



Full length paper

Variability in cosmological hydrodynamical simulations: How stochastic processes, numerical effects, and reproducibility limits impact predictability

Chaitra ^{a,b} ^{*}, Antonio Ragagnin ^a, Milena Valentini ^{b,a,c,d}, Giuseppe Murante ^{a,e,d}, Stefano Borgani ^{b,a,c,e,d}, Giuliano Taffoni ^{a,d,b}

^a INAF (Italian National Institute for Astrophysics) – Astronomical Observatory of Trieste, Trieste, 34143, Italy

^b Department of Physics, University of Trieste, Trieste, 34143, Italy

^c Istituto Nazionale di Fisica Nucleare, Trieste, 34127, Italy

^d Italian Research Center on High Performance Computing, Big Data and Quantum Computing, Casalecchio di Reno, 40033, Italy

^e Institute for Fundamental Physics of the Universe, Trieste, 34151, Italy

ARTICLE INFO

Keywords:

Chaos
Cosmological hydrodynamical simulations
Galaxy clusters
Opengadget
Run-to-run variability

ABSTRACT

Cosmological hydrodynamical simulations are powerful tools for studying galaxy formation, yet their predictive precision is limited by stochastic variability and numerical uncertainty. We quantify this variability using four identical realizations of a zoom-in galaxy-cluster simulation evolved with OPENGADGET3 under tightly controlled compiler, library, and hardware settings. Variability is measured through the properties of matched galaxies across repeated runs, including a mixed linear model that separates run-to-run variation from within-run noise.

Variations of approximately 10–25% are found in galaxy dark matter and stellar masses for the baseline simulations. The variability trending above the shot-noise floor reflects the combined effects of stochastic star formation and feedback regulation, and is further amplified when black hole physics is included.

Furthermore, our results indicate that feedback acts to regulate variability, reducing scatter in both stellar and black hole masses. Our inference from run-to-run variation indicates a noise-dominated regime that remains statistically reproducible, despite individual realization differences.

These results establish baseline, noise-dominated variability estimates at low resolution, demonstrate how feedback modulates predictability, and provide a statistical framework for future studies of reproducibility in cosmological hydrodynamical simulations.

1. Introduction

Cosmological hydrodynamical simulations have become essential tools for understanding galaxy formation and evolution, enabling researchers to explore complex physical processes that span vast ranges in time and scale. Major simulation projects like Illustris, EAGLE, and IllustrisTNG (Vogelsberger et al., 2014; Schaye et al., 2015; Pillepich et al., 2018) have provided crucial insights into how galaxies form within the cosmic web, how feedback processes regulate star formation, and how black holes coevolve with their host galaxies.

However, a fundamental challenge emerges when using these simulations to make quantitative predictions: their outcomes exhibit inherent variability that can significantly affect scientific conclusions. Even when simulations are initialized with seemingly identical conditions, individual galaxy properties can vary substantially between runs, raising critical questions about the reliability and interpretation of simulation predictions (see e.g. Thiébaud et al. (2008), Su et al. (2018),

Keller et al. (2019), Genel et al. (2019), Davies et al. (2021, 2022, 2023), Borrow et al. (2023)).

This variability in cosmological simulations arises from multiple distinct sources. At the most fundamental level, gravitational N-body systems can exhibit sensitive dependence on initial conditions (a hallmark of chaotic behavior), as first recognized in astrophysical contexts by e.g. Miller (1964), Standish (1968). However, modern cosmological simulations incorporate additional complexity through stochastic sub-grid physics models for star formation, stellar feedback, and black hole accretion, each introducing algorithmic stochasticity that can amplify small differences into large-scale variations in galaxy properties.

Variability in cosmological simulations has been investigated in several studies using different simulation codes (e.g. AREPO, SWIFT, GASOLINE2, RAMSES). While small-scale perturbations are expected to average out in ensemble statistics of phase-space quantities, individual galaxy properties can remain sensitive to such perturbations. These

^{*} Corresponding author at: INAF (Italian National Institute for Astrophysics) – Astronomical Observatory of Trieste, Trieste, 34143, Italy.

E-mail address: chaitra.chaitra@inaf.it (Chaitra).

works have reported repeated-run variations of approximately 2–25% in various galaxy properties (e.g. Genel et al. (2019), Keller et al. (2019), Borrow et al. (2023), Pakmor et al. (2025)). The magnitude of the variation depends on both resolution and subgrid physics, with more stable quantities (e.g. halo mass, M_{dm} , V_{max}) showing smaller scatter, while higher variability is associated with processes such as mergers, stochastic or “bursty” feedback, and black hole growth and dynamics (In A, we provide a brief overview of the numerical and physical mechanisms that contribute to variability in cosmological simulations and summarize relevant findings from the literature).

Pure N-body simulations can be argued to be exhibiting chaotic behavior despite the presence of round-off error. But, when extending to hydrodynamic simulations incorporating subgrid models (i.e. for star formation, black hole etc.), any observed “chaotic-like” behavior is strongly influenced by stochastic processes. Thus, in cosmological simulations, the presence of stochasticity complicates or even prevents the unambiguous identification of effects of chaos, as well as knowing whether stochastic processes suppress or amplify underlying chaotic effects.

A more precise description would treat the observed divergence as a consequence of stochastic variability, possibly with features of stochastic-chaos, but not attributable to deterministic chaos in the classical sense.

Understanding the sources of simulation variability is crucial for several reasons. First, it determines the appropriate uncertainty estimates for simulation predictions, affecting how we compare simulations to observations. Second, misidentifying the dominant sources of unpredictability can lead to incorrect conclusions about the underlying physics of galaxy formation.

This study expands on earlier work in three key directions. First, it provides the first variability analysis performed using OPENGADGET3, a widely used cosmological code for which stochastic divergence has not yet been quantified. Second, we introduce a controlled, module-based comparison that isolates the impact of individual physical processes such as stellar physics (cooling, star formation, stellar winds, and supernova (SN) feedback) and black hole physics (seeding and active galactic nucleus (AGN) feedback) on variability growth. This helps us address the question on whether feedback prescriptions amplify or regulate the growth of variations. Third, we apply a mixed linear modeling framework to statistically separate numerical noise from intrinsic run-to-run divergence, providing a structured way to assess convergence across repeated simulations.

The paper is structured as follows: We present the simulation setup and data processing in Section 2, followed by the statistical methods for variability analysis in Section 3. Section 4 presents the results for simulations without black holes, including both the baseline and feedback variation tests. Section 5 extends this analysis to simulations with black holes, comparing different feedback strengths and also discussing the effects observed in mixed precision configurations. Finally, Section 6 summarizes the main findings and outlines the implications for future simulation studies. Supplementary details are provided in the Appendices. These include results for dark-matter-only, dark-matter-plus-hydrodynamical simulations, tests using varied MPI task counts, and additional variability results for simulations with black holes and mixed precision configuration. Finally, we provide a technical description of the stochastic subgrid physics and random number generator (RNG) implementation.

2. Methodology

Our aim is to quantify a baseline of total variability for cosmological hydrodynamical simulations using OPENGADGET3. A key aspect of our methodology is the systematic isolation of astrophysical processes to determine their individual contributions to this divergence. We proceed progressively: first establishing baselines with stellar feedback alone, then incorporating black hole physics to assess their additional impact.

Table 1

Simulation configuration and cosmological parameters.

Dianoga D1-run configuration			
Resolution	m_{DM}	$m_{\text{gas,initial}}$	ϵ_{DM}
1×	$8.43 \times 10^8 M_{\odot} h^{-1}$	$1.56 \times 10^8 M_{\odot} h^{-1}$	5.62 kpc h^{-1} for $z > 2$: 16.87 kpc h^{-1}
Λ CDM Model Parameters			
Ω_M	Ω_B	h	σ_8
0.24	0.04	0.72	0.8

This allows us to disentangle the influence of specific mechanisms rather than merely observing their combined effect, providing clearer insights into the drivers of simulation variability.

The experimental framework consists of four main components: (1) a suite of identical simulations to establish baseline variability under controlled conditions, (2) systematic variations in feedback parameters to isolate the effects of different algorithmic stochasticity, (3) a galaxy matching procedure to track individual objects across simulation realizations, and (4) statistical methods to decompose total variability into its constituent sources.

2.1. OPENGADGET3: Code and physical modules

All simulations in this study were conducted using OPENGADGET3, a novel implementation of the Gadget code (Springel et al., 2001). We employ a stable version of OPENGADGET3 (consistent with Groth et al. (2023) and configured similarly to Bassini et al. (2020)) that provides reliable performance and flexibility for extensive parameter exploration. The code uses TreePM for gravitational forces and Smoothed Particle Hydrodynamics (SPH) for gas dynamics, with adaptive time stepping to handle the vast range of dynamical timescales in cosmological simulations¹.

The simulation code incorporates comprehensive subgrid physics following the implementations detailed in Bassini et al. (2020). Key physical processes include:

- Cooling and metal enrichment (Tornatore et al., 2003, 2007)
- Star formation following Springel and Hernquist (2003) with Schmidt law-like prescriptions
- Stellar feedback including supernova explosions, galactic winds, and chemical enrichment
- Black hole physics (Springel et al., 2005) with Bondi-Hoyle-Lyttleton accretion and thermal feedback (when included)

A summary of the stochastic subgrid physics and RNG implementation used in this study is provided in F. For our controlled experiments, we systematically enable and disable specific components to isolate their individual contributions to simulation variability.

2.1.1. Initial conditions

The zoom-in initial conditions used in this project for all simulations are the ones described by Bonafede et al. (2011). They are the so-called “Dianoga-set” of simulations (see Ragone-Figueroa et al. (2013), Rasia et al. (2015), Planelles et al. (2017), Biffi et al. (2017), Biffi et al. (2018), Ragone-Figueroa et al. (2018), Bassini et al. (2019) and Bassini et al. (2020)) that were generated from a parent low-resolution dark

¹ For reference, the project’s repository is: <https://www.space-coe.eu/codes/opengadget.php>. The specific production version used in this work is hosted at: <https://www.ict.inaf.it/gitlab/tscosmo/opengadget3-triesteproductionversion>. At the time of publishing of this paper, OPENGADGET3 is a private code with its community-wide release planned in the forthcoming months; we provide these repositories for reference until the code base is available publicly. The code will also be shared on request.

matter only simulation of comoving cubic volume $1 \text{ (Gpc } h^{-1})^3$ executed with GADGET-2 code and adopting a Λ CDM cosmological model (see Table 1). The dark matter particle has a mass of approximately $10^9 h^{-1} M_\odot$. This provides a relatively coarse initial resolution, sufficient to identify and track the formation of large-scale structures, including massive galaxy clusters.

The box is then processed with Friends-of-Friends (FOF; Davis et al. (1985)) to identify the most massive clusters: 24 in number, with $M_{200} > 8 \times 10^{14} h^{-1} M_\odot$ and 5 relatively less massive with M_{200} ranging from $1.4 \times 10^{14} h^{-1} M_\odot$. The volumes selected for each object are then simulated at different (higher) resolutions with zoomed initial condition technique (ZIC code, see Tormen et al. (1997) and Bonafede et al. (2011)). Each initial condition (IC) for the selected objects was generated by tracing the position of all particles within this Lagrangian region (a sphere of 5–7 virial radius at $z = 0$ centered on the respective object) all the way up to the beginning. The region traced in the parent box is divided into 64^3 cells, and the cells are resampled with higher resolution collisionless particles.

For simulations with full physics (gas, star formation, black holes etc.), baryons are added to the ICs along with dark matter. Thus ICs of different resolutions are available for running both dark matter only and full physics simulations.

For our work, we focus on one galaxy cluster (D1/g0016649) with virial mass $M_{\text{vir}} = 1.6 \times 10^{15} M_\odot$, simulated at the lowest resolution level (Dianoga 1x). Dark matter particle mass is $8.43 \times 10^8 M_\odot h^{-1}$, initial gas particle mass is $1.56 \times 10^8 M_\odot h^{-1}$, and gravitational softening lengths are $5.62 \text{ kpc } h^{-1}$ for gas/dark matter and $3.0 \text{ kpc } h^{-1}$ for stars/black holes.

2.1.2. Computational controls

In studies of simulation variability, it is critical to tightly constrain numerical factors because even subtle variations in code setup, compiler optimizations, libraries, or hardware architecture can affect results. Different code versions may contain algorithmic changes, bug fixes, or optimizations that alter operation ordering or numerical precision handling.

Library and Compiler: We fix all library versions (FFTW 3.3.10, GSL 2.7.1, HDF5 1.12.2) and use GCC 11.2.0 across all simulations to ensure reproducible floating-point operations. The exact compiler flags which fix the C++ standard and optimization levels, used for all simulations were: `-std=c++11 -Wall -g -O2`

We adopt OpenMPI 4.1.3 and maintain consistent parallelization schemes to avoid variations in operation ordering that can arise from different MPI implementations. Processes are not bound to specific cores or sockets (with the `-bind-to none` flag) and OpenMP was inactive.

Hardware and Execution Environment: All simulations run on the Hotcat cluster (Taffoni et al., 2020; Bertocco et al., 2020) using identical Intel Xeon CPU E5-2697v4 processors (2.30 GHz) with 256 GB memory. We maintain consistent computational resources: 4 computing nodes with 36 cores each, using 1 MPI task per core (144 total cores). Identical executables are used for all runs except when feedback parameter modifications require recompilation.

2.1.3. Parameter variations

Our experimental design systematically varies feedback parameters while maintaining all other conditions identical, enabling isolation of specific physical process contributions to simulation variability. For each parameter configuration, we run four identical simulations to quantify repeated-run variability. This sample size balances computational feasibility (each simulation test-set requires a total of $\approx 1.65 \times 10^4$ CPU core-hours, including postprocessing and analysis) with adequate sampling to characterize variability.

The Fiducial Model: The fiducial configuration incorporates comprehensive hydrodynamics and standard subgrid physics. For stellar feedback, the fiducial wind velocity is $v_w = 350 \text{ km s}^{-1}$, corresponding

Table 2

Feedback parameter configurations tested.

Configuration	Stellar feedback $v_w \text{ (km s}^{-1}\text{)}$	Black hole feedback ϵ_r
<i>Phase 1: Stellar feedback only (no black holes)</i>		
FIDUCIAL	350	N/A
LOW	200	N/A
HIGH	500	N/A
ZERO	0	N/A
<i>Phase 2: Combined stellar and black hole feedback</i>		
FIDUCIAL	350	0.1
HIGH	500	0.2
ZERO	0	0.0001

to approximately 50% of supernova energy converted into kinetic energy carried by galactic winds (Barai et al., 2015; Bassini, 2021).

Stellar Feedback Parameter Space: To isolate stellar feedback effects, we first conduct experiments without black hole physics by setting the critical halo mass for black hole seeding to an unphysical high value. The parameter ranges (Table 2) span physically motivated values consistent with previous cosmological simulations (e.g. Barai et al. (2015), Bassini (2021)).

Black Hole Feedback Parameter Space: In the second phase, we introduce black hole physics while maintaining stellar feedback configurations. The radiative efficiency values encompass the range commonly used in cosmological simulations (e.g. Springel et al. (2005)).

2.2. Galaxy identification and matching

This study quantifies variability through changes in macroscopic galaxy properties rather than individual particle trajectory divergence, capturing the practical impact of simulation unpredictability on observable cosmological phenomena. In practice, this approach tracks matched “clone” galaxies across simulation runs using unique particle IDs.

Fig. 1 provides a visual demonstration of the variability explored in this work, showing the same galaxy cluster at $z = 0$ across four identical runs of our Fiducial model (for comparison, analogous visualizations for dark-matter-only and hydro-only configurations are provided in B). Although the central galaxy remains morphologically similar across realizations, measurable differences are present. We focus our analysis on the Brightest Cluster Galaxies (BCGs) and their progenitors at high redshifts. These galaxies, which are the central subhalos of their respective halos, are chosen for several methodological reasons: they represent the most massive and luminous galaxies in each cluster and are better resolved in our low-resolution simulations, allowing for reduced shot noise. In contrast, the visible differences in satellite positions and structures across runs (highlighted in Fig. 1) are apparent. Satellites are particularly sensitive to small perturbations in timing, merger histories, and environmental interactions. Thus our approach deliberately limits the analysis to objects expected to show lower scatter, although it may underestimate total system variability.

The zoom-in region of the D1 cluster contains a large catalog of halos. We analyze the central subhalo of every identified halo within this volume, obtaining approximately 450–600 galaxies per snapshot (see Section 3.1). Galaxy identification is achieved through post-processing using FOF and SUBFIND algorithms (Springel, 2010; Dolag et al., 2009). Starting from one simulation as reference, we identify matching galaxies by selecting the most massive central galaxy within each parent halo across identical runs. For this work we use g3read and g3matcha (Ragagnin, 2025).

Our matching procedure involves two steps: (1) identify clone halos by requiring $> 60\%$ overlap in particle IDs and $> 50\%$ mass agreement, and (2) apply similar criteria to central galaxies within matched halos. We impose a minimum mass threshold of $10^{11} M_\odot$ to ensure robust

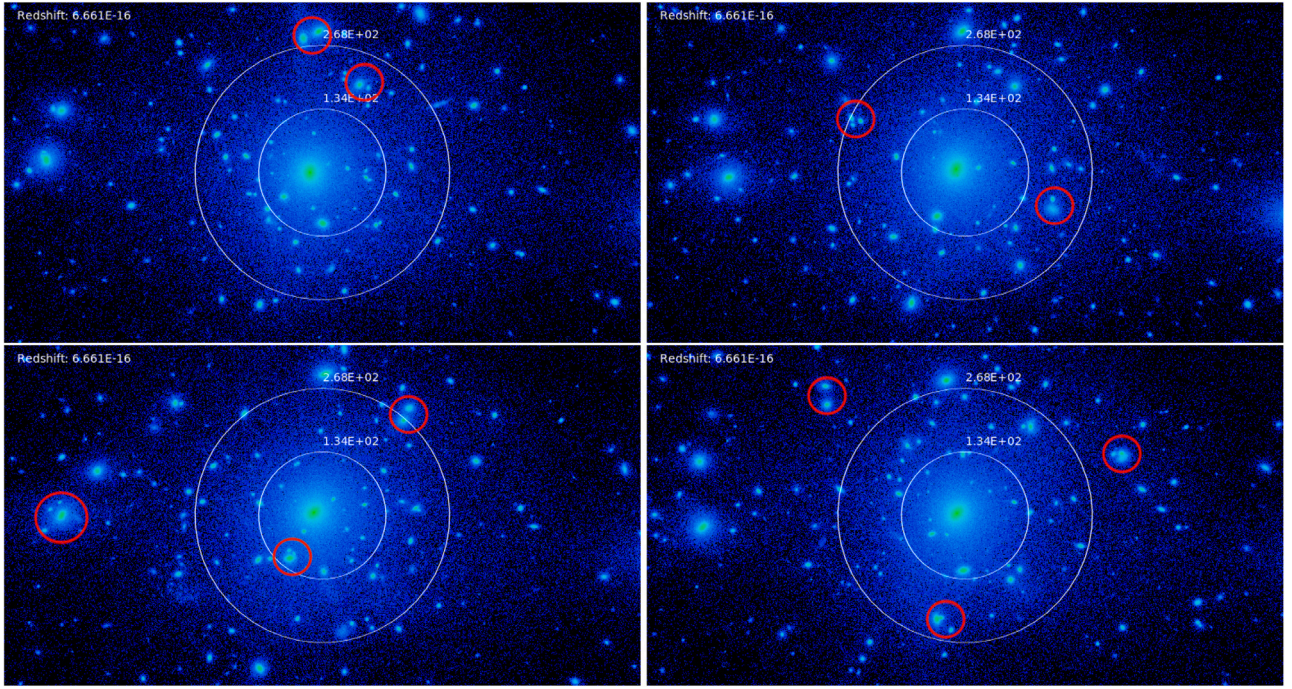


Fig. 1. Visual representation of the Dianoga simulation for the Fiducial model showing 4 realizations of the same galaxy cluster at $z = 0$. Substructure differences are encircled in red.

comparisons. These thresholds were chosen to balance completeness with accuracy based on preliminary tests.

While comprehensive matching methods such as Angulo and White (2010), Schaller et al. (2015), Lovell et al. (2018), Bose et al. (2019), Genel et al. (2019) and Borrow et al. (2023) can accommodate complex galaxy evolutionary processes (including merger events), this adapted approach provides a practical balance between accuracy and computational efficiency (considering the large catalog of galaxies in the D1 cluster). 95% – 98% of galaxies are matched with the adapted technique; a higher refinement in the matching algorithm will be left for future work, particularly for the study of impacts of mergers which is not in the scope of current work. For the matched galaxies, we extract properties from both SUBFIND catalogs (dark matter mass, stellar mass, black hole mass) and snapshot files (cold gas mass, additional stellar mass measurements using 50 physical kpc apertures, and black hole mass of the most massive black hole within a 30 pkpc aperture from the galaxy center).

3. Statistical analysis framework

Following the simulation execution and galaxy matching procedures described in Section 2, we obtain catalogs of matched clone galaxies with measured properties across identical simulation runs. We conduct statistical analysis for each set of simulations, following a methodology similar to that employed by Genel et al. (2019), where the variation is computed for a sample of galaxies matched across identical runs. This section presents our statistical framework for quantifying simulation variability.

3.1. Data preparation and preprocessing

Our dataset consists of matched galaxy catalogs from four identical simulations for each parameter configuration. The number of matched galaxies varies with redshift, averaging approximately 500 per snapshot across all runs. This provides a sufficiently large sample for statistical analyses, including the examination of variability for specific samples (e.g. variability for different halo mass bins). For cases where the

number of matched galaxies falls below 50 at a given redshift, we do not report results to avoid biases from poor sampling. As expected, less regulated feedback models (ZERO and LOW) yield larger numbers of substructures, resulting in a higher average of matched galaxies (≈ 600), whereas more efficient feedback runs (FID and HIGH) produce fewer systems, with averages of ≈ 450 – 550 . Nonetheless, provided the sample size exceeds the adopted threshold, central tendencies remain reliable for comparative analysis across feedback configurations.

Normalization: To ensure consistent measurements across simulations, galaxy properties are normalized using the within-subject mean (i.e. matched-galaxy mean), thus accounting for differences in scaling between galaxies while preserving relative differences between the runs. This approach isolates repeated-run variations from intrinsic galaxy-to-galaxy differences. The data are log-transformed (using base-10) to stabilize variance and improve the approximation to normality required for statistical modeling.

Values with z-scores exceeding ± 10 are excluded. This threshold removes only a small number of extreme outliers (typically 1–2 isolated points) that could disproportionately inflate the results. However, cases with numerous or systematically large deviations, which may reflect physically meaningful differences (e.g. due to varying feedback mechanisms), are retained deliberately.

Galaxy Properties Analyzed: We compute variation for galaxy properties including dark matter mass (M_{dm}) and stellar mass (M_*) available from SUBFIND catalogs, as well as cold gas mass extracted from 50 physical kpc apertures centered on matched galaxies from snapshot files. This size is a commonly adopted choice in OPENGADGET-based studies that captures the central galaxy while minimizing intracluster light (ICL) contamination and allowing for comparison with observational data (Ragone-Figueroa et al., 2018; Bassini, 2021; Marini et al., 2021) (although stellar mass variations computed for quantities extracted from 50 pkpc apertures compared to those of the SUBFIND catalog were not found to be markedly different).

3.2. Method 1: Pooled standard deviation

Our primary approach computes pooled standard deviations within bins of time rather than at individual simulation snapshots. $\Delta t = 0.2$ in

lookback time was adopted after testing for different bin sizes with an average of 40 bins depending on the test case. Binning mitigates noise from low-number statistics at individual snapshots while preserving temporal evolution of variability.

The pooled standard deviation for each time bin is computed as:

$$\sigma_{\text{pooled}} = \sqrt{\frac{\sum_{i=1}^N (n_i - 1) \sigma_i^2}{\sum_{i=1}^N (n_i - 1)}} \quad (1)$$

where σ_i is the standard deviation (in log10 space) across identical runs for galaxy i , and n_i is the number of available runs for that galaxy.

3.3. Method 2: Mixed linear model decomposition

The variation in galaxy properties is affected not just by run-to-run variation but also by noise within individual simulations. While run-to-run variation reflects systematic differences between simulations that would otherwise be identical, within-simulation noise originates from stochastic processes and computational limitations.

Method 2 employs Mixed Linear Models (MLM; Gelman (2006)) to separate these components. MLMs are particularly suited for hierarchical data where observations are nested within groups: in our case, each galaxy is measured across different simulation realizations (repeated measures).

Model Specification: The model decomposes galaxy properties into two variance components:

$$y_{e,i} = u_e + v_{e,i} \quad (2)$$

where:

- $y_{e,i}$ represents a given property for matched galaxy i in simulation run e
- $u_e \sim N(0, \sigma_u^2)$ is the random effect capturing variation between simulation runs, representing differences in overall mean galaxy properties between identical simulations i.e. run-to-run variation.
- $v_{e,i} \sim N(0, \sigma_v^2)$ is the residual error capturing variation within each simulation run, i.e. total noise

The model takes as input the dependent variable (e.g. matched galaxy stellar mass; see Section 3.1) arranged by simulation run. Thus, each observation consists of a galaxy ID, its simulation run label ($e \in 1,2,3,4$), and its measured property value.

Parameter Estimation: For each galaxy property, the MLM is fitted using Restricted Maximum Likelihood Estimation (REML), where fixed effects (group means) are first estimated, then variance components are estimated based on residual data. The total variation is partitioned into run-to-run variation and residual noise: both σ_u^2 and σ_v^2 are estimated. Here, “group means” refer to the estimated mean property value within each simulation run (i.e. the mean across all galaxies in run e).

Total Variance Decomposition: The total variability is expressed as:

$$\sigma_{\text{total}}^2 = \sigma_u^2 + \sigma_v^2 \quad (3)$$

This decomposition enables assessment of relative contributions from systematic differences between identical simulations versus random noise within individual runs.

Only results from converged models are reported, using the statsmodels package (Seabold and Perktold, 2010).

3.4. Shot noise estimation

As discussed in A, a baseline level of variability is expected from finite particle sampling i.e. shot noise. An approximate lower limit for shot noise in measurements of galaxy dark matter mass and stellar mass is given by:

$$\sigma_{\text{shot}} \approx \frac{1}{\sqrt{N}} \quad (4)$$

where N is the number of stellar or dark matter particles contributing to the galaxy property measurement. This serves as a lower limit on measurable total scatter, below which apparent variation cannot be meaningfully interpreted as physical or numerical divergence. Note, however, that individual components of the variability decomposition (e.g. the run-to-run variance from the MLM) may fall below this threshold, since the shot noise estimate constrains the total measurable variance, not its sub-components.

3.5. Connection to physical interpretation

A total σ value thus computed for a given galaxy property with the described methods can be interpreted as a fractional scatter across otherwise identical realizations of the same system. In this sense, the results serve analogously to an “error bar” on the predicted quantity, representing the effective precision limit of the simulation.

Given the numerous sources of variability in cosmological simulations (as discussed in A), it remains challenging (if not impossible) to disentangle the contributions of numerical noise from those of run-to-run divergence to these “error bars”. In this regard we employ a mixed linear model as an attempt to provide a statistical decomposition of the two. While the run-to-run or “chaotic-like” variability term may, in part, reflect divergence related to stochastic chaos, we adopt the neutral terminology “run-to-run variability” throughout for consistency. Note that in previous literature, the term “run-to-run variability” is often used to denote the total observed scatter, whereas here it specifically refers to the decomposed component of the total variability. We do not precisely quantify the level of each contribution at this stage, as such an analysis would require more complete tests including systematic resolution-based studies. Instead we focus on mainly comparative analysis between different simulation test sets, reporting only trends that are sufficiently evident rather than marginal changes. The second method is therefore intended as a framework for future investigations aimed at disentangling the relative roles of noise and “chaotic-like” divergence in driving variability.

4. Variability without black holes

We begin by examining the simulations excluding black holes to establish a baseline measure of variability arising solely from stellar physics, feedback, and numerical noise. This provides a controlled reference against which the impact of black holes i.e. accretion, feedback and black hole dynamics (presented in the next section), can be directly assessed. By comparing the two sets, we can disentangle variability intrinsic to our simulation framework, specific subgrid physics modules, and their respective feedback processes.

4.1. Ensemble averaged properties in the simulation volume

Galaxy stellar mass function (GSMF) for the Fiducial simulation (see Table 2) set is shown in Fig. 2. The GSMF is constructed for all galaxies identified by SUBFIND at $z = 0$. The lower panel shows the GSMF from each individual realization (clone) expressed as a fractional difference from the ensemble mean. The values cluster tightly around zero across most of the mass range, indicating very small clone-to-clone variation. Fig. 3 shows the star formation history (SFH) constructed by following the evolution of star formation rate for the identical simulations in the Fiducial set. The star formation rate (SFR) presented here is not instantaneous SFR, but derived from the stellar ages of the particles and represents the global SFH of the simulation volume at $z = 0$. The relative error of each realization (bottom panel) with respect to the mean SFR is minimal, signifying that the global SFH is a robust representation of the Fiducial model.

Together, these results provide a statistical overview across the population of galaxies in the four cloned simulations. The negligible discrepancies observed between the properties of galaxies among the

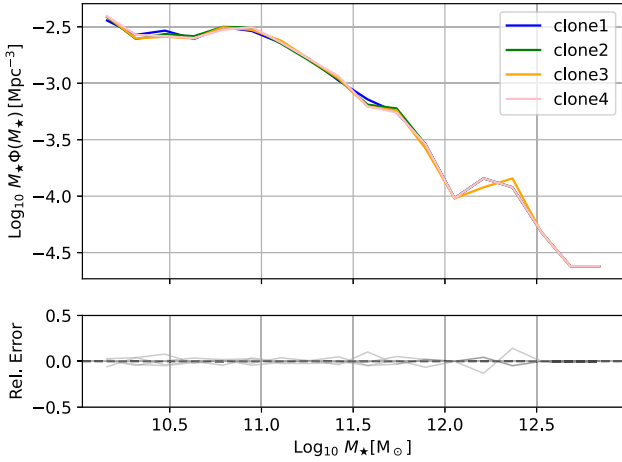


Fig. 2. Galaxy stellar mass function (GSMF) at $z \approx 0$ for the Fiducial simulation set. Each colored line corresponds to an identical run (clone) in the set, considering aggregate of stellar mass reported by SUBFIND for all galaxies in the respective run. The lower panel shows the relative differences of each clone's GSMF with respect to the mean of the ensemble, illustrating the minimal clone-to-clone variation. The horizontal dashed line indicates zero deviation.

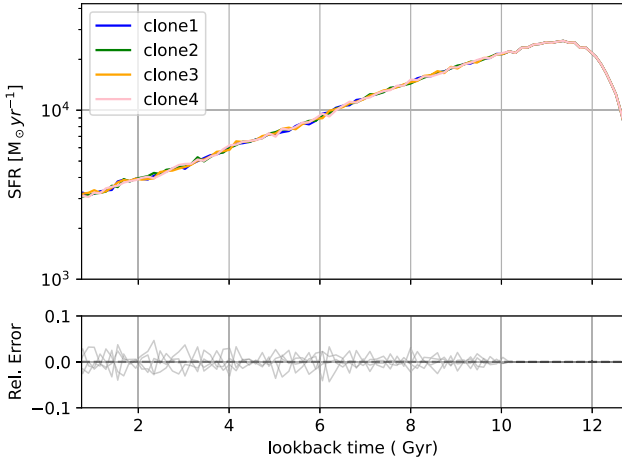


Fig. 3. Global star formation history across the four clones of the Fiducial simulation set as a function of lookback time (top panel), where each colored line corresponds to an individual realization (clone). Bottom panel shows the relative variation of each clone's star formation rate compared to the ensemble mean.

clones, across a large population and on large scales, demonstrate that the ensemble averages of galaxy properties are reproduced by the simulations despite underlying run-to-run variability. Indeed, while ensemble averages demonstrate minimal variation, the properties of individual galaxies can exhibit significant fluctuations. The following section explores these individual variations in greater detail.

4.2. Variation in fiducial simulations without black hole

Fig. 5 illustrates the variations in galaxy dark matter mass (left panel) and stellar mass (right panel), respectively, between matched galaxies in the identical simulations. These results pertain to the Fiducial simulation set run without black holes.

For both, the variations are quantified using the two methods: pooled standard deviation (**Method 1**, marked with black asterisks) and Mixed Linear Model (**Method 2**, colored, starting with the label “total”), as described in Section 3. The total standard deviation (cyan

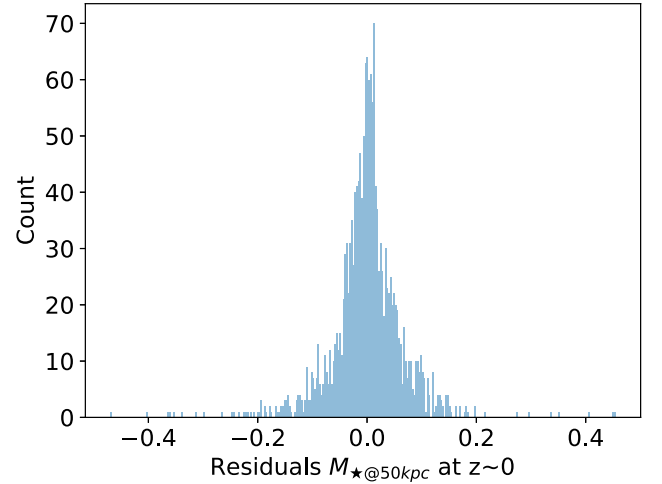


Fig. 4. Distribution of residuals in galaxy stellar mass for the Fiducial simulation set at $z = 0$.

circles) is calculated by summing both the residual error (i.e. variation within each simulation) and the run-to-run variation obtained using Method 2. The run-to-run variation is also plotted in the figures (in magenta-x), along with a rough estimate of the shot noise (dashed line). The data evolution over time is plotted with respect to lookback time in gigayears (Gyr).

Fig. 4 shows that the distribution of residuals for galaxy stellar mass at $z = 0$. The distribution is approximately Gaussian, centered near zero, and displays no strong skewness. This supports the assumption of normally distributed errors required by the Mixed Linear Model and indicates that the model fit is statistically well-behaved.

From **Fig. 5**, we observe that the variations computed using the two methods are consistent, although the data exhibit a scattered rather than smooth evolution. Star formation peaks before 10 Gyr; however, for the adopted setup, **OPENGADGET3** with full-physics is not numerically deterministic (although the individual modules for gravity, hydrodynamics, and star formation are). Consequently, minute variations remain negligible until approximately 9 Gyr for the Fiducial model, after which they grow and stabilize at $\sigma_{M_{\text{dm}}} = 0.03\text{--}0.05$ and $\sigma_{M_{\text{s}}} = 0.04\text{--}0.08$. In both cases, the total variation closely approaches the level expected from shot noise, indicating that shot noise dominates for these low-resolution runs. This behavior is most evident for $\sigma_{M_{\text{dm}}}$, while $\sigma_{M_{\text{s}}}$ remains slightly above the shot-noise limit.

The higher $\sigma_{M_{\text{s}}}$ reflects the cumulative effects of highly nonlinear and localized processes such as star formation, cooling, gas dynamics, and in particular, feedback mechanisms (as demonstrated in the succeeding section through feedback-related tests). At late times, the variation plateaus, likely driven by gas exhaustion, which suppresses star formation and limits further divergence. This physical constraint explains the observed saturation. Additionally, we tested the sensitivity of the system to small numerical perturbations at the outset by performing additional runs in which the number of MPI tasks was varied. This varied-MPI test set is conceptually similar to introducing small perturbations through alternative mechanisms (e.g. explicitly introducing small differences in the initial conditions) and provides one approach to probing variability in simulations that may otherwise yield byte-identical results. As expected, the varied-MPI runs show greater divergence, particularly at the peak of star formation, which is not visible in the identical-MPI-task set. However, the system still plateaus similarly to the unperturbed case, due to gas exhaustion regulating the growth of variations (see **C**). In comparison to $\sigma_{M_{\text{s}}}$, the plateauing of dark matter mass aligns closely with the shot noise floor, reflecting the inherently stable nature of collisionless dark matter dynamics.

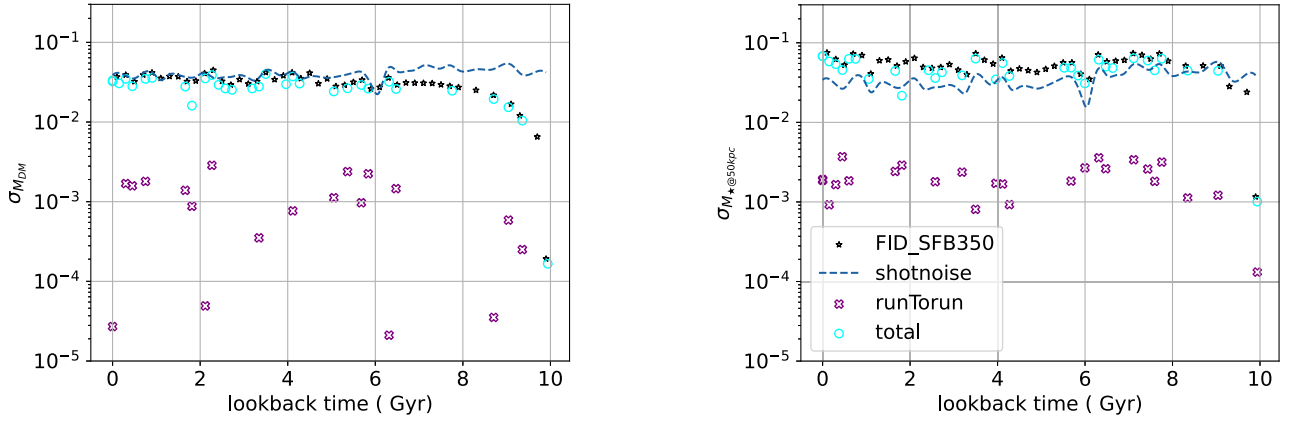


Fig. 5. Variation in galaxy dark matter mass (left panel) and stellar mass (right panel: stellar mass is comprised of all stellar particles within an aperture of 50 pkpc from galaxy center) for Fiducial simulations plotted as a function of lookback time in Gyrs. Variation as computed from Method 1 (pooled standard deviation; black stars) and Method 2 (total and run-to-run; cyan-circles and magenta-x respectively) is presented along with an approximate, lower limit estimate of shot noise (dashed colored line). The close overlap between the total variation computed from the two methods is visually apparent and highlights the consistency of the two approaches.

The run-to-run variation for each of these properties are similar and less than $\sigma = 0.005$. This is expected that the noise within each set of simulations is very high, considering the simulations are of very low resolution; we do not see any interesting increase in scatter between runs.

4.3. Feedback tests

Fig. 6 (top panel) shows the variation in dark matter mass for the 4 flavors of stellar feedback for runs: Fiducial, LOW, ZERO and HIGH as listed in Table 2. The variation is computed with Method 1, i.e. pooled standard deviation. Here we can see that the variation is not very different for each feedback case. The effect of changing stellar feedback on dark matter mass is negligible. Although the scatter in dark matter mass is slightly less pronounced towards later times (< 2 Gyr) for HIGH feedback than others, all the cases remain more or less the similar, with variation growing from 11–10 Gyr and plateauing to a range of $\sigma_{M_{dm}} = 0.03$ –0.06 with some scatter.

Fig. 6 (middle panel) shows the variation in stellar mass for each feedback case, where the total variation is computed with Method 1. The LOW and ZERO runs show reduced scatter compared to the Fiducial (already discussed), whereas HIGH run shows increased scatter, with σ_{M_*} between 0.06–0.1.

Why the increase in σ_{M_} for higher stellar feedback?* As already discussed for the Fiducial case, the run-to-run variation is low, σ_{M_*} under 0.005. Thus, the increase in σ_{M_*} is not due to higher feedback driving an increase in run-to-run variation. Instead, it indicates an increase in noise within each simulation, reflecting higher stochastic variation in each simulation associated with the feedback.

Stellar feedback can lead to “bursty” star formation because feedback from supernovae or winds can temporarily heat and expel gas, quenching star formation in some regions, while other regions may still have gas available to form stars. This leads to increased noise in the stellar mass associated with each simulation. The effect can be further assessed by checking the impact of stellar feedback driving high variation in the cold gas mass as shown in Fig. 6 (bottom panel). The total variation computed with Method 1 shows consistently high variability in the cold gas mass ($\sigma_{M_{cold}} \approx 0.4$ –0.8), with the highest scatter occurring in simulations with stronger stellar feedback, particularly at early times. This drives subsequent variations in the SFR, which ultimately results in higher variation in galaxy stellar mass in each simulation. Thus high stellar feedback, in simulations without black holes, increases the stochastic variation of a given simulation.

Does that mean that reduced or zero feedback is better in terms of reducing stochastic variation? With reduced feedback or lack thereof (cases LOW and ZERO), the only dominant effect is star formation, unregulated, it leads to excessive gas cooling, leading to overproduction of stars and thus rapid depletion of gas. Thus these galaxies show less variability in their star formation histories (albeit unrealistically high SFR), as star formation proceeds smoothly and continuously until all the gas is exhausted. This reasoning is particularly discussed by Keller et al. (2019) who run dwarf galaxy simulations with and without feedback. They state that for unregulated star formation, variations are dampened, considering the heightened star formation efficiency (SFE); but, the effect does not hold when this one-to-one relation is compromised. They test this by varying the gas fraction for different sets of simulations testing for both cases of with and without stellar feedback. They find that for lower gas fractions, the variation is higher for unregulated star formation and only dampened for certain higher values of gas fraction. But with stellar feedback, the fluctuations in stellar mass are maintained regardless of the gas fraction.

In addition to varying initial gas fractions, the presence of black holes can also disrupt the observed dynamics of high SFE, gas depletion and reduced σ_{M_*} . The variability introduced by stellar feedback dominates in the absence of additional regulating mechanisms such as black hole feedback, which, when included, can suppress star formation and thereby is expected to influence the overall variability.

4.3.1. Variation in different galaxy mass bins

The results for σ_{M_*} shown so far considers a large distribution of galaxies above $M_{SUB} > 10^{11} M_{\odot}$. In this part, we consider how the stellar mass can vary within galaxies belonging to different mass bins. In Fig. 7, σ_{M_*} is plotted for galaxies belonging to mass bins [$11.5 < \log M_{SUB}(M_{\odot}) < 12.0$] (top panel), [$12.0 < \log M_{SUB}(M_{\odot}) < 12.5$] (middle panel) and [$12.5 < \log M_{SUB}(M_{\odot}) < 13.0$] (bottom panel) respectively. Each time and mass bin always considers a sample size > 50 . As seen, in Fig. 7 top and middle panel, the results of the full distribution (Fig. 6 (middle panel)) is still reflected. Thus for the galaxies of relatively lower mass bins [$\log M_{SUB}(M_{\odot}) < 12.5$], the stochastic variation is higher for stronger feedback compared to LOW and ZERO feedback. But, when it comes to galaxies belonging to higher end of the mass bins [$\log M_{SUB}(M_{\odot}) > 12.5$], see bottom panel of Fig. 7 (containing galaxies of mass bins: [$12.5 < \log M_{SUB}(M_{\odot}) < 13.0$]), the variation for all the feedback types are scattered but with σ_{M_*} more or less similar without a significant trend. This result reflects the differential impact stellar feedback has on galaxies depending on their mass. Lower-mass galaxies experience a stronger impact from stellar

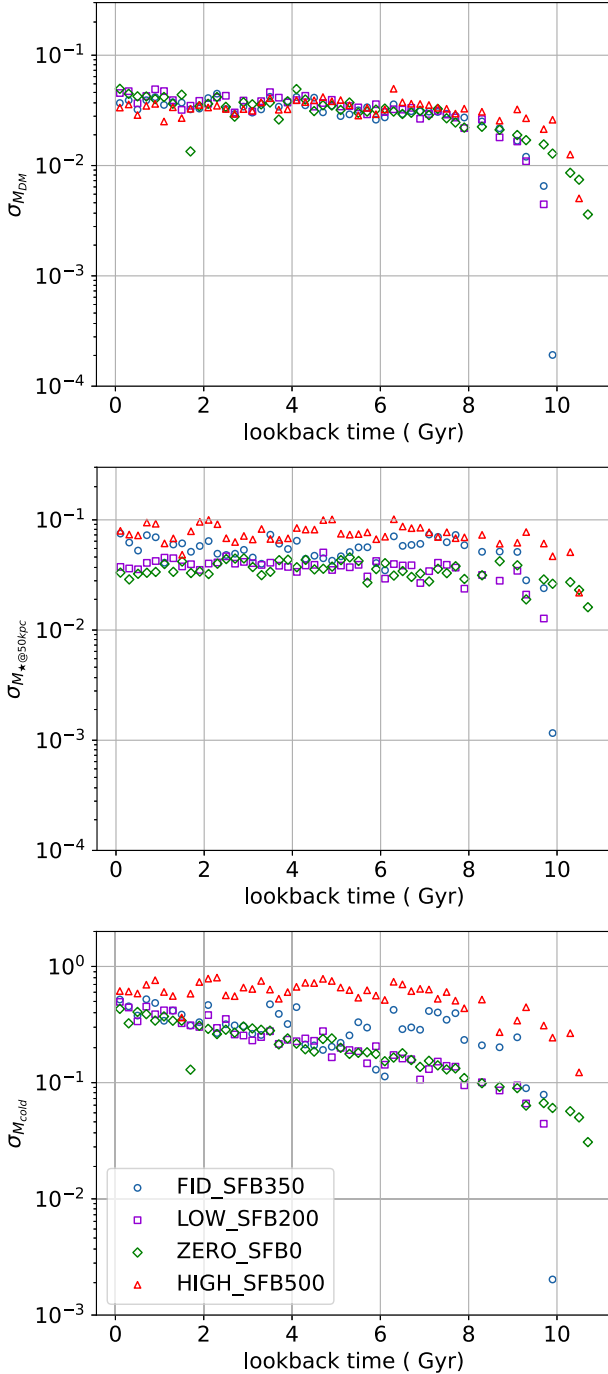


Fig. 6. Total variation (Method 1) in galaxy dark matter mass (top panel), stellar mass (middle panel) and cold gas mass (bottom panel) as a function of lookback time (Gyr), shown for the 4 feedback models: Fiducial, LOW, ZERO, HIGH stellar feedback. Each configuration is represented by a distinct color and marker, as indicated in the legend, and this scheme is maintained consistently throughout the paper.

feedback due to shallower gravitational potential wells; the energy and momentum of supernovae feedback or winds can drive gas out of the galaxy more efficiently, thus limiting the available gas for star formation and therefore having an impact on the galaxy's stellar mass. The varying trends in σ_{M_*} with respect to the feedback is a reflection of this.

On the contrary, higher-mass galaxies with deeper gravitational potential wells are less affected by stellar feedback. Although stellar feedback still plays a role, it is considerably less efficient in driving out significant amounts of gas. Thus, for the relatively larger mass bin [$12.5 < \log M_{SUB}(M_\odot) < 13.0$], the effect of HIGH feedback is not evident and thus the stochastic variation is more or less similar to a lower or unregulated star formation.

5. Variability including black holes

Fig. 8 shows results analogous to those in Section 4.2 (i.e. Fig. 5), but for simulations including black holes. Here, total variation (Method 2: sum of residual and run-to-run variation, in cyan-circles), run-to-run variation (in magenta-x) is represented for the Fiducial simulations with black holes for galaxy dark matter (left panel) and stellar mass (right panel). Also, the variation computed with Method 1 (black stars) is provided along with an approximate lower limit estimate of shot noise (colored dashed line).

Variations increase steadily from ≈ 9 Gyr, plateauing at $\sigma_{M_{dm}} \approx 0.04$ and $\sigma_{M_*} \approx 0.1$ thereafter (corresponding to $\approx 10\%$ and $\approx 25\%$ scatter, respectively). The total variation computed using both methods is consistent. For dark matter mass, the total variation closely approaches the lower limit set by shot noise, similar to the no-black-hole (NOBH) case.

For stellar mass, shot noise does not appear to be the sole contributor to the total variation at later times (considering the total stellar mass variation is systematically higher than in the corresponding NOBH case from ≈ 6 Gyr onward). The run-to-run component for both properties remains negligible compared to the total variation, with $\sigma \lesssim 0.01$.

The increased variations are driven by the inclusion of black hole subgrid physics, which is, no doubt, a significant source of variation. This is considering that black hole positions are highly sensitive to the gravitational interaction with surrounding particles, which can even lead to phenomena such as the wandering black hole (Ragone-Figueroa et al., 2018; Damiano et al., 2024). This can introduce a significant source of variation that can affect variations in stellar mass buildup. Additionally, both star formation and black hole accretion compete for gas; consequently, we can expect an increased sensitivity of stellar mass to local variation, and thus we observe higher scatter compared to simulations without black holes.

Note that across all properties examined here (including the baseline variation for our NOBH-tests), the residual variance dominates the run-to-run variance by at least an order of magnitude. Thus, considering that our low-resolution runs are noise-dominated: repeated runs converge to a well-defined expectation value for each galaxy property. By contrast, if the run-to-run component were to become non-negligible, the interpretation would change qualitatively. In that regime, each individual trajectory realized by the simulation should be regarded as equally valid. This distinction between noise-dominated and run-to-run variance has direct consequences for whether individual galaxies can be robustly studied within a given numerical setup. In the present work, we do not attempt to define a quantitative threshold for this transition (for e.g., in terms of the relative effect size or the fraction of run-to-run variance relative to the total variance), since this requires systematic resolution-scaling tests, which we defer to future work. Notably, the run-to-run component derived from Method 2 represents a statistical variance rather than a purely physical divergence; it can itself respond to changes in the underlying noise floor. Its relative effect size therefore provides a useful indicator of when repeated-run variability becomes non-negligible. Nonetheless, it is important to emphasize that determining whether variability is noise dominated or run-to-run dominated is scientifically significant: in the former case ensemble averaging converges to a true expectation value, while in the latter case ensemble statistics must be interpreted probabilistically, reflecting fundamental limits on predictability at the level of individual galaxies or sub-samples.

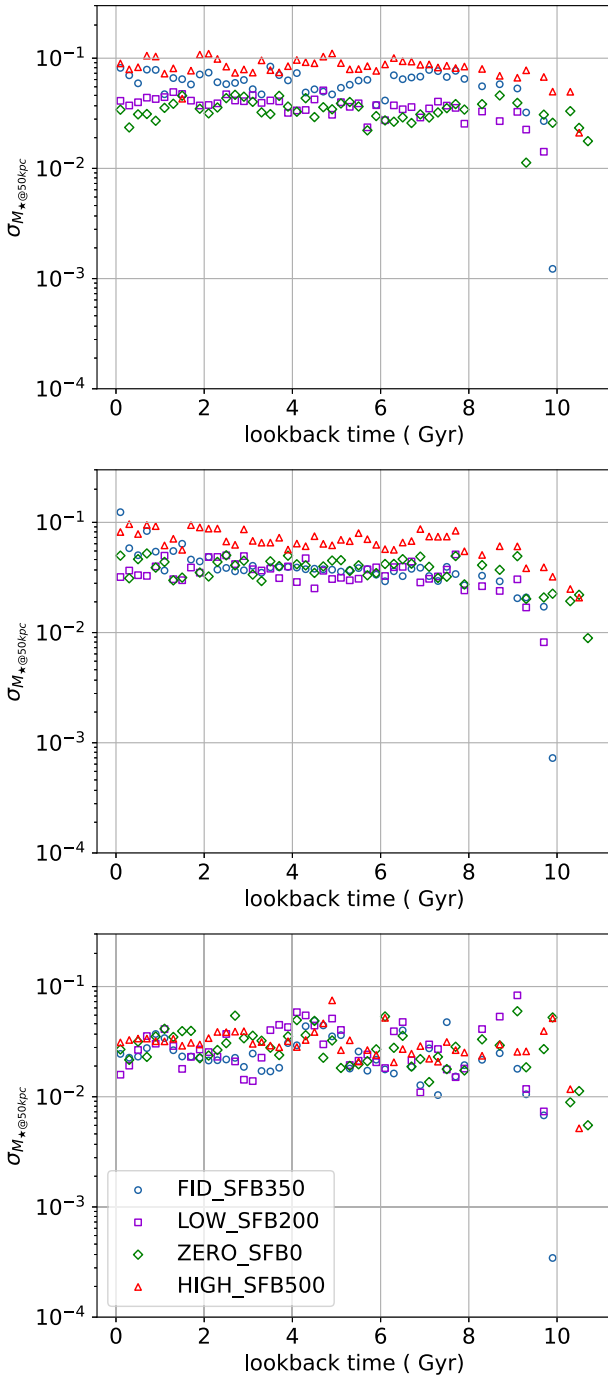


Fig. 7. Total variation (Method 1) in galaxy stellar mass as a function of lookback time in Gyrs for galaxies within a mass bin of $[11.5 < \log M_{SUB}(M_\odot) < 12.0]$ (top panel), $[12.0 < \log M_{SUB}(M_\odot) < 12.5]$ (middle panel) and $[12.5 < \log M_{SUB}(M_\odot) < 13.0]$ (bottom panel) shown for the four feedback models: Fiducial, LOW, ZERO, HIGH stellar feedback. M_{SUB} denotes the total mass summed from all the particles bound to the galaxy, taken from SUBFIND catalog.

5.1. Feedback tests

Fig. 9 shows the total variation in galaxy dark matter mass (computed with Method 1) from the feedback tests including black holes for ZERO, HIGH feedback tests. ZERO feedback case without black hole is provided for reference (see Table 2 for the respective feedback parameter values used).

The variation grows from around 7–8 Gyrs and plateaus to $\sigma_{M_{dm}} \approx 0.02$ –0.06. The plot clearly show that with stronger feedback, there is a reduction in the scatter in the galaxy dark matter mass; $\sigma_{M_{dm}}$ in HIGH plateaus below 0.03 with some scatter. Both the cases of zero feedback with and without black hole have similar trends with $\sigma_{M_{dm}}$ above 0.03 up to 0.06 with some scatter.

For the specific choice of feedback parameters in the HIGH feedback case, the impact on the galaxy population is most pronounced at the low-mass end ($M \lesssim 10^{12} M_\odot$). As shown in Fig. 10, where the GSMF is plotted for the FID, HIGH, and ZERO runs along with the (Bernardi et al., 2013) observational data provided for reference; considering the stellar mass ascribed by SUBFIND to all the galaxies in a single run in each respective test. GSMF is normalized following (Bassini et al., 2020). As can be observed in the Figure, the HIGH feedback case exhibits the characteristic flattening of the low-mass slope (Dekel and Silk, 1986; Larson, 1974; White and Frenk, 1991) much more prominent than Fiducial.

This selection effect results in the sample of HIGH-galaxies being, on average, denser than in the other cases (see Fig. 11 for a comparison of maximum circular velocity between HIGH and ZERO). The maximum circular velocity is defined as $V_{max} = \max \left[\sqrt{GM(<r)/r} \right]$ as function of radius r ; $M(<r)$ is the enclosed subhalo mass and G is the gravitational constant. This bias effectively lowers the shot noise in the sample and thus reduces $\sigma_{M_{dm}}$.

For stellar mass results, Fig. 12 (left panel) shows the total variation in galaxy stellar mass computed using Method 1 for the FID, ZERO, and HIGH feedback test cases. The ZERO feedback case exhibits variability comparable to the Fiducial. In contrast, the HIGH feedback case shows reduced variation at earlier epochs (≥ 5 Gyr), remaining below the FID and ZERO cases throughout. While the FID and ZERO runs approach a plateau of $\sigma_{M_*} = 0.1$ at $z = 0$, the HIGH feedback case continues to increase.

The scatter in σ_{M_*} is influenced by both stellar and black hole feedback. Results from the no-black-hole-feedback tests (Section 4.3) show that unregulated star formation leads to rapid gas exhaustion, resulting in a reduced σ_{M_*} . However, when black holes are included, this trend is reversed: the high-feedback case exhibits lower σ_{M_*} than the Fiducial setup. More efficient feedback better suppresses the scatter in star formation. Each feedback scenario eventually reaches a plateau in σ_{M_*} , which coincides with the onset of gas exhaustion. The timing and extent of this exhaustion are jointly controlled by stellar and black hole feedback. When gas is exhausted, star formation is quenched, leading to a convergence and stabilization of σ_{M_*} across the different runs (Gas is not as exhausted for HIGH feedback case as is the case for other cases, thus σ_{M_*} does not plateau).

An inverse correlation can be observed between the black hole mass scatter, $\sigma_{M_{bh}}$, and the strength of black hole feedback in Fig. 12 (right panel). In simulations with more efficient feedback, the final black hole masses exhibit significantly reduced scatter. For instance, in the HIGH-feedback configuration, $\sigma_{M_{bh}}$ is constrained to ≈ 0.5 , whereas in the unregulated feedback scenario, $\sigma_{M_{bh}} \approx 1$. Although the latter represents an unphysical extreme, it highlights the critical role of feedback in regulating black hole growth. Despite the scatter in instantaneous accretion rates being comparable across different feedback strengths, the cumulative effect of suppressed feedback leads to excessively high accretion over cosmic time (See Fig. D.19 right panel in D as well as variation in galaxy cold gas mass provided in the left panel). Consequently, black holes in the ZERO-feedback case grow to masses that are, on average, two orders of magnitude larger than those in regulated runs. The unrealistic massive ZERO-black holes accrete all of the cold gas, thereby suppressing the star formation. This results in a reduction of σ_{M_*} which explains why σ_{M_*} in the ZERO-feedback case is not as elevated as might otherwise be expected (Fig. 12 left panel).

Further details: the overall trend in the ZERO-feedback case reflects two competing effects. For galaxies with halo masses in the range $[12.0 < \log M_{SUB}(M_\odot) < 13.0]$, black holes grow sufficiently massive to

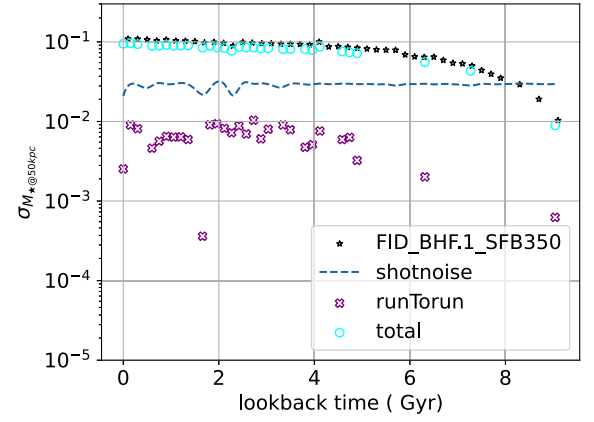
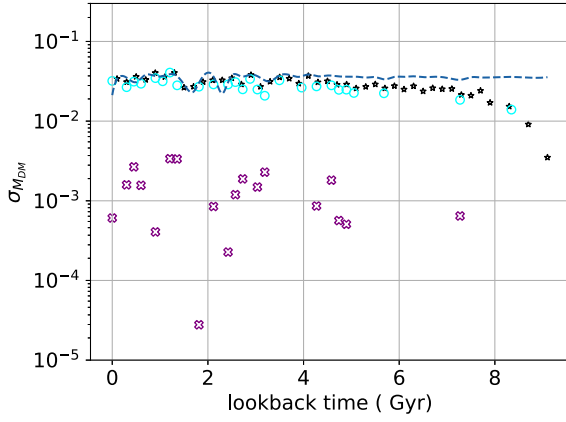


Fig. 8. Same as Fig. 5, but for Fiducial simulations including black holes. Variation in galaxy dark matter mass (left panel) and stellar mass (right panel) is shown.

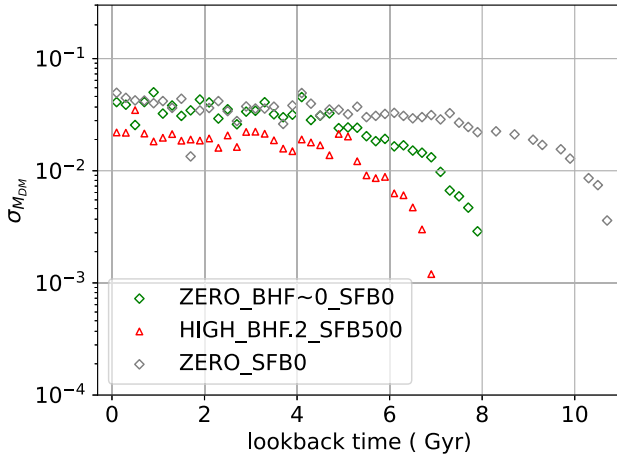


Fig. 9. Total variation (only Method 1) in galaxy dark matter mass as a function of lookback time (Gyr), shown for the feedback models: ZERO, HIGH feedback with black holes (colored lines and markers as indicated in the legend), and ZERO feedback without black hole (in gray) provided for comparison.

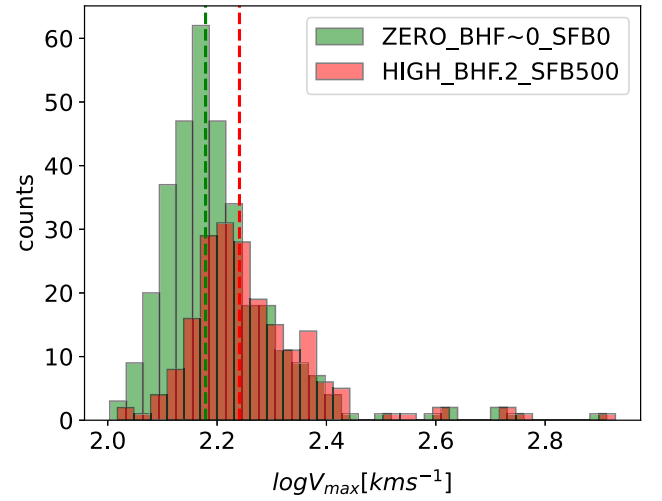


Fig. 11. Histogram of maximum circular velocities at $z = 0$ for BCGs in one of the runs of the ZERO and HIGH feedback respectively. Dashed vertical lines indicate the mean $V_{\max}$ for each distribution.

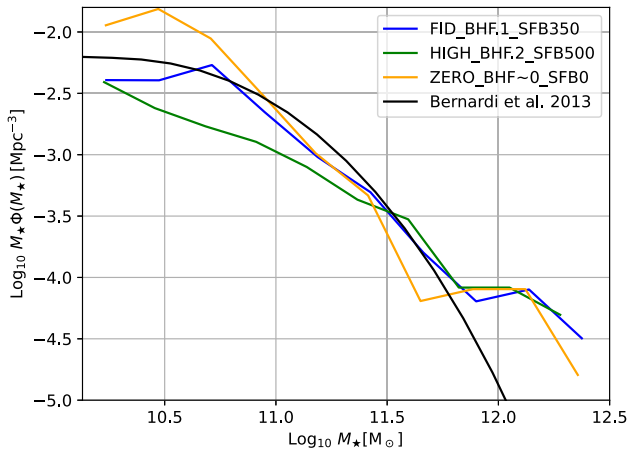


Fig. 10. GSMF for the feedback models: FID, ZERO, HIGH feedback at $z = 0$. GSMF is normalized as described in Bassini et al. (2020). Observational data from Bernardi et al. (2013) in black.

strongly suppress star formation, lowering the apparent scatter. In this regime, the behavior resembles the ZERO-NOBH or LOW-NOBH case, though here the suppression arises from the high black hole accretion and gas exhaustion rather than from unregulated star formation. At lower halo masses [$M_{SUB}(M_{\odot}) < 12.0$], star formation still competes effectively for the available gas. This manifests as a set of outliers in the ZERO-feedback runs that drive $\sigma_{M_{bh}}$ to higher values, even though the bulk of the distribution remains comparable to the HIGH and FID cases (with $\sigma_{M_{bh}} \approx 0.5$).

5.2. Change in precision

The choice of numerical precision in simulations reflects a balance between computational cost and the control of accumulated round-off error. In systems where higher-order terms or conservation properties play an important role, round-off error becomes increasingly relevant as the number of operations increases. For instance, in idealized tests of SPH density estimation, the intrinsic spatial accuracy of the method (e.g. $\approx 0.1\%$ density reconstruction accuracy with sufficient neighbors; see Dehnen and Aly (2012)) might suggest that lower numerical precision is adequate for some operations. However, this conclusion does not account for long-term integration effects in a complex, coupled system.

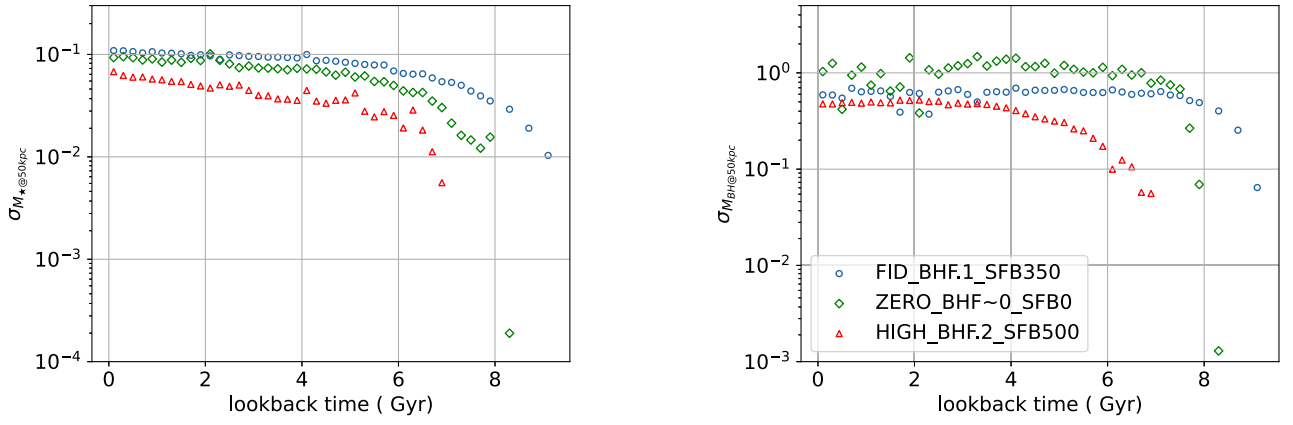


Fig. 12. Total variation (Method 1) in galaxy stellar mass (left panel) and black hole mass (right panel) as a function of lookback time (Gyr), shown for the feedback models: FID, ZERO and HIGH.

Our work utilizes cosmological simulations that are non-linear and feature a significantly more complex numerical setup including coupled gravity, adaptive time-stepping, a modern SPH solver, and stochastic subgrid models. In this environment, small numerical differences arising from round-off error compound over thousands of integration steps. To ensure numerical stability and minimize accumulated integration error, higher precision arithmetic is therefore required for key quantities.

Throughout the preceding sections of this study, we have utilized the full double precision (64-bit) configuration as our Fiducial baseline. Our rationale for this choice is to establish a more numerically stable environment. Having established this baseline, we now investigate the impact of a more computationally efficient mixed precision setup.

Mixed precision (MP) refers to the use of multiple floating point precision levels, typically, single precision (32-bit) and double precision (DP) within the same computational workflow. The purpose of this approach is to balance numerical accuracy with computational efficiency, given that some operations are performed in lower precision to significantly reduce memory usage and to increase speed. Critical calculations requiring higher numerical stability retain double precision: e.g. particle positions, variables in SPH calculations such as entropy, pressure gradients. Gravitational force calculations (accumulation of forces in long range and direct summation), other quantities like particle mass, velocity updates are in single precision.

Given its advantages in computational speed and reduced memory footprint, mixed precision arithmetic is currently widely adopted in simulations using `OPENGADGET3`. In this section, we investigate how mixed precision impacts variability in simulation outcomes.

Fig. 13 compares variation in galaxy stellar mass and cold gas mass (variation in galaxy dark matter mass and black hole mass provided in **Fig. E.20**) for the Fiducial simulation set including black holes run in double precision (already discussed) with its mixed precision counterpart. In addition to the change in the numerical precision, the snapshots in the latter are also stored with reduced numerical precision.

The difference in round-off errors between the two simulations sets are evident with the errors accumulating more rapidly in the mixed precision runs well before ≈ 9 Gyr. σ_{M_*} (**Fig. 13** left panel) plateaus to 0.1; similar to, but with a slight enhancement over the double precision scatter. Note that we do not observe a corresponding increase in the variation of cold gas mass (**Fig. 13** right panel, $\sigma_{M_{\text{cold}}} \approx 1$) or in the instantaneous star formation rates.

An enhancement in mixed precision σ_{M_*} can be easily visualized in **Figs. 14**, where the stellar masses of the most massive BCGs at $z = 0$ are shown for each realization in both the DP (**Fig. 14** left panel) and MP (**Fig. 14** right panel) test sets. Compared to DP, the increased scatter among the fourth most massive galaxies is starkly evident in case of MP runs. The time evolution of this divergence is illustrated in **Fig.**

15. Here, a selection of the most massive galaxy stellar masses are plotted, each presented with vertical error bars indicating the variation among clones over a ~ 1.4 Gyr interval in the MP runs (different colors maintained for galaxies across the snapshots). The figure demonstrates that, in this case, the divergence in stellar mass is largely persistent over time, suggesting a systematic growth of variability.

The effect occurs because one clone (Clone 1) hosts a black hole at $z \approx 3.42$, while another (Clone 2) does not. When Clone 2 is finally seeded (at $z \approx 2.31$), its black hole accretes rapidly and largely catches up to the other clones by the end of the simulation, with final black hole masses remaining of the same order. In contrast, the impact on stellar mass is more pronounced due to the extended period during which Clone 2 evolves without a black hole. Stellar mass variation increases during this interval and continues until the black hole in Clone 2 grows sufficiently massive ($z \approx 1.7$) for feedback to stymie further growth. By this stage, the stellar mass in Clone 2 has already grown substantially relative to the other clones. As a result, $\sigma_{M_{bh}}$ remains relatively unaffected compared to the corresponding higher variation in stellar mass (see **Fig. E.20**, left panel).

Though outwardly an interesting case, the root cause of this divergence has been attributed to a too-aggressive precision truncation in the Barnes-Hut MPI exchange of accelerations. The newer version of `OPENGADGET3` alleviates this issue by adopting a larger number of double precision quantities compared to our older version of the code. We specifically carried out tests in mixed precision, since this mode is widely used and forms the basis of the long-standing “PROD” version of the code employed in many previous studies. The significantly large divergence observed in this configuration implies that results from such studies are also affected; thus, while large-scale properties remain robust, ‘error bars’ for individual object properties would be broader compared to double precision simulations. Consequently, researchers analyzing such simulations of individual galaxies must exercise caution, as their results are subject to greater numerical sensitivity and potentially larger uncertainties than those found in higher precision setups. The recently implemented change is expected to mitigate this behavior, although we have not independently verified whether they are directly related to the divergence we observe.

6. Summary and conclusions

We investigated variability in galaxy cluster simulations using `OPENGADGET3` code. Identical simulations at the lowest resolution setting (1x) for one of the Dianoga galaxy clusters were executed. Individual galaxies (only central galaxies) were matched across these runs, and variations in their properties were analyzed using two statistical methods: pooled standard deviation (Method 1) and a mixed linear model (Method 2). Here we list some of our key results:

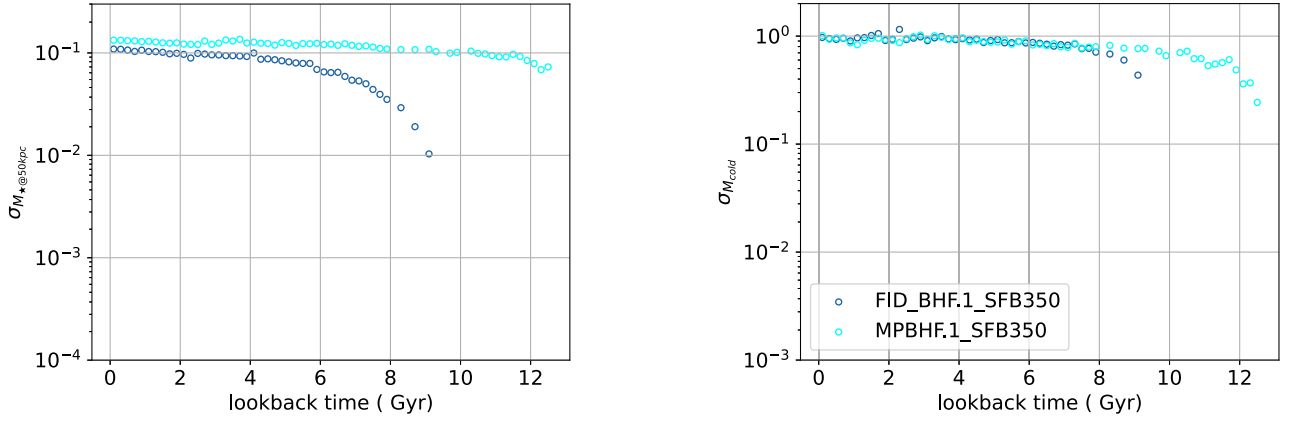


Fig. 13. Total variation (Method 1) in galaxy stellar mass (left panel) and cold gas mass (right panel); as a function of lookback time (Gyr), shown for Fiducial double (in blue) and mixed precision sets (MP, in cyan).

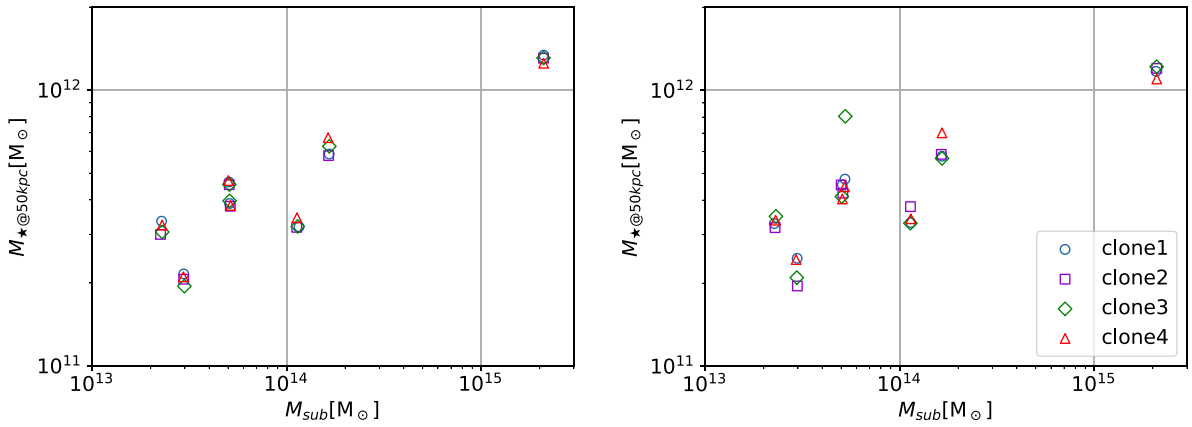


Fig. 14. Stellar mass versus subhalo mass correlation for few of the most massive BCGs in Fiducial double precision runs (left panel) and mixed precision runs (right panel) at $z = 0$. Each identical run in the respective set is represented by a different marker and color.

Without black holes.

- For our *Fiducial* model: We find variations of ($\sigma_{M_{\text{dm}}} = 0.03\text{--}0.05$) and ($\sigma_{M_*} = 0.04\text{--}0.08$) (Fig. 5; relatively higher variation in stellar mass is attributed to shot noise and the influence of highly non-linear processes like star formation and feedback).
- Feedback tests (see Table 2): Negligible effect on the variation in dark matter mass (Top panel of Fig. 6) but significantly impacts the variation in stellar mass (Middle panel of Fig. 6). Higher stellar feedback has increased scatter ($\sigma_{M_*} \approx 0.08\text{--}0.1$) due to increased stochastic variation (results consistent with Keller et al. (2019)).

With black holes.

- For our *Fiducial* model: yields a $\sigma_{M_{\text{dm}}} \approx 0.04$ (Fig. 8; consistent with its Fiducial no-black-hole counterpart). However, σ_{M_*} saturates at a significantly higher value of ≈ 0.1 , indicating increased sensitivity of star formation in the presence of black holes.
- Feedback tests (see Table 2): The HIGH-feedback model exhibits a reduced scatter, with $\sigma_{M_*} \approx 0.07\text{--}0.08$ and a slight decrease in $\sigma_{M_{\text{bh}}} \approx 0.5$, compared to the Fiducial model ($\sigma_{M_*} = 0.1$ and $\sigma_{M_{\text{bh}}} \approx 0.6$; see Fig. 12).

Our results indicate the regulatory role of feedback in moderating both star formation and black hole growth. However, disentangling the relative contributions of feedback, subgrid physics, and stochastic effects to the overall variability is inherently complex. For instance, Genel et al. (2019) reported enhanced variability in simulations with feedback, though part of this trend may reflect differences in the inclusion

of black holes between their feedback and no-feedback runs. Further parameter exploration will be valuable to better characterize these dependencies and test the generality of our findings.

Run-to-run variation is minimal ($\sigma \lesssim 10^{-2}$), indicating that inherent numerical noise within each simulation is the dominant source of variability. At the resolution used, the simulations are stable with respect to run-to-run divergence, and the observed variations are primarily due to stochastic noise and shot noise. In practical terms, this means that repeated runs converge toward an expected mean behavior, rather than representing distinct but equally valid realizations of the system. A higher systematic variation would instead imply that each simulation outcome could diverge meaningfully over time. Differentiating between these two regimes (noise-dominated or systematically divergent) is crucial for assessing the reproducibility of cosmological simulations. We refrain from quantifying such distinctions here, as this would require a more complete set of tests, but the framework serves as a potential tool for future studies aimed at assessing the limits of reproducibility and predictive reliability in cosmological simulations.

The results presented here are specific to the configuration and feedback implementation of the `OPENGADGET3` simulation, emphasizing the need for careful consideration when extrapolating findings to other simulation setups. To strengthen the generalization of our conclusions, future studies will include resolution-based convergence tests to assess the stability of these variation patterns at higher resolutions. Mergers, an important source of variability in galaxy properties, were not included in this analysis but represent a key factor that will be addressed in subsequent research to provide a complete picture of variability in galaxy cluster simulations.

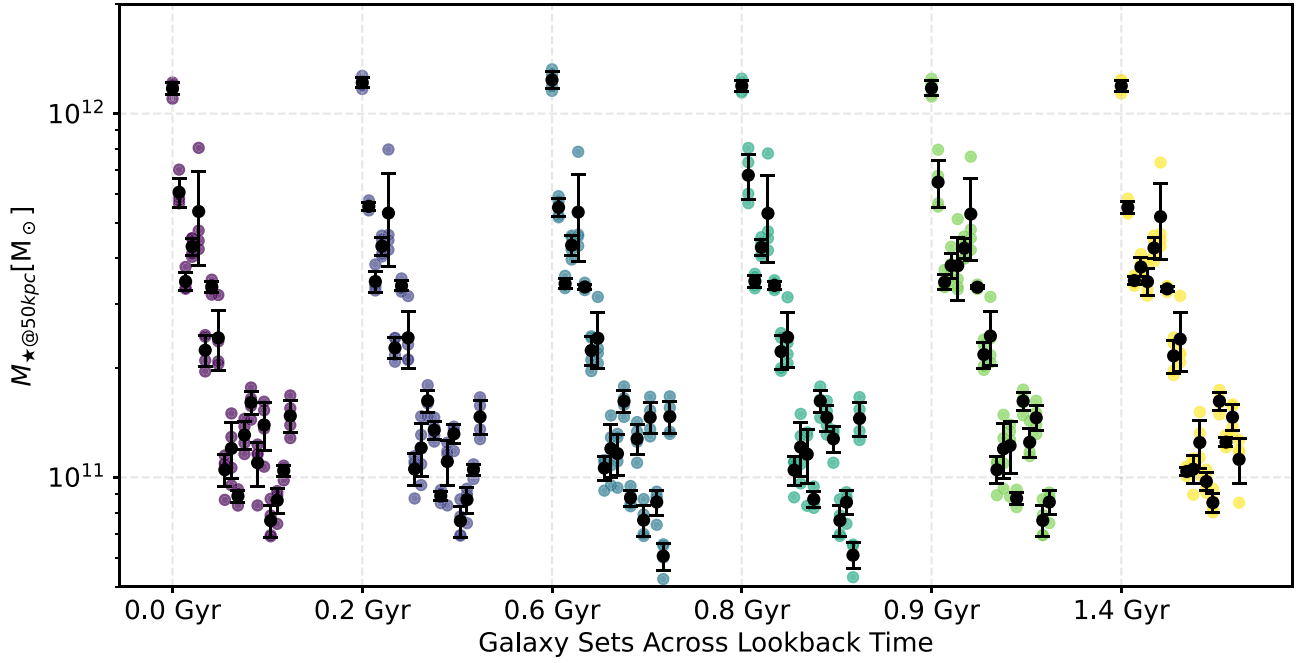


Fig. 15. Stellar masses of the N most massive galaxies across a ~ 1.4 Gyr interval, sampled at five snapshots for mixed precision simulations. Matched galaxies are plotted and for each matched galaxy set, the points indicate the stellar mass from the representative clones, while the vertical error bars roughly illustrate the variability between each set. Different colors correspond to individual galaxies, and the same color is maintained across snapshots to allow visual tracking of whether a galaxy's relative scatter persists over time or fluctuates. This highlights persistent differences (rather than fluctuating) in stellar mass in matched galaxies across identical simulations.

In conclusion, this study provides a benchmark for the expected magnitude of variation in simulations run with `OPENGADGET3`. The results highlight the importance of considering repeated-run variability when interpreting simulation outcomes, especially for properties sensitive to highly non-linear processes like star formation, black hole dynamics and feedback. Understanding these variations is crucial for accurately comparing simulation results with observational data and improving the reliability of their predictions.

Future work will extend this analysis to higher-resolution simulations; additionally, the exploration of a broader range of subgrid models and feedback parameters will help clarify the conditions under which feedback mitigates or exacerbates variability, thereby enhancing our understanding and interpretation of results of galaxy evolution in cosmological simulations.

CRediT authorship contribution statement

Chaitra: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation. **Antonio Ragagnin:** Writing – review & editing, Software. **Milena Valentini:** Writing – review & editing. **Giuseppe Murante:** Writing – review & editing. **Stefano Borgani:** Project administration. **Giuliano Taffoni:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

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

Acknowledgments

We appreciate the anonymous referees for their helpful suggestions. We sincerely thank Francesco Pauli for helpful discussions and timely

advice, which greatly contributed to clarifying statistical methods used in this study. This work has been supported by the Spoke-3 “FutureHPC & BigData” of the ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing – and hosting entity, funded by European Union – NextGenerationEU. Supported by Italian Research Center on High Performance Computing Big Data and Quantum Computing (ICSC), project funded by European Union – NextGenerationEU – and National Recovery and Resilience Plan (NRRP) – Mission 4 Component 2 within the activities of Spoke 2 (Fundamental Research and Space Economy), (CN 00000013 - CUP C53C22000350006). This work has also been supported by the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 1409 published on 14.9.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU – Project Title “Space-based cosmology with Euclid: the role of High-Performance Computing” – CUP J53D23019100001 - Grant Assignment Decree No. 962 adopted on 30/06/2023 by the Italian Ministry of University and Research (MUR). AR acknowledges EuroHPC Joint Undertaking for awarding the project ID EHPC-REG-2024R01-029 access to Leonardo at CINECA, Italy. AR acknowledges ISCRA for awarding this project access to the LEONARDO supercomputer, owned by the EuroHPC Joint Undertaking, hosted by CINECA, Italy (HP10BUFI59). This work is also supported by “Bando per il finanziamento della Ricerca Fondamentale 2024 dell’Istituto Nazionale di Astrofisica” Decreto n. 8/2024 MiniGrant “Numerical Chaos in Cosmological Simulations”. MV also acknowledges partial financial support from the INFN Indark Grant .

Appendix A. Mechanisms contributing to variability and sensitivity in cosmological simulations

The sensitivity of cosmological simulations to divergent outcomes arises from several distinct mechanisms. While some represent fundamental numerical uncertainties (such as discretization errors, precision errors), others (such as subgrid physics models) act as triggers that amplify small perturbations into macro-scale variations.

We emphasize a distinction established in previous studies regarding bit-wise reproducibility: while “numerical determinism” can be enforced via specific implementation strategies (such as strict operation ordering or 128-bit summation with subsequent truncation), these methods impose significant computational overhead. Additionally, for complex subgrid models such as MUPPI (Murante et al., 2010, 2014), which utilize Runge–Kutta integration with adaptive time steps, achieving such reproducibility is technically difficult and often impractical compared to simpler threshold-based subgrid schemes. Ultimately, such determinism is primarily a tool for debugging and code verification. It does not imply that a physical solution is unique or insensitive to perturbations. It simply means that the same numerical approximations are reproduced across runs; however, when execution conditions differ (e.g. MPI task configuration, checkpoint–restart cycles, or hardware architecture), small perturbations can arise and may later be amplified by nonlinear dynamics. As demonstrated by Genel et al. (2019), even when a code produces byte-identical results, the introduction of infinitesimal perturbations to the initial conditions leads to divergent galaxy properties.

This section establishes a systematic framework for understanding different mechanisms of simulation unpredictability with a particular focus on cosmological simulations (see e.g. Hernquist et al. (1993), Binney and Tremaine (2008), Barnes and Hut (1989)).

Poisson noise from finite particle sampling. In particle-based simulations, continuous matter fields are represented by discrete particles, introducing statistical fluctuations in measured properties; commonly referred to as shot noise. This represents a fundamental precision limit that depends solely on resolution, not on the underlying physics.

Numerical method and precision effects.

1. Discretization Errors: Approximating continuous equations with discrete time steps or spatial grids introduces errors that depend on resolution choices. Large time steps may fail to capture system dynamics accurately.
2. Round-off Errors: Finite precision computer arithmetic accumulates small errors at each computational step, particularly problematic in long-term integrations of sensitive systems.
3. Truncation Errors: Approximating infinite series or integrals by finite sums can introduce significant errors if higher-order terms are important but neglected.
4. Inadequate Resolution: Insufficient spatial or temporal resolution can lead to unphysical results, such as failing to resolve shock waves in hydrodynamic simulations or missing important small-scale physics.
5. Parallel Execution and Operation Ordering: Because floating-point arithmetic is non-associative, differences in the order of operations (e.g. in distributed computing, multi-threading, vectorization) can introduce small but measurable numerical differences between runs.

N-body specific numerical effects. Gravitational simulations face additional technical challenges:

1. Force Softening: The softening length ϵ prevents numerical divergences from close encounters but introduces a trade-off between stability and accuracy. Inappropriate choices can either over-smooth important physics or reintroduce numerical instabilities.
2. Truncation of Multipole Expansion: Tree codes approximate distant forces using multipole expansions. While typically well-controlled, these approximations can contribute to accumulated errors.
3. Improper Boundary Conditions: Incorrectly specified boundaries can introduce spurious artifacts that destabilize the simulation.

Probabilistic subgrid implementations. Modern cosmological simulations incorporate probabilistic algorithms to model physical processes below the resolution limit, such as star formation, stellar feedback, and black hole accretion. While these models are typically implemented using random number generators (RNGs) with fixed seeds to ensure numerical determinism, their formulation involves probabilistic decision criteria. For example, small numerical perturbations can shift a gas particle’s state across a star formation threshold, leading to differences in the timing or ordering of star formation events between realizations. In nonlinear regimes, such differences can be amplified, allowing small-scale perturbations to manifest as macroscopic variations between otherwise identical runs.

Event-driven amplification. Discrete nonlinear events (such as galaxy mergers, major feedback episodes, or dynamical instabilities) can amplify small pre-existing differences into large, persistent variations. These effects produce step-function changes in system evolution that persist over long timescales (Keller et al., 2019).

Chaos. Chaotic behavior refers to the sensitivity of a deterministic system to small perturbations in its initial conditions, which can propagate and lead to complex, large-scale changes in its evolution (Studies of chaos was pioneered by Lorenz (1963), Mandelbrot (1982), May (1976), Feigenbaum (1978)). This phenomenon, termed deterministic chaos, is characterized by positive Lyapunov exponents (Strogatz, 2019). Numerical simulations are inherently chaotic systems, and their long-term predictability is fundamentally limited by the amplification of small perturbations (Aarseth, 1963; Miller, 1964; Standish, 1968; Henon and Heiles, 1964). It was found that the Lyapunov timescale (λ^{-1}) generally decreases with increasing N , but the trend saturates for $N \gtrsim 30$, indicating a regime where the divergence rate becomes largely independent of particle number (The question of how chaotic behavior scales with N , referred to as “ N -dependence”, see some early works for example, Gurzadian and Savvidi (1984), Gurzadian and Savvidy (1986), Heggie (1991), Kandrup (1989, 1990b,a)).

A.1. Variability studies in cosmological simulation

Systematic quantification of variability is not yet standard practice when reporting simulation results; examples of studies that have explicitly focused on quantifying or establishing baselines for such effects include (Su et al., 2017, 2018; Oh et al., 2020; Granato et al., 2021; Keller and Kruijssen, 2022; Valentini et al., 2023). Studying variability is particularly important in cosmological simulations, as it arises from the complex nonlinear interplay of gravity, hydrodynamics, and probabilistic subgrid physics such as star formation and feedback. These processes, together with mergers and environmental influences, can amplify small perturbations and lead to divergent galaxy properties across otherwise identical runs.

Davies et al. (2021, 2022, 2023) examined the role of assembly history on galactic evolution using the EAGLE model (Schaye et al., 2015). They used a technique called genetic modification (Roth et al., 2016; Rey and Pontzen, 2018; Hoffman and Ribak, 1991) to create different assembly histories for the galaxies. They found that mergers can significantly affect black hole (BH) growth and feedback, which in turn impacts the circumgalactic medium (CGM). The 2022 and 2023 follow-up studies expanded on the earlier work by focusing on the dynamic interaction between galaxy mergers, BH activity, and the CGM. Through more detailed simulations, they revealed how the timing and intensity of mergers influence the thermodynamic properties of the CGM (resulting in order-of-magnitude variations between their simulations) and the subsequent evolutionary path of the galaxy. The results suggest that understanding the subtleties of merger histories is crucial for accurately modeling galaxy formation and evolution, especially in relation to BH feedback and its long-term effects on galaxy properties.

One of the earliest systematic studies of variability within the context of cosmological simulations was conducted by Thiébaud et al.

(2008), who demonstrated that in Λ CDM cosmologies, chaos arises predominantly at nonlinear scales. Using simulations of $100 h^{-1}$ Mpc boxes with perturbed initial particle positions and velocities, they quantified the critical transition scale for simulation variability to be around $3.5 h^{-1}$ Mpc. On these scales, integrative quantities such as halo mass remained robust, while properties like spin orientation exhibited sensitivity to initial conditions.

Building on these ideas, Genel et al. (2019) performed analysis of variability in cosmological volumes ($25\text{--}50 \text{ Mpc}^3/h$) using the AREPO code and the full IllustrisTNG feedback model. Their resolution-scaled simulations demonstrated that tiny perturbations ($\approx 10^{-7}$ pc) in initial positions result in divergent galaxy properties of 2%–25% over cosmic time. They established that in simulations with full feedback, variability persists even at high resolution, whereas in no-feedback runs, variations reduce with resolution. Additionally, they conclude that the repeated-run divergence contributes significantly to the overall scatter in key scaling relations such as the Tully-Fisher and star formation rate–mass relations.

Keller et al. (2019) quantified variability in simulated galaxies by analyzing 128 identical simulations using GASOLINE2 and RAMSES codes. In Milky Way-like galaxies with $M_{\text{halo}} \approx 6.5 \times 10^{11} M_{\odot}$, they found a typical scatter in stellar mass $\sigma_{M_*} \approx 5\%$ due to feedback-regulated star formation. Importantly, they noted that in the absence of feedback or under unregulated star formation, variations grow substantially, especially when gas is not rapidly exhausted. Additionally, the timing of mergers was shown to cause stellar mass variations by factors of 2, persisting for over a Gyr, thus linking such divergence to long-term galactic histories.

Borrow et al. (2023) performed analysis with 16 identical simulations using the SWIFT code with cosmological volumes (25 Mpc^3), finding stellar mass variations up to 25%. Their results confirmed that while population-wide scaling relations remain robust, individual galaxy evolution, especially during “bursty” feedback or black hole accretion, can diverge significantly. Notably, they illustrated how random variability causes the correlation between galaxy stellar mass and black hole mass to essentially disappear at the highest black hole masses ($M_{\text{BH}} > 10^8 M_{\odot}$), underscoring the influence of non-linear feedback on black hole and galaxy co-evolution.

Pakmor et al. (2025) investigated the variability of the Auriga galaxy formation model using seven Milky Way-like zoom-in realizations, differing only by random seeds. They found most global properties at $z = 0$ (e.g. stellar mass, disc mass, star formation history) to be robust within $\leq 10\%$, while instantaneous quantities such as star formation rate and internal disc structure showed larger variations (up to a factor of two). Additionally, they found that although satellite infall properties were consistent across runs, their subsequent orbital evolution diverged after the first pericentric passage. Importantly, the study further showed that changes in numerical resolution (by a factor of eight) introduced far greater and more systematic differences than the intrinsic stochastic variability.

These studies collectively demonstrate that cosmological simulations exhibit substantial variability arising from multiple, often coupled, sources. Isolating the contribution of each source individually is frequently infeasible and beyond the scope of most production simulation setups. Accordingly, A is intended to provide context for the inherent numerical and physical complexity of cosmological simulations, motivating our analysis of the cumulative impact of these factors on overall variability in simulation results.

Appendix B. Variability in DMO and hydro-only simulations

In this section, we demonstrate the impact of intrinsic numerical variability by excluding the stochasticity associated with the subgrid physics used in our full-physics runs. We performed controlled tests for Dark Matter Only (DMO) and Dark Matter plus hydrodynamics (Hydro) simulations using our low-resolution setup (see Table 1) with

a mixed precision configuration. For each test set, we evolved three realizations using identical initial conditions and numerical parameters. Two runs executed with identical MPI task counts (144 tasks) yielded byte-identical results. A third run was evolved using a different domain decomposition (180 MPI tasks), introducing small perturbations in the ordering of floating-point operations relative to the 144-task run. Fig. B.16 provides a visual comparison of the most massive halo at $z = 0$ between the 144-task and 180-task runs for both DMO (top panel) and Hydro (bottom panel) cases. In contrast to the full-physics simulations presented in the main text (Fig. 1), these realizations exhibit no visually discernible differences in their substructure. Furthermore, Fig. B.17 shows the Halo Mass Functions (HMFs) for the DMO (left panel) and Hydro (right panel) sets, with the relative difference between realizations provided in the lower panels. Both HMFs are statistically indistinguishable across the entire mass range, demonstrating that these outcomes are highly consistent and show low sensitivity to round-off level perturbations.

Appendix C. Tests with varied MPI task counts

To assess the sensitivity of the results to small numerical perturbations at the outset, we repeated the Fiducial simulations (excluding black holes) with different MPI task decompositions. This set comprised four runs, executed with 144, 180, 216, and 252 MPI tasks, respectively. In Fig. C.18, σ (computed with Method 1) for the varied-MPI-task Fiducial set is compared to the baseline Fiducial set performed with identical MPI task counts.

For σ_{M_*} (right panel), the varied-MPI-task-runs exhibit an earlier onset of measurable variation, beginning at lookback times $> 12.5 \text{ Gyr}$, and $\sigma_{M_*} > 0.2$ that coincides with the higher shot noise and increased stochastic noise which can be attributed to peak of star formation. The variation then decreases and subsequently plateaus, approaching a level similar to the unperturbed baseline but remaining slightly higher (the difference likely reflects additional noise introduced by the altered MPI decomposition).

In contrast, the variation in dark matter mass (left panel of Fig. C.18) shows no marked difference between the perturbed and unperturbed runs: though the variations in the varied-MPI-task set accumulate from $> 12.5 \text{ Gyr}$, both plateau at nearly identical levels. The lack of sensitivity to small numerical perturbations reflects the inherently stable nature of collisionless dark matter dynamics. For stellar mass, however, excess round-off error can act as an additional noise contribution, which may compound with the stochasticity of the star formation and feedback subgrid models.

Appendix D. Variability including black holes: Galaxy cold gas mass and black hole accretion rate

We show here the total variation (Method 1) in galaxy cold gas mass (D.19 left panel) and black hole accretion rate (BHMD, D.19 right panel) for the Fiducial, ZERO and HIGH feedback cases.

The variation in galaxy cold gas mass is broadly similar across all feedback configurations, plateauing at $\sigma_{M_{\text{cold}}} \approx 1$. For BHMD, though variation in black hole mass accretion exhibits a slightly lower scatter for the ZERO-feedback case, likely due to the presence of more massive black holes that largely depleted their gas reservoirs; overall, $\sigma_{\text{bhmd}} \approx 4\text{--}6$ across all cases. These variations demonstrate that black hole growth remains a highly stochastic process even when feedback strength varies. Note that these quantities are highly variable because they are instantaneous rather than time-averaged and are strongly influenced by short-term environmental changes.

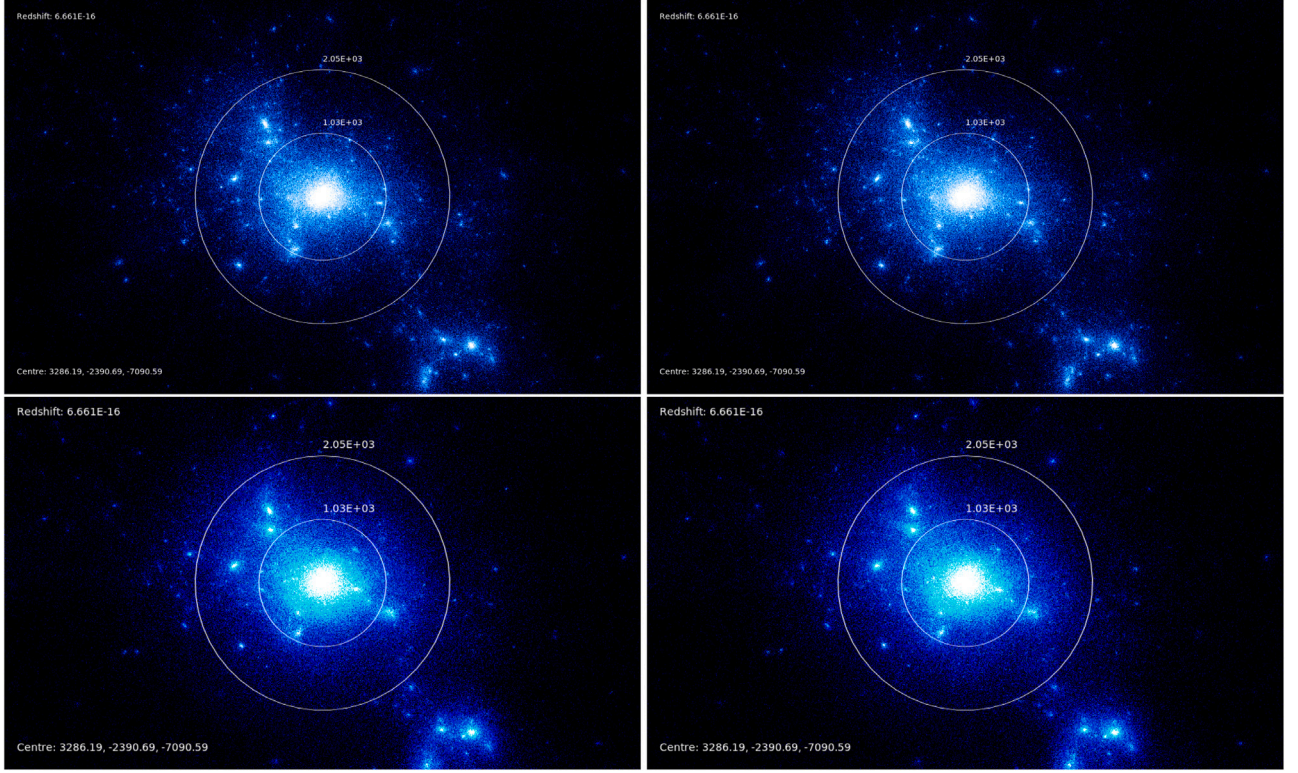


Fig. B.16. Visual comparison of the Dianoga D1 (see Table 1 and Section 2) dark-matter-only (top panels) and dark-matter and hydrodynamical simulations (bottom panels) at $z = 0$, each with two realizations evolved from nearly identical set-up. The left panels shows the runs executed with 144 MPI tasks, while the right panels shows the run executed with 180 MPI tasks. No visually discernible differences are observed in the positions or morphology of the halo and its substructures in the respective repeated runs for dark-matter-only and as well as dark-matter and hydrodynamical simulations.

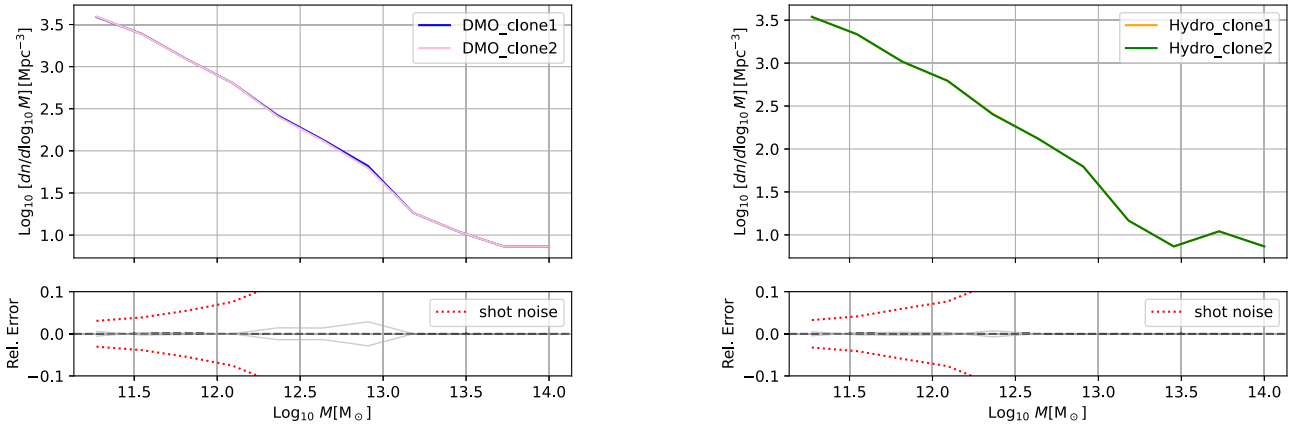


Fig. B.17. Top panel shows Halo mass functions (HMFs) for the dark-matter-only (left panel, prefixed “DMO”) and dark-matter and hydrodynamical simulations (right panel, prefixed “Hydro”). Each test set is comprised of 2 identical runs, executed with 144 MPI tasks (DMO_clone1 and Hydro_clone1) and 180 MPI tasks (DMO_clone2 and Hydro_clone2). The bottom panel shows the relative difference between the respective identical-realizations (gray) and the shot noise (red dotted line), demonstrating statistical agreement across the resolved mass range.

Appendix E. Mixed precision variability: Galaxy dark matter and black hole mass

This Appendix provides supplementary figures in support of Section 5.2. Fig. E.20 displays the total variation (σ) computed via Method 1 for galaxy dark matter mass (left panel) and black hole mass (right panel) across both the double and mixed precision test sets. While the accumulation of round-off errors occurs more rapidly in the mixed precision runs (well before 9 Gyr), the resulting variability in these specific properties is not significantly varied. Both $\sigma_{M_{\text{dm}}}$ and $\sigma_{M_{\text{mbh}}}$

plateau at levels nearly identical to their double precision counterparts, demonstrating minimal sensitivity to numerical precision effects.

Appendix F. Subgrid physics implementation details

F.1. Star formation and associated feedback

Star formation and stellar feedback are incorporated through a set of subgrid models that convert gas into stars and model the impact

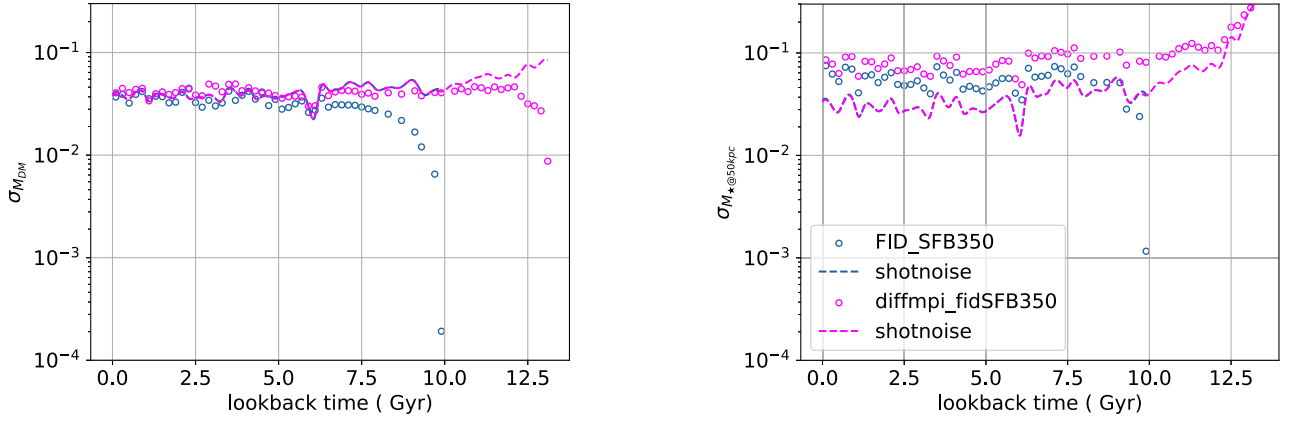


Fig. C.18. Total variation (Method 1) in galaxy dark matter mass (left panel) and stellar mass (right panel) as a function of lookback time (Gyr), shown for the Fiducial-NOBH model. The blue lines represent the standard Fiducial set (fixed MPI tasks, as in Fig. 6), while the pink lines represent the varied MPI task set (label prefixed with 'diffmpi_fid'). Dashed colored lines indicate the approximate shot noise floor for each test set.

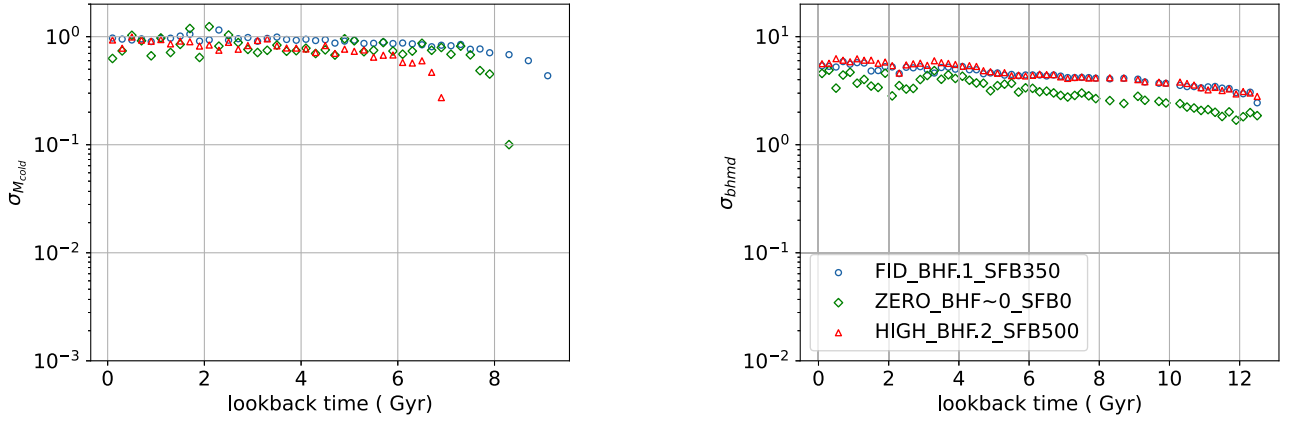


Fig. D.19. Same as Fig. 12 but for galaxy cold gas mass (left panel) and black hole accretion rate (right panel).

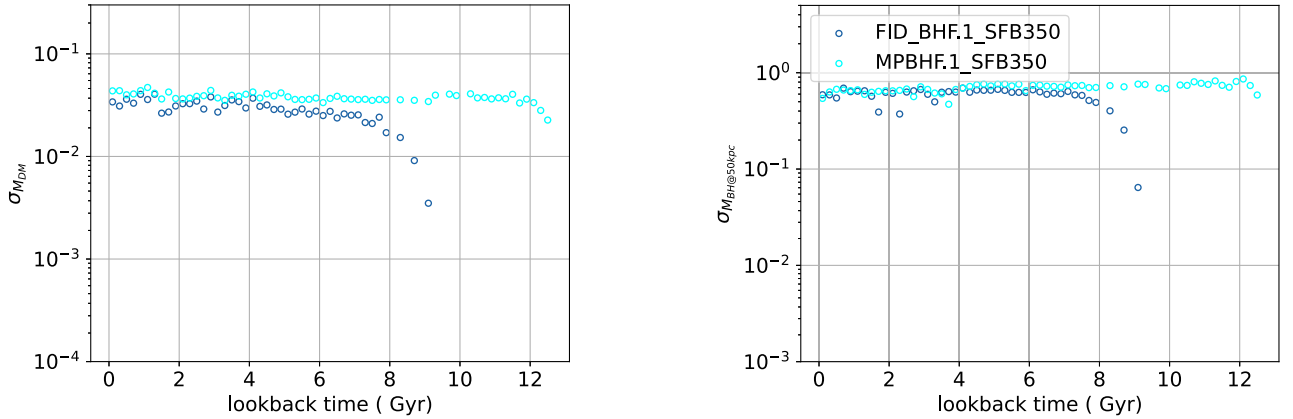


Fig. E.20. Total variation (Method 1) in galaxy dark matter mass (left panel) and black hole mass (right panel); as a function of lookback time (Gyr), shown for Fiducial double (in blue) and mixed precision sets (MP, in cyan).

of stellar feedback on the surrounding gas. For the reference model of star formation and feedback implemented, please check (Springel and Hernquist, 2003; Tornatore et al., 2003, 2007). Here we only summarize some implementation details.

The Interstellar Medium (ISM) refers to the diffuse matter i.e. gas (atomic, molecular, and ionized) and dust filling the space between

stars within a galaxy. The ISM plays a critical role in star formation, as stars form from the gravitational collapse of dense regions within the ISM. This aspect is modeled with gas particles, where a given gas particle is defined to be multiphase with respect to a certain density threshold and temperature threshold (which defines a hot or cold phase). Thus,

- A density threshold: A gas particle can form stars if its density exceeds a threshold value ρ_{th} . This threshold is chosen from Schmidt law-like prescription and derived from density function sustaining a self-regulated star formation regime.
- A temperature criterion: Gas particles eligible for star formation must also be sufficiently cool, typically below a temperature threshold T_{th} . This ensures that only cold, dense gas, which is capable of collapsing under its own gravity, forms stars.

Star formation rate (SFR). quantifies the rate at which a galaxy forms new stars over a given period of time, typically expressed in solar masses per year ($M_{\odot} \text{ yr}^{-1}$). In GADGET-3, once the particle is converted to a star particle it follows a Schmidt law-like prescription, where the rate at which gas converts into stars is proportional to the gas density. The instantaneous SFR density for a gas particle i is given by:

$$\dot{\rho}_{\star,i} = (1 - \beta) \frac{\rho_i}{t_{\star}} \quad (\text{F.1})$$

where:

- $\dot{\rho}_{\star,i}$ is the star formation rate density,
- β is the mass fraction of massive stars (dependent on the Initial Mass Function, IMF),
- ρ_i is the gas density of particle i ,
- $t_{\star} = \frac{t_{\text{dyn}}}{c_{\star}}$ is the star formation timescale configured according to Schmidt's law (Schmidt, 1959).

IMF used in this work is Chabrier (2003). The actual process of star formation is implemented stochastically, converting gas particles into star particles according to the computed SFR.

Stellar feedback. There are a variety of stellar feedback models implemented in GADGET-3. Below, we focus only on the ones presented by Springel and Hernquist (2003), Tornatore et al. (2007). Stellar feedback in GADGET-3 encompasses thermal and kinetic feedback, including primarily supernova explosions, winds, and chemical enrichment. These processes inject energy, momentum, and heavy elements into the surrounding gas, influencing its thermal and dynamical state. The thermal energy released is computed based on the IMF adopted. The feedback heats up the hot gas as well as evaporates the gas in the cold phase via thermal conduction. This recycles in the dense gas clouds back to the ambient phase. The evaporated cold gas mass follows,

$$\dot{\rho}_c = -A\beta \frac{\rho_c}{t_{\star}} \quad (\text{F.2})$$

where, A is the evaporation efficiency and assumed to follow a theoretically modeled dependence on the ISM density ($A \propto \rho^{-\frac{4}{5}}$)

Radiative cooling of the hot gas replenishes the cold gas clouds. The cycle continues sustaining a self-regulated process of star formation.

Instead of explicitly modeling the detailed mass exchange between phases (cold clouds, hot gas, and stars), the method assumes equilibrium conditions are achieved rapidly. The model evolves gas particles probabilistically into star particles based on the computed star formation rate. For each timestep, a new star particle is spawned if a random number drawn uniformly falls below the probability:

$$p_{\star} = \frac{m}{m_{\star}} \left(1 - \exp \left[-\frac{(1 - \beta)x\Delta t}{t_{\star}} \right] \right), \quad (\text{F.3})$$

where $m_{\star}$ is the mass of a single star particle, and x is the fraction of gas in cold clouds.

In the multiphase model for ISM supporting SF and feedback, winds are an additional implementation following the phenomenological prescription for galactic winds, also, as detailed in Springel and Hernquist (2003). Hot gas associated with supernovae is converted into wind particles following a mass loss proportional to the SFR :

$$\dot{M}_w = \eta \dot{M}_{\star} \quad (\text{F.4})$$

where: η is the efficiency parameter (typically $\eta = 2$ is assumed). This conversion depends on probability of a generated random number to be less than a limit of:

$$p_w = 1 - \exp \left[-\frac{\eta(1 - \beta)x\Delta t}{t_{\star}} \right] \quad (\text{F.5})$$

where:

- p_w is the probability that a gas particle is converted into a wind particle within the time interval Δt .
- Δt is the simulation time step.
- $t_{\star}$ is the star formation timescale.

Depending on the gradient of the gravitational potential of the galaxy, the wind particles are ejected orthogonally, along the rotational axis of the galaxy. The velocity of the winds are configured, a $v_w = 350 \text{ km s}^{-1}$ corresponds to a fixed fraction of 50% of the supernovae energy converted into the kinetic energy carried by the wind (Bassini, 2021).

Supernovae feedback is implemented together with metal enrichment following the prescriptions presented in Tornatore et al. (2007). A metal dependent cooling function is used to compute the respective energy radiated due to the quantity of metals produced. H and the by products He, C, Ca, O, N, Ne, Mg, S, Si, Fe, Na, Al, Ar are all followed and computed specific to the IMF and predictions from theoretical modeling respective to type-Ia, II and asymptotic giant branch (AGB) stars; see Matteucci and Padovani (1993), Thielemann et al. (2003), Woosley and Weaver (1995), Romano et al. (2010) and Karakas (2010).

F.2. Black hole and associated feedback

The base reference for black hole implementation is Springel et al. (2005). Black holes are initially seeded in dark matter halos once they reach a critical mass or are otherwise designated to host a BH. Typically, seed BHs have masses around $10^5 - 10^6 M_{\odot}$, placed in the center of newly formed halos.

The accretion rate onto a black hole is determined using the Bondi-Hoyle-Lyttleton formula (refer (Bondi, 1952)), modified for a realistic astrophysical context:

$$\dot{M}_{\text{BH}} = \alpha \frac{4\pi G^2 M_{\text{BH}}^2 \rho}{(c_s^2 + v^2)^{3/2}} \quad (\text{F.6})$$

where:

- $\dot{M}_{\text{BH}}$ is the accretion rate onto the black hole,
- α is a dimensionless parameter that accounts for the efficiency of the accretion process,
- G is the gravitational constant,
- M_{BH} is the black hole mass,
- ρ is the local gas density,
- c_s is the sound speed of the gas,
- v is the relative velocity between the black hole and the gas.

This accretion rate is capped by the Eddington limit to avoid super-Eddington accretion scenarios:

$$\dot{M}_{\text{Edd}} = \frac{4\pi G M_{\text{BH}} m_p}{\epsilon_r \sigma_T c} \quad (\text{F.7})$$

where:

- $\dot{M}_{\text{Edd}}$ is the Eddington accretion rate,
- m_p is the proton mass,
- ϵ_r is the radiative efficiency,
- σ_T is the Thomson scattering cross-section,
- c is the speed of light.

Feedback mechanisms. The energy feedback from accreting black holes is modeled as thermal energy injected into the surrounding gas. The rate of energy injection is given by:

$$\dot{E}_{\text{feed}} = \epsilon_f \epsilon_r \dot{M}_{\text{BH}} c^2 \quad (\text{F.8})$$

where:

- $\dot{E}_{\text{feed}}$ is the feedback energy rate,
- ϵ_f is the fraction of the radiated energy that couples to the ISM,
- ϵ_r is the radiative efficiency,
- $\dot{M}_{\text{BH}}$ is the black hole accretion rate.

This energy transfers to the closest 200 of gas particles, increasing the pressure and temperature, which can suppress further star formation. ϵ_f and ϵ_r are configured to match the empirical $\dot{M}_{\text{BH}} - M_\star$ correlation obtained from star-forming regions in galaxies

Local gas properties such as density, temperature, and bulk velocity are calculated using the SPH smoothing kernel (similar to SPH particles), providing the necessary inputs for computing the accretion rate using the Bondi-Hoyle-Lyttleton formula. Gas particles within the smoothing kernel of a black hole are assigned a probability p_j of being accreted:

$$p_j = w_j \frac{\dot{M}_{\text{BH}} \Delta t}{\rho}, \quad (\text{F.9})$$

where:

- w_j is the SPH kernel weight,
- $\dot{M}_{\text{BH}}$ is the accretion rate,
- Δt is the timestep, and
- ρ is the local gas density.

A random number determines whether a particle is accreted.

Apart from accretion, the rate of feedback is also computed from a given black hole. Feedback energy proportional to the accreted mass-energy ($\dot{E}_{\text{feed}} = \epsilon_f \epsilon_r \dot{M}_{\text{BH}} c^2$) is deposited into surrounding gas particles, weighted by the SPH kernel.

Black hole dynamics. Black holes are treated as sink particles that interact gravitationally with their environment. Their mass grows through accretion and mergers with other black holes. The dynamics of BH particles is followed through the simulation using standard N-body techniques, ensuring accurate gravitational interactions.

F.3. Details of the random number generation

The Random Number Generator (RNG) is used to model stochastic processes such as the probability of formation of star or wind particle (detailed in the above [Appendices F and F.2](#)). OPENGADGET3 utilizes the GNU Scientific Library (GSL), specifically the high-quality ranlxd1 algorithm (a variant of the high-quality, long-period pseudo-random number generator Ranlux algorithm ([Marsaglia and Zaman, 1991](#))). To guarantee parallel-safety and bit-wise reproducibility of the random number assignment across parallel executions, the code employs a deterministic lookup system. The global GSL RNG is initialized with a fixed global seed and used once at the start to populate a static lookup table. When a particle requires a random number, the value is retrieved from this table using a deterministic index which is calculated based on the particle's unique ID. This deterministic ID-based indexing ensures that the random number associated with a given particle and process is invariant across different parallel execution environments (MPI ranks, load balancing).

This implementation ensures numerical reproducibility for a fixed execution path; however, in other cases, small differences can cause RNG sequences across the entire volume to become uncorrelated between the two runs. An alternative approach in the SWIFT code

([Schaller et al., 2024](#)), utilizes a combination of particle IDs and discrete timestamps to ensure more localized and time-consistent random sequences. While the relative contribution of the RNG scheme versus the inherent non-linearity of the subgrid physics remains an interesting open question for future benchmarking, our results establish a baseline for the variability present in the current widely-used framework. A possible future direction of investigation will be to include different approaches to RNG within OPENGADGET3, so as to assess any impact that this can have on our simulation results.

Data availability

Data will be made available on request.

References

- Aarseth, S.J., 1963. Dynamical evolution of clusters of galaxies, i. Mon. Not. R. Astron. Soc. 126, 223. doi:10.1093/mnras/126.3.223. Publisher: OUP ADS Bibcode: 1963MNRAS.126..223A. URL <https://ui.adsabs.harvard.edu/abs/1963MNRAS.126..223A>.
- Angulo, R.E., White, S.D.M., 2010. One simulation to fit them all - changing the background parameters of a cosmological N-body simulation. Mon. Not. R. Astron. Soc. doi:10.1111/j.1365-2966.2010.16459.x, arXiv:0912.4277 [astro-ph]. URL <http://arxiv.org/abs/0912.4277>.
- Barai, P., Monaco, P., Murante, G., Ragagnin, A., Viel, M., 2015. Galactic outflow and diffuse gas properties at $z \geq 1$ using different baryonic feedback models. Mon. Not. R. Astron. Soc. 447 (1), 266–286. doi:10.1093/mnras/stu2340.
- Barnes, J.E., Hut, P., 1989. Error Analysis of a Tree Code. Astrophys. J. Suppl. Ser. 70, 389. doi:10.1086/191343, Publisher: IOP ADS Bibcode: 1989ApJS...70..389B. URL <https://ui.adsabs.harvard.edu/abs/1989ApJS...70..389B>.
- Bassini, L., 2021. Buchi neri e formazione stellare negli ammassi di galassie: una vista dalle simulazioni. T13:26:06Z Publisher: Università degli Studi di Trieste. URL <https://arts.units.it/handle/11368/2981312>. (Accepted: 2021 March 04).
- Bassini, L., Rasia, E., Borgani, S., Granato, G.L., Ragone-Figueroa, C., Biffi, V., Ragagnin, A., Dolag, K., Lin, W., Murante, G., Napolitano, N.R., Taffoni, G., Tornatore, L., Wang, Y., 2020. The DIANO simulation of galaxy clusters: characterising star formation in protoclusters. Astron. Astrophys. 642, A37. doi:10.1051/0004-6361/202038396, Publisher: EDP Sciences. URL <https://www.aanda.org/articles/aa/abs/2020/10/aa38396-20/aa38396-20.html>.
- Bassini, L., Rasia, E., Borgani, S., Ragone-Figueroa, C., Biffi, V., Dolag, K., Gaspari, M., Granato, G.L., Murante, G., Taffoni, G., Tornatore, L., 2019. Black hole mass of central galaxies and cluster mass correlation in cosmological hydro-dynamical simulations. Astron. Astrophys. 630, A144. doi:10.1051/0004-6361/201935383, Publisher: EDP Sciences. URL <https://www.aanda.org/articles/aa/abs/2019/10/aa35383-19/aa35383-19.html>.
- Bernardi, M., Meert, A., Sheth, R.K., Vikram, V., Huertas-Company, M., Mei, S., Shankar, F., 2013. The massive end of the luminosity and stellar mass functions: dependence on the fit to the light profile. Mon. Not. R. Astron. Soc. 436 (1), 697–704. doi:10.1093/mnras/stt1607.
- Bertocco, S., Goz, D., Tornatore, L., Ragagnin, A., Maggio, G., Gasparo, F., Vuerli, C., Ta, G., 2020. INAF Trieste Astronomical Observatory Information Technology Framework.
- Biffi, V., Planelles, S., Borgani, S., Fabjan, D., Rasia, E., Murante, G., Tornatore, L., Dolag, K., Granato, G.L., Gaspari, M., Beck, A.M., 2017. The history of chemical enrichment in the intracluster medium from cosmological simulations. Mon. Not. R. Astron. Soc. 468 (1), 531–548. doi:10.1093/mnras/stx444.
- Biffi, V., Planelles, S., Borgani, S., Rasia, E., Murante, G., Fabjan, D., Gaspari, M., 2018. The origin of ICM enrichment in the outskirts of present-day galaxy clusters from cosmological hydrodynamical simulations. Mon. Not. R. Astron. Soc. 476 (2), 2689–2703. doi:10.1093/mnras/sty363.
- Binney, J., Tremaine, S., 2008. Galactic Dynamics: Second Edition REV - Revised, 2nd ed. Princeton University Press, doi:10.2307/j.ctvc778ff, URL <https://www.jstor.org/stable/j.ctvc778ff>.
- Bonafede, A., Dolag, K., Staszyszyn, F., Murante, G., Borgani, S., 2011. A non-ideal magnetohydrodynamic GADGET: simulating massive galaxy clusters. Mon. Not. R. Astron. Soc. 418, 2234–2250. doi:10.1111/j.1365-2966.2011.19523.x, ADS Bibcode: 2011MNRAS.418.2234B. URL <https://ui.adsabs.harvard.edu/abs/2011MNRAS.418.2234B>.
- Bondi, H., 1952. On Spherically Symmetrical Accretion. Mon. Not. R. Astron. Soc. 112 (2), 195–204. doi:10.1093/mnras/112.2.195.
- Borrow, J., Schaller, M., Bahé, Y.M., Schaye, J., Ludlow, A.D., Ploekinger, S., Nobels, F.S.J., Altamura, E., 2023. The impact of stochastic modelling on the predictive power of galaxy formation simulations. Mon. Not. R. Astron. Soc. 526 (2), 2441–2457. doi:10.1093/mnras/stad2928.

- Bose, S., Frenk, C.S., Jenkins, A., Fattahi, A., Gómez, F.A., Grand, R.J.J., Marinacci, F., Navarro, J.F., Oman, K.A., Pakmor, R., Schaye, J., Simpson, C.M., Springel, V., 2019. No cores in dark matter-dominated dwarf galaxies with bursty star formation histories. *Mon. Not. R. Astron. Soc.* 486 (4), 4790–4804. doi:10.1093/mnras/stz1168.
- Chabrier, G., 2003. Galactic Stellar and Substellar Initial Mass Function. *Publ. Astron. Soc. Pac.* 115, 763–795. doi:10.1086/376392, Publisher: IOP ADS Bibcode: 2003PASP..115..763C. URL <https://ui.adsabs.harvard.edu/abs/2003PASP..115..763C>.
- Damiano, A., Valentini, M., Borgani, S., Tornatore, L., Murante, G., Ragagnin, A., Ragone-Figueroa, C., Dolag, K., 2024. Dynamical friction and the evolution of black holes in cosmological simulations: A new implementation in OpenGadget3. *Astron. Astrophys.* 692, A81. doi:10.1051/0004-6361/202450021, Publisher: EDP Sciences. URL <https://www.aanda.org/articles/aa/abs/2024/12/aa50021-24/aa50021-24.html>.
- Davies, J.J., Crain, R.A., Pontzen, A., 2021. Quenching and morphological evolution due to circumgalactic gas expulsion in a simulated galaxy with a controlled assembly history. *Mon. Not. R. Astron. Soc.* 501 (1), 236–253. doi:10.1093/mnras/staa3643.
- Davies, J.J., Pontzen, A., Crain, R.A., 2022. Galaxy mergers can initiate quenching by unlocking an AGN-driven transformation of the baryon cycle. *Mon. Not. R. Astron. Soc.* 515 (1), 1430–1443. doi:10.1093/mnras/stac1742, arXiv:2203.08157 [astro-ph]. URL <http://arxiv.org/abs/2203.08157>.
- Davies, J.J., Pontzen, A., Crain, R.A., 2023. Are the fates of supermassive black holes and galaxies determined by individual mergers, or by the properties of their host haloes? arXiv:2301.04145 [astro-ph]. URL <http://arxiv.org/abs/2301.04145>.
- Davis, M., Efstathiou, G., Frenk, C.S., White, S.D.M., 1985. The evolution of large-scale structure in a universe dominated by cold dark matter. *Astrophys. J.* 292, 371–394. doi:10.1086/163168, Publisher: IOP ADS Bibcode: 1985ApJ...292..371D. URL <https://ui.adsabs.harvard.edu/abs/1985ApJ...292..371D>.
- Dehnen, W., Aly, H., 2012. Improving convergence in smoothed particle hydrodynamics simulations without pairing instability. *Mon. Not. R. Astron. Soc.* 425 (2), 1068–1082. doi:10.1111/j.1365-2966.2012.21439.x, arXiv:1204.2471 [astro-ph]. URL <http://arxiv.org/abs/1204.2471>.
- Dekel, A., Silk, J., 1986. The Origin of Dwarf Galaxies, Cold Dark Matter, and Biased Galaxy Formation. *Astrophys. J.* 303, 39. doi:10.1086/164050, ADS Bibcode: 1986ApJ...303...39D. URL <https://ui.adsabs.harvard.edu/abs/1986ApJ...303...39D>.
- Dolag, K., Borgani, S., Murante, G., Springel, V., 2009. Substructures in hydrodynamical cluster simulations. *Mon. Not. R. Astron. Soc.* 399 (2), 497–514. doi:10.1111/j.1365-2966.2009.15034.x.
- Feigenbaum, M., 1978. Quantitative universality for a class of nonlinear transformations. *J. Stat. Phys.* 19, 25–52. doi:10.1007/BF01020332.
- Gelman, A., 2006. Data Analysis Using Regression And Multilevel/Hierarchical Models. In: Cambridge University Press, vol. 3. doi:10.1017/CBO9780511790942, Journal Abbreviation: Cambridge University Press.
- Genel, S., Bryan, G.L., Springel, V., Hernquist, L., Nelson, D., Pillepich, A., Weinberger, R., Pakmor, R., Marinacci, F., Vogelsberger, M., 2019. A Quantification of the Butterfly Effect in Cosmological Simulations and Implications for Galaxy Scaling Relations. *Astrophys. J.* 871 (1), 21. doi:10.3847/1538-4357/aaf4bb, Publisher: The American Astronomical Society.
- Granato, G.L., Ragone-Figueroa, C., Taverna, A., Silva, L., Valentini, M., Borgani, S., Monaco, P., Murante, G., Tornatore, L., 2021. Dust evolution in zoom-in cosmological simulations of galaxy formation. *Mon. Not. R. Astron. Soc.* 503 (1), 511–532. doi:10.1093/mnras/stab362.
- Groth, F., Steinwandel, U.P., Valentini, M., Dolag, K., 2023. The cosmological simulation code OPENGADGET3 - implementation of meshless finite mass. *Mon. Not. R. Astron. Soc.* 526, 616–644. doi:10.1093/mnras/stad2717, Publisher: OUP ADS Bibcode: 2023MNRAS.526..616G. URL <https://ui.adsabs.harvard.edu/abs/2023MNRAS.526..616G>.
- Gurzadian, V.G., Savvidi, G.K., 1984. On the problem of relaxation of stellar systems.. *Akad. Nauk. SSSR Dokl.* 277, 69–73, ADS Bibcode: 1984DokSSR.277...69G. URL <https://ui.adsabs.harvard.edu/abs/1984DokSSR.277...69G>.
- Gurzadian, V.G., Savvidi, G.K., 1986. Collective relaxation of stellar systems.. *Astron. Astrophys.* 160, 203–210, ADS Bibcode: 1986A&A...160..203G. URL <https://ui.adsabs.harvard.edu/abs/1986A&A...160..203G>.
- Heggie, D.C., 1991. Chaos in the N-body problem of stellar dynamics. 272, pp. 47–62, Conference Name: Predictability, Stability, and Chaos in N-Body Dynamical Systems ADS Bibcode: 1991ASIB..272...47H. URL <https://ui.adsabs.harvard.edu/abs/1991ASIB..272...47H>.
- Henon, M., Heiles, C., 1964. The applicability of the third integral of motion: Some numerical experiments. *Astron. J.* 69, 73. doi:10.1086/109234, URL http://adsabs.harvard.edu/cgi-bin/bib_query?1964AJ.....69...73H.
- Hernquist, L., Hut, P., Makino, J., 1993. Discreteness Noise versus Force Errors in N-Body Simulations. *Astrophys. J.* 402, L85. doi:10.1086/186706, Publisher: IOP ADS Bibcode: 1993ApJ...402L..85H. URL <https://ui.adsabs.harvard.edu/abs/1993ApJ...402L..85H>.
- Hoffman, Y., Ribak, E., 1991. Constrained Realizations of Gaussian Fields: A Simple Algorithm. *Astrophys. J.* 380, L5. doi:10.1086/186160, Publisher: IOP ADS Bibcode: 1991ApJ...380L...5H. URL <https://ui.adsabs.harvard.edu/abs/1991ApJ...380L...5H>.
- Kandrup, H.E., 1989. The time scale for “mixing” in a stellar dynamical system. *Phys. Lett. A* 140, 97–100. doi:10.1016/0375-9601(89)90497-0, ADS Bibcode: 1989PhLA..140...97K. URL <https://ui.adsabs.harvard.edu/abs/1989PhLA..140...97K>.
- Kandrup, H.E., 1990a. Divergence of Nearby Trajectories for the Gravitational N-Body Problem. *Astrophys. J.* 364, 420. doi:10.1086/169425, Publisher: IOP ADS Bibcode: 1990ApJ...364..420K. URL <https://ui.adsabs.harvard.edu/abs/1990ApJ...364..420K>.
- Kandrup, H.E., 1990b. How fast can a galaxy “mix”? *Phys. A Stat. Mech. Appl.* 169, 73–94. doi:10.1016/0378-4371(90)90217-G, ADS Bibcode: 1990PhyA..169...73K. URL <https://ui.adsabs.harvard.edu/abs/1990PhyA..169...73K>.
- Karakas, A.I., 2010. VizieR Online Data Catalog: Updated stellar yields from AGB models (Karakas, 2010). *VizieR Online Data Cat.* 740, J/MNRAS/403/1413. ADS Bibcode: 2010yCat..74031413K. URL <https://ui.adsabs.harvard.edu/abs/2010yCat..74031413K>.
- Keller, B.W., Kruijssen, J.M.D., 2022. Uncertainties in supernova input rates drive qualitative differences in simulations of galaxy evolution. *Mon. Not. R. Astron. Soc.* 512 (1), 199–215. doi:10.1093/mnras/stac511, arXiv:2004.03608 [astro-ph]. URL <http://arxiv.org/abs/2004.03608>.
- Keller, B.W., Wadsley, J.W., Wang, L., Kruijssen, J.M.D., 2019. Chaos and variance in galaxy formation. *Mon. Not. R. Astron. Soc.* 482 (2), 2244–2261. doi:10.1093/mnras/sty2859.
- Larson, R.B., 1974. Effects of Supernovae on the Early Evolution of Galaxies. *Mon. Not. R. Astron. Soc.* 169 (2), 229–245. doi:10.1093/mnras/169.2.229.
- Lorenz, E.N., 1963. Deterministic Nonperiodic Flow. (ISSN: 1520-0469) Section: Journal of the Atmospheric Sciences. URL https://journals.ametsoc.org/view/journals/atsc/20/2/1520-0469_1963_020_0130_dnf_2_0_co_2.xml.
- Lovell, M.R., Pillepich, A., Genel, S., Nelson, D., Springel, V., Pakmor, R., Marinacci, F., Weinberger, R., Torrey, P., Vogelsberger, M., Alabi, A., Hernquist, L., 2018. The fraction of dark matter within galaxies from the IllustrisTNG simulations. *Mon. Not. R. Astron. Soc.* 481 (2), 1950–1975. doi:10.1093/mnras/sty2339.
- Mandelbrot, B.B., 1982. The fractal geometry of nature, Revised ed. 7876824, San Francisco, OCLC. URL <http://catdir.loc.gov/catdir/bios/hol056/81015085.html>.
- Marini, I., Borgani, S., Saro, A., Granato, G.L., Ragone-Figueroa, C., Sartoris, B., Dolag, K., Murante, G., Ragagnin, A., Wang, Y., 2021. Velocity dispersion of brightest cluster galaxies in cosmological simulations. *Mon. Not. R. Astron. Soc.* 507 (4), 5780–5795. doi:10.1093/mnras/stab2518.
- Marsaglia, G., Zaman, A., 1991. A New Class of Random Number Generators. *Ann. Appl. Probab.* 1 (3), 462–480. doi:10.1214/aop/1177005878, Publisher: Institute of Mathematical Statistics. URL <https://projecteuclid.org/journals/annals-of-applied-probability/volume-1/issue-3/A-New-Class-of-Random-Number-Generators/10.1214/aop/1177005878.full>.
- Matteucci, F., Padovani, P., 1993. Chemical Evolution of Galaxies and Quasar Metallicities. *Astrophys. J.* 419, 485. doi:10.1086/173502, Publisher: IOP ADS Bibcode: 1993ApJ...419..485M. URL <https://ui.adsabs.harvard.edu/abs/1993ApJ...419..485M>.
- May, R., 1976. Simple Mathematical Models With Very Complicated Dynamics. *Nature* 26, 457. doi:10.1038/261459a0.
- Miller, R.H., 1964. Irreversibility in Small Stellar Dynamical Systems.. *Astrophys. J.* 140, 250. doi:10.1086/147911, Publisher: IOP ADS Bibcode: 1964ApJ...140..250M. URL <https://ui.adsabs.harvard.edu/abs/1964ApJ...140..250M>.
- Murante, G., Monaco, P., Borgani, S., Tornatore, L., Dolag, K., Goz, D., 2014. Simulating realistic disk galaxies with a novel sub-resolution ISM model. doi:10.48550/arXiv.1411.3671, arXiv:1411.3671 [astro-ph]. URL <http://arxiv.org/abs/1411.3671>.
- Murante, G., Monaco, P., Giovalli, M., Borgani, S., Diaferio, A., 2010. A subresolution multiphase interstellar medium model of star formation and supernova energy feedback. *Mon. Not. R. Astron. Soc.* 405 (3), 1491–1512. doi:10.1111/j.1365-2966.2010.16567.x, Publisher: Oxford Academic.
- Oh, B.K., Smith, B.D., Peacock, J.A., Khochfar, S., 2020. Calibration of a star formation and feedback model for cosmological simulations with enzo. *Mon. Not. R. Astron. Soc.* 497 (4), 5203–5219. doi:10.1093/mnras/staa2318.
- Pakmor, R., Bieri, R., Fragkoudi, F., Gómez, F.A., Grand, R.J.J., Simpson, C.M., Talbot, R.Y., van de Voort, F., Werhahn, M., 2025. Quantifying the intrinsic variability due to randomness of the Auriga galaxy formation model. *Mon. Not. R. Astron. Soc.* 543 (2), 1761–1774. doi:10.1093/mnras/staf1542.
- Pillepich, A., Springel, V., Nelson, D., Genel, S., Naiman, J., Pakmor, R., Hernquist, L., Torrey, P., Vogelsberger, M., Weinberger, R., Marinacci, F., 2018. Simulating galaxy formation with the IllustrisTNG model. *Mon. Not. R. Astron. Soc.* 473 (3), 4077–4106. doi:10.1093/mnras/stx2656.
- Planelles, S., Fabjan, D., Borgani, S., Murante, G., Rasia, E., Biffi, V., Truong, N., Ragone-Figueroa, C., Granato, G.L., Dolag, K., Pierpaoli, E., Beck, A.M., Steinborn, L.K., Gaspari, M., 2017. Pressure of the hot gas in simulations of galaxy clusters. *Mon. Not. R. Astron. Soc.* 467 (4), 3827–3847. doi:10.1093/mnras/stx318.
- Ragagnin, A., 2025. Aragagnin/g3read. original-date: 2020-05-08T22:11:19Z. URL <https://github.com/aragagnin/g3read>.
- Ragone-Figueroa, C., Granato, G.L., Ferraro, M.E., Murante, G., Biffi, V., Borgani, S., Planelles, S., Rasia, E., 2018. BCG mass evolution in cosmological hydro-simulations. *Mon. Not. R. Astron. Soc.* 479 (1), 1125–1136. doi:10.1093/mnras/sty1639.

- Ragone-Figueroa, C., Granato, G.L., Murante, G., Borgani, S., Cui, W., 2013. Brightest cluster galaxies in cosmological simulations: achievements and limitations of active galactic nuclei feedback models. *Mon. Not. R. Astron. Soc.* 436 (2), 1750–1764. doi:10.1093/mnras/stt1693.
- Rasia, E., Borgani, S., Murante, G., Planelles, S., Beck, A.M., Biffi, V., Ragone-Figueroa, C., Granato, G.L., Steinborn, L.K., Dolag, K., 2015. COOL CORE CLUSTERS FROM COSMOLOGICAL SIMULATIONS. *Astrophys. J.* 813 (1), L17. doi:10.1088/2041-8205/813/1/L17, Publisher: American Astronomical Society. URL <https://doi.org/10.1088/2041-8205/813/1/L17>.
- Rey, M.P., Pontzen, A., 2018. Quadratic genetic modifications: a streamlined route to cosmological simulations with controlled merger history. *Mon. Not. R. Astron. Soc.* 474 (1), 45–54. doi:10.1093/mnras/stx2744, URL <http://academic.oup.com/mnras/article/474/1/45/4563625>.
- Romano, D., Karakas, A.I., Tosi, M., Matteucci, F., 2010. Quantifying the uncertainties of chemical evolution studies - II. Stellar yields. *Astron. Astrophys.* 522, A32. doi:10.1051/0004-6361/201014483, Publisher: EDP Sciences. URL <https://www.aanda.org/articles/aa/abs/2010/14/aa14483-10/aa14483-10.html>.
- Roth, N., Pontzen, A., Peiris, H.V., 2016. Genetically modified haloes: towards controlled experiments in Λ CDM galaxy formation. *Mon. Not. R. Astron. Soc.* 455 (1), 974–986. doi:10.1093/mnras/stv2375, URL <http://academic.oup.com/mnras/article/455/1/974/986737/Genetically-modified-haloes-towards-controlled>.
- Schaller, M., Borrow, J., Draper, P.W., Ivkovic, M., McAlpine, S., Vandenbroucke, B., Bahé, Y., Chaikin, E., Chalk, A.B.G., Chan, T.K., Correa, C., van Daalen, M., Elbers, W., Gonnet, P., Hausammann, L., Helly, J., Huško, F., Kegerreis, J.A., Nobels, F.S.J., Ploekinger, S., Revaz, Y., Roper, W.J., Ruiz-Bonilla, S., Sandnes, T.D., Uyttenhove, Y., Willis, J.S., Xiang, Z., 2024. SWIFT: A modern highly-parallel gravity and smoothed particle hydrodynamics solver for astrophysical and cosmological applications. *Mon. Not. R. Astron. Soc.* 530 (2), 2378–2419. doi:10.1093/mnras/stae922, arXiv:2305.13380 [astro-ph]. URL <http://arxiv.org/abs/2305.13380>.
- Schaller, M., Frenk, C.S., Bower, R.G., Theuns, T., Jenkins, A., Schaye, J., Crain, R.A., Furlong, M., Dalla Vecchia, C., McCarthy, I.G., 2015. Baryon effects on the internal structure of Λ CDM haloes in the EAGLE simulations. *Mon. Not. R. Astron. Soc.* 451, 1247–1267. doi:10.1093/mnras/stv1067, Publisher: OUP ADS Bibcode: 2015MNRAS.451.1247S. URL <https://ui.adsabs.harvard.edu/abs/2015MNRAS.451.1247S>.
- Schaye, J., Crain, R.A., Bower, R.G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C.S., McCarthy, I.G., Helly, J.C., Jenkins, A., Rosas-Guevara, Y.M., White, S.D.M., Baes, M., Booth, C.M., Camps, P., Navarro, J.F., Qu, Y., Rahmati, A., Sawala, T., Thomas, P.A., Trayford, J., 2015. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *Mon. Not. R. Astron. Soc.* 446 (1), 521–554. doi:10.1093/mnras/stu2058.
- Schmidt, M., 1959. The Rate of Star Formation. *Astrophys. J.* 129, 243. doi:10.1086/146614, Publisher: IOP ADS Bibcode: 1959ApJ...129..243S. URL <https://ui.adsabs.harvard.edu/abs/1959ApJ...129..243S>.
- Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and Statistical Modeling with Python. In: 9th Python in Science Conference. doi:10.25080/Majora-92bf1922-011, URL <https://proceedings.scipy.org/articles/Majora-92bf1922-011>.
- Springel, V., 2010. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *Mon. Not. R. Astron. Soc.* 401 (2), 791–851. doi:10.1111/j.1365-2966.2009.15715.x.
- Springel, V., Di Matteo, T., Hernquist, L., 2005. Modelling feedback from stars and black holes in galaxy mergers. *Mon. Not. R. Astron. Soc.* 361, 776–794. doi:10.1111/j.1365-2966.2005.09238.x, ADS Bibcode: 2005MNRAS.361..776S. URL <https://ui.adsabs.harvard.edu/abs/2005MNRAS.361..776S>.
- Springel, V., Hernquist, L., 2003. Cosmological SPH simulations: A hybrid multi-phase model for star formation. *Mon. Not. R. Astron. Soc.* 339 (2), 289–311. doi:10.1046/j.1365-8711.2003.06206.x, arXiv:astro-ph/0206393. URL <http://arxiv.org/abs/astro-ph/0206393>.
- Springel, V., Yoshida, N., White, S.D.M., 2001. GADGET: a code for collisionless and gasdynamical cosmological simulations. *New Astron.* 6 (2), 79–117. doi:10.1016/S1384-1076(01)00042-2, URL <https://www.sciencedirect.com/science/article/pii/S1384107601000422>.
- Standish, Jr., E.M., 1968. Numerical Studies of the Gravitational Problem of N Bodies. (Ph.D. thesis). Yale University, Publication Title: Ph.D. Thesis ADS Bibcode: 1968PhDT.....6S. URL <https://ui.adsabs.harvard.edu/abs/1968PhDT.....6S>.
- Strogatz, S.H., 2019. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, second ed. CRC Press, Boca Raton, doi:10.1201/9780429492563.
- Su, K.-Y., Hopkins, P.F., Hayward, C.C., Faucher-Giguère, C.-A., Kereš, D., Ma, X., Robles, V.H., 2017. Feedback first: the surprisingly weak effects of magnetic fields, viscosity, conduction and metal diffusion on sub- l^* galaxy formation. *Mon. Not. R. Astron. Soc.* 471 (1), 144–166. doi:10.1093/mnras/stx1463.
- Su, K.-Y., Hopkins, P.F., Hayward, C.C., Ma, X., Boylan-Kolchin, M., Kasen, D., Kereš, D., Faucher-Giguère, C.-A., Orr, M.E., Wheeler, C., 2018. Discrete effects in stellar feedback: Individual Supernovae, Hypernovae, and IMF Sampling in Dwarf Galaxies. *Mon. Not. R. Astron. Soc.* 480 (2), 1666–1675. doi:10.1093/mnras/sty1928.
- Taffoni, G., Becciani, U., Garilli, B., Maggio, G., Pasian, F., Umana, G., Smareglia, R., Vitello, F., 2020. CHIPP: INAF pilot project for HTC, HPC and HPDA. doi:10.48550/arXiv.2002.01283, arXiv:2002.01283 [astro-ph]. URL <http://arxiv.org/abs/2002.01283>.
- Thiébaud, J., Pichon, C., Sousbie, T., Prunet, S., Pogosyan, D., 2008. On the onset of stochasticity in Λ cold dark matter cosmological simulations. *Mon. Not. R. Astron. Soc.* 387 (1), 397–406. doi:10.1111/j.1365-2966.2008.13250.x.
- Thielemann, F.-K., Argast, D., Brachwitz, F., Hix, W.R., Höflich, P., Liebendörfer, M., Martínez-Pinedo, G., Mezzacappa, A., Nomoto, K., Panov, I., 2003. Supernova Nucleosynthesis and Galactic Evolution. In: Hillebrandt, W., Leibundgut, B. (Eds.), From Twilight To Highlight: The Physics of Supernovae. Springer, Berlin, Heidelberg, pp. 331–343. doi:10.1007/10828549_46.
- Tormen, G., Bouchet, F.R., White, S.D.M., 1997. The structure and dynamical evolution of dark matter haloes. *Mon. Not. R. Astron. Soc.* 286 (4), 865–884. doi:10.1093/mnras/286.4.865.
- Tornatore, L., Borgani, S., Dolag, K., Matteucci, F., 2007. Chemical enrichment of galaxy clusters from hydrodynamical simulations: Chemical enrichment of galaxy clusters. *Mon. Not. R. Astron. Soc.* 382 (3), 1050–1072. doi:10.1111/j.1365-2966.2007.12070.x, URL <https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2007.12070.x>.
- Tornatore, L., Borgani, S., Springel, V., Matteucci, F., Menci, N., Murante, G., 2003. Cooling and heating the intracluster medium in hydrodynamical simulations. *Mon. Not. R. Astron. Soc.* 342 (4), 1025–1040. doi:10.1046/j.1365-8711.2003.06631.x, URL <https://academic.oup.com/mnras/article/342/4/1025/956309>.
- Valentini, M., Dolag, K., Borgani, S., Murante, G., Maio, U., Tornatore, L., Granato, G.L., Ragone-Figueroa, C., Burkert, A., Ragagnin, A., Rasia, E., 2023. Impact of H2-driven star formation and stellar feedback from low-enrichment environments on the formation of spiral galaxies. *Mon. Not. R. Astron. Soc.* 518, 1128–1147. doi:10.1093/mnras/stac2110, Publisher: OUP ADS Bibcode: 2023MNRAS.518.1128V. URL <https://ui.adsabs.harvard.edu/abs/2023MNRAS.518.1128V>.
- Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G.F., Nelson, D., Hernquist, L., 2014. Introducing the Illustris Project: Simulating the coevolution of dark and visible matter in the Universe. *Mon. Not. R. Astron. Soc.* 444 (2), 1518–1547. doi:10.1093/mnras/stu1536, arXiv:1405.2921 [astro-ph]. URL <http://arxiv.org/abs/1405.2921>.
- White, S.D.M., Frenk, C.S., 1991. Galaxy Formation through Hierarchical Clustering. *Astrophys. J.* 379, 52. doi:10.1086/170483, ADS Bibcode: 1991ApJ...379...52W. URL <https://ui.adsabs.harvard.edu/abs/1991ApJ...379...52W>.
- Woosley, S.E., Weaver, T.A., 1995. The Evolution and Explosion of Massive Stars. II. Explosive Hydrodynamics and Nucleosynthesis. *Astrophys. J. Suppl. Ser.* 101, 181. doi:10.1086/192237, Publisher: IOP ADS Bibcode: 1995ApJS...101..181W. URL <https://ui.adsabs.harvard.edu/abs/1995ApJS...101..181W>.