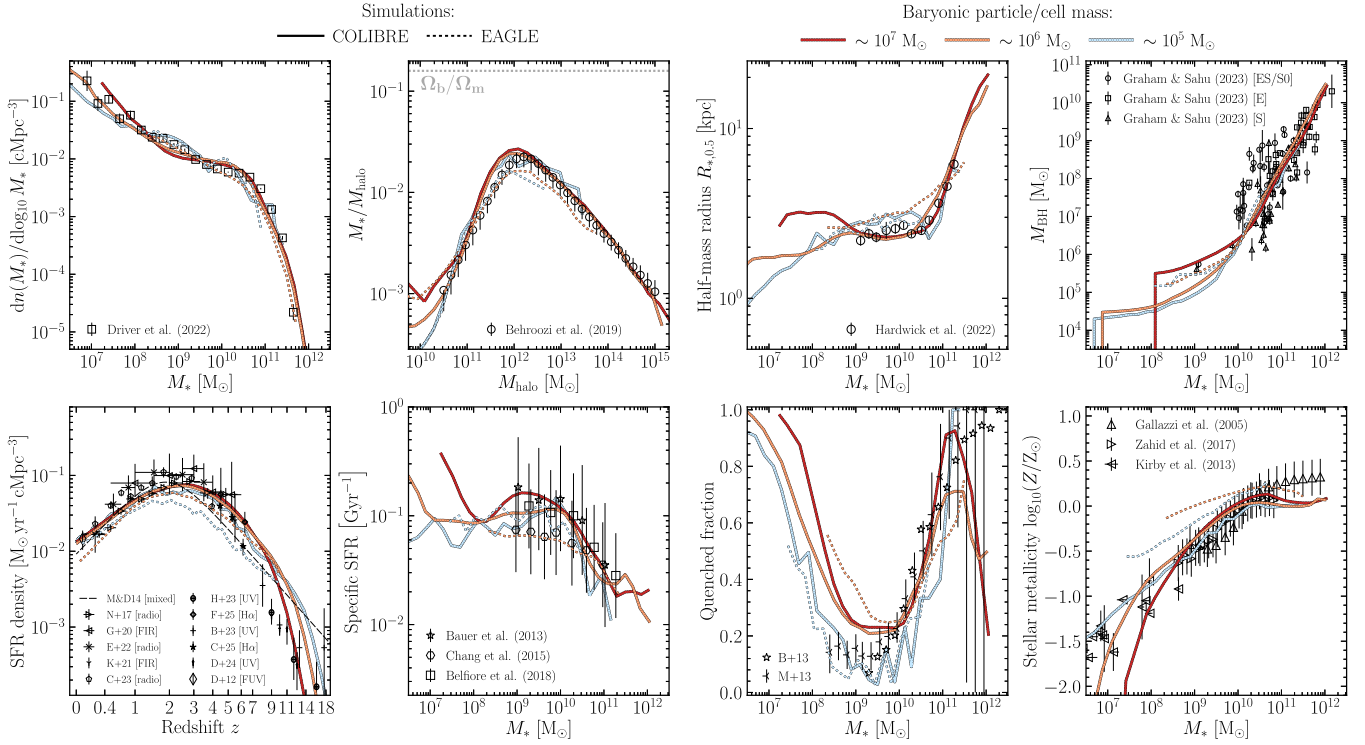


Figure F1. Comparison of COLIBRE (solid) and EAGLE (dashed) for galaxy observables that are direct results for both sets of simulations. From the top left to the bottom right, the different panels show the galaxy SMF, the stellar-to-halo mass ratio for central galaxies, galaxy sizes (i.e. projected half-mass radii), BH masses, the cosmic star formation history, sSFRs of star-forming galaxies, quenched fractions, and stellar metallicities. Except for the star formation history, all results are for $z = 0$. Line colours indicate the approximate baryonic particle mass, which is similar to COLIBRE’s L200m6 for EAGLE Ref_L100N1504 (orange) and similar to COLIBRE’s L025m5 for EAGLE Recal_L025N752 (blue). There is no EAGLE simulation with m7 resolution. Note that for EAGLE the corresponding DM mass resolution is a factor of four times worse than for the baryons, whereas they are approximately the same for COLIBRE. For EAGLE, all quantities are computed using 30 kpc apertures, which is the fiducial aperture choice for EAGLE, whereas we use 50 kpc for COLIBRE. The data points show the same observations as shown in Section 7. Halo mass is consistently defined as the mass within the radius for which the mean internal density is equal to the critical overdensity provided by G. L. Bryan & M. L. Norman (1998). Different from the plots shown in that section, none of the predicted stellar masses were perturbed to mimic the effect of random errors on the observed stellar masses.

$10^{11} M_{\odot}$ (e.g. J. Schaye et al. 2015; E. Chaikin et al. 2025b). As for COLIBRE, galaxy SFRs are instantaneous and computed from the gas particles. Different from the plots in the main text, we do not impose random stellar mass errors on any of the predicted masses because such errors were not considered in the calibration of EAGLE and TNG. However, as shown in Appendix D, the effect of the 0.1 dex scatter that we included in the plots shown in the main text is very small. The TNG300 stellar masses are the original predictions as opposed to the masses corrected for resolution effects.

Fig. F1 shows that for the SMF and BH masses the numerical convergence and agreement with the data are similarly good for EAGLE and COLIBRE, although at m6 resolution EAGLE slightly undershoots the knee of the SMF. Unlike COLIBRE, EAGLE m6, though not m5, predicts slightly too large galaxy sizes for $M_{*} < 10^{11} M_{\odot}$, but the difference is likely smaller than the systematics affecting the comparison of EAGLE with observations (A. de Graaff et al. 2022). The cosmic star formation history is better converged and reproduced by COLIBRE, with EAGLE underestimating the peak, particularly for m5 resolution. At $M_{*} \sim 10^{10} M_{\odot}$, the main sequence of star-forming galaxies is a factor ≈ 2 lower for EAGLE than for COLIBRE, but at m5 resolution the two are in good agreement. For the quenched fractions, the numerical convergence is worse for EAGLE, and

for m6 resolution, the agreement with the data is worse than for COLIBRE. However, at m5 resolution, EAGLE and COLIBRE predict similar quenched fractions. Finally, the stellar mass-metallicity relation is much shallower for EAGLE, and both the agreement with the data and the numerical convergence are clearly poorer than for COLIBRE.

Fig. F2 shows that the numerical convergence tends to be substantially worse for TNG than for COLIBRE, as might be expected from the fact that TNG was only calibrated at m6 resolution (TNG100). Galaxy stellar masses are a factor $\approx 2 - 4$ too high for TNG50, while they are too low for TNG300 (see e.g. A. Pillepich et al. 2018b; C. Engler et al. 2021 for a discussion of the effects of resolution on galaxy stellar masses in IllustrisTNG). While galaxy sizes agree well with COLIBRE and the data for $M_{*} \lesssim 10^{10} M_{\odot}$ at m6 resolution, at higher masses, they are too large for all TNG resolutions. TNG50 sizes only agree with COLIBRE and the observations for $M_{*} \approx 10^9 M_{\odot}$. While BH masses are similar for COLIBRE and TNG at the high-mass end, $M_{*} > 10^{11} M_{\odot}$, TNG predicts much higher BH masses in low-mass galaxies. For $M_{*} \lesssim 10^{10} M_{\odot}$ the BH masses in TNG appear too high compared to the data. For m6 resolution, the cosmic star formation histories of TNG and COLIBRE are in good agreement, but at m7 resolution the SFR in TNG is significantly below that of COLIBRE and the observations, while

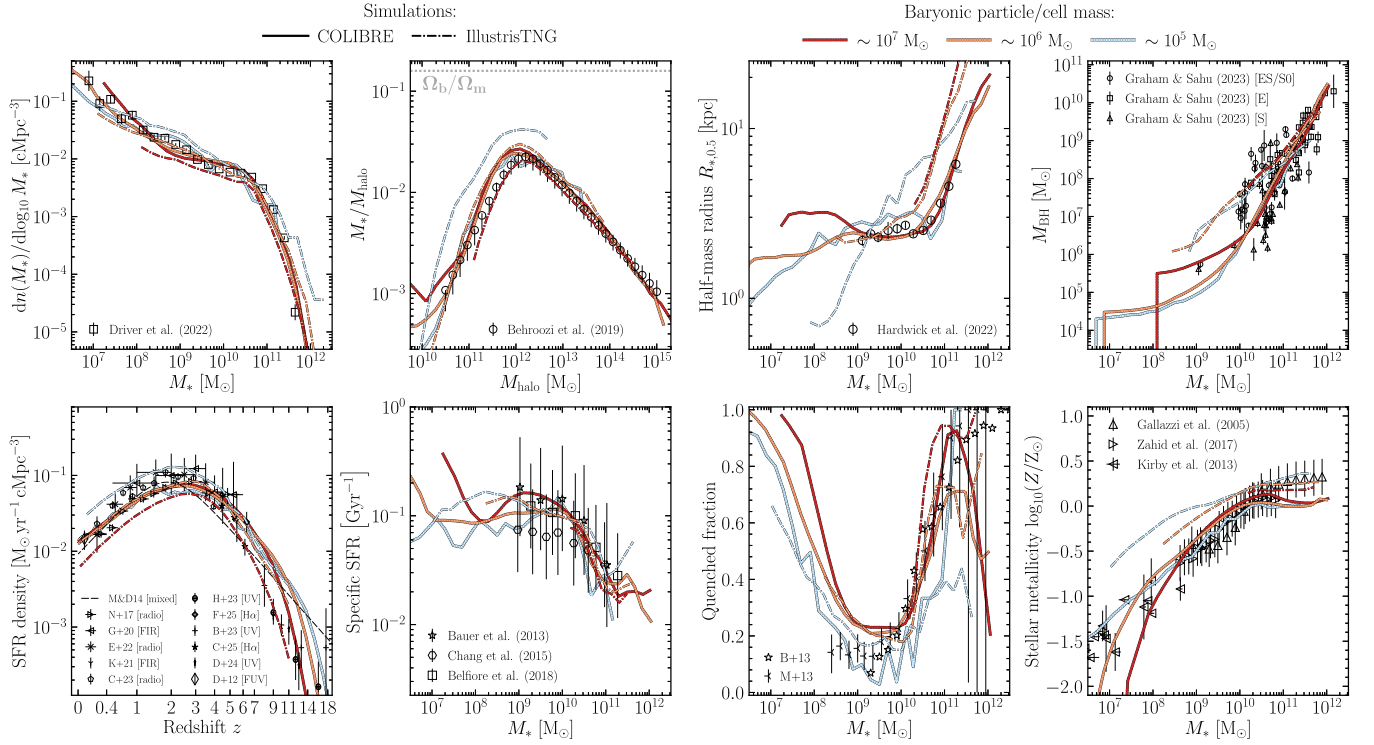


Figure F2. As Fig. F1, but comparing to IllustrisTNG instead of EAGLE. Line colours indicate the approximate baryonic particle/cell mass, which is similar to our L400m7 for TNG300 (red), to L200m6 for TNG100 (orange), and to L025m5 for TNG50 (blue). Note that for TNG, the corresponding DM mass resolution is a factor of four times worse than for the baryons, whereas they are approximately the same for COLIBRE. For TNG, galaxy properties are computed using 30 kpc apertures, the fiducial TNG value for e.g. the mass function, whereas we use 50 kpc for COLIBRE. The exception is the stellar mass in the stellar mass–halo mass relation, which for TNG300 we have taken from R. Pakmor et al. (2023) because they computed the relation down to lower masses than available in the public catalogue, but who use all bound mass for this relation. The TNG300 SMF, which has also been taken from R. Pakmor et al. (2023), does however use 30 kpc apertures. For COLIBRE and the data from P. Behroozi et al. (2019) halo mass is defined as the mass within the radius for which the mean internal density is equal to the critical overdensity provided by G. L. Bryan & M. L. Norman (1998), while for TNG, it is equal to 200 times the critical density.

at m5 resolution, it is significantly above both COLIBRE and the data for $z < 3$. For $M_* > 10^9 M_\odot$, the sSFRs and quenched fractions are in broad agreement for each resolution, although TNG50 predicts lower quenched fractions at the highest stellar masses (see M. Donnari et al. 2021, for a discussion of the impact of resolution and statistical sampling). Finally, like EAGLE, TNG

predicts a substantially shallower mass–metallicity relation than COLIBRE, resulting in too high metallicities for $M_* \lesssim 10^{10} M_\odot$.

This paper has been typeset from a $\text{\LaTeX}$ file prepared by the author.