UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVI CICLO DEL DOTTORATO DI RICERCA IN FISICA

Chaotic effects in cosmological hydrodynamical simulations and archive

Settore scientifico-disciplinare: **FIS/05**

DOTTORANDO / A
Chaitra

COORDINATORE
PROF. Francesco Longo

SUPERVISORE DI TESI
PROF. Giuliano Taffoni

PROF. Stefano Borgani

ANNO ACCADEMICO 2023/2024

Contents

Abstract

This PhD thesis addresses two distinct challenges in the field of astrophysical simulations: quantifying variability in simulations and developing standards-compliant data-sharing frameworks for simulation outputs. In the first part, I investigate the chaotic effects inherent in galaxy cluster simulations. To understand the range of variability that arises from minor differences in initial conditions and numerical precision, I ran a set of identical simulations on the HotCat computing cluster using the OpenGadget3 code. I focused on quantifying the variation in galaxy properties by matching galaxies across runs and analyzing the variations using two statistical methods. My findings for low resolution galaxy cluster simulation indicate that noise, primarily due to Poisson sampling, is the dominant source of variability in galaxy properties, with only minimal run-to-run differences observed in baseline and feedback tests. Stronger stellar feedback result in an increase in stochastic noise within given simulation and this correlation depends on galaxies within specific mass bins that are more sensitive to the respective feedback processes. These insights set the first foundational basis for understanding the limitations of numerical simulations and set the stage for studying the effects of chaos for accurately interpreting results from cosmological simulations.

The second part of this thesis focuses on enhancing the accessibility and interoperability of simulation data by developing a standards-compliant archive for astrophysical simulations. Recognizing the need for efficient data sharing, the primary objectives for the part 2 of my thesis were established. For this work, I implemented a custom database and interface as a first step, followed by the integration of key Virtual Observatory (VO) data standards, such as VOTable for tabular data representation, Universal Worker Service (UWS) for managing long-running processes, and preliminary templates based on the Simulation Data Model (SimDM). However, there were several challenges encountered with VO standards, particularly regarding the SimDM and SimDAL (Simulation Data Access Layer) standards, which lack explicit guidelines for managing large, distributed datasets typical of cosmological simulations. In collaboration with VO standards authors, a systematic review will be initiated to address these limitations, with a specific focus on adapting SimDM templates for complex simulation datasets. This part of my work contributes to the broader goals of the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles by refining the standards and practices for astrophysical simulation data sharing, ultimately advancing the capabilities of researchers to analyze, interpret, and build on existing simulation results.

Chapter 1

Context of this thesis

The research activities conducted during this PhD focus on two main topics in the field of numerical astrophysics and cosmology. The first topic centers on the study of chaotic effects in cosmological hydrodynamical simulations, aiming to understand the sensitivity and stochasticity in formation of galaxies. The second topic is a data science initiative, aimed at the development of an archive for cosmological hydrodynamical simulations. This archive is intended to support the broader astrophysics community by providing standardized, accessible datasets for analysis and further simulation work. In this chapter a broad introduction of the two topics and the thesis structure is outlined.

1.1 Chaotic effects in cosmological hydrodynamical simulations

This project aims to study the effects of chaos in Numerical Cosmological Simulations. Chaotic effects, also termed as the butterfly effect, is an inherent feature of many complex systems where minute differences are amplified over time, leading to large divergences in outcomes. In astrophysical and cosmological simulations, sources of chaos include parallel reductions, variations in the random number generator, slight perturbations in initial conditions, and, when coupled with sub-resolution models that are highly non-linear, these small changes can grow and lead to significantly different outcomes.
In the astrophysical context, particularly in the study of clusters of galaxies and brightest cluster galaxies (BCGs), simulations often incorporate a large number of astrophysical modules to model various physical processes. Galaxy clusters are massive structures in the universe composed of hundreds to thousands of galaxies bound together by gravity, with a significant portion of their mass in the form of dark matter. These clusters are crucial for understanding large-scale structure formation and the dynamics of cosmic evolution. Within these clusters, BCGs are of particular interest as they are typically the largest and most massive galaxies, often located near the cluster's center. BCGs play a key role in astrophysics and cosmology by acting as tracers for understanding the assembly history and growth of galaxy clusters, as well as their interactions with dark matter and

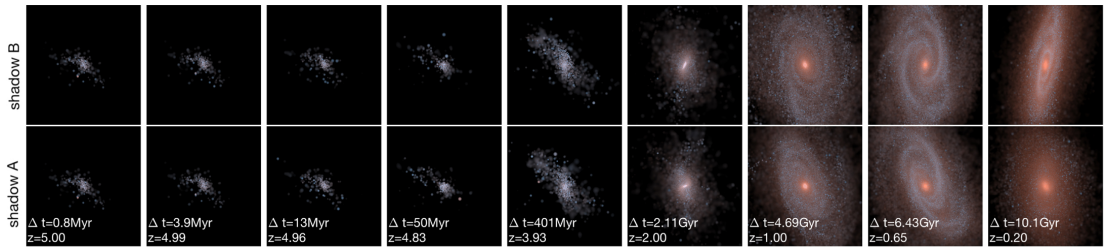

Figure 1.1.1: Snapshots of galaxies, shadow A and shadow B matched between identical simulations, but with minute perturbation in initial conditions. The evolution from z=5 to z=0.2 is shown. Credits: [48].

intracluster gas. Their study helps to probe the influence of galaxy mergers, feedback mechanisms, and large-scale structure formation in the universe.

Simulations of galaxy cluster and galaxies include gas dynamics, stellar evolution, feedback mechanisms, black hole interactions, chemical enrichment, all of which contribute to the complexity of the system. Thus, astrophysical simulations are intrinsically chaotic, and even tiny variations can produce divergent results. For example in Figure 1.1.1, the galaxies evolve in simulations of identical setup except for the introduction of minute perturbation in their initial condition, end-up having significant differences in their structure and properties [48]. The final result of one galaxy still actively star forming (shadow B) and the other (shadow A), mostly quenched.

In this work, the aim is to quantify the impact of chaotic variations on key observables, such as galaxy dark matter mass, stellar mass, star formation rates, etc., specifically to lay out the baselines using OpenGadget3 code and also, to explore how, or if, certain astrophysical processes mitigate the effects of chaos. By understanding the relationship between chaos and physical processes, an assessment of the reliability of simulation results can be made and ensure that they accurately reflect underlying astrophysical phenomena rather than being a consequence of numerical artifacts.

Furthermore, understanding the effects of chaos on high performance computing (HPC) platforms is vital for theoretical-observational comparisons, especially in the exascale era, where these effects can magnify due to supercomputer complexity.

This study represents a crucial foundational step in understanding the effects of chaos across different configuration parameters, initial conditions (ICs), platforms, and sub-grid physics, laying the groundwork for more detailed investigations to follow.

The efforts to understand chaotic effects involves an extensive set of tests across multiple configurations, examining not just a single setup but a range of tests with different configurations and parameters. Tests include varying number of MPI tasks, checkpointing, varying random number generator seed, runs on different computing infrastructure, varying resolutions etc,. Early tests were made on galaxy simulations, which proved challenging due to long runtimes needed to achieve a sufficient sample size (refers to number data points collected in a study) to estimate the variation on matched galaxies. Resolution-based tests were also

attempted, but instability in the code led to switching to an older version, ultimately running cluster simulations to produce the first results presented in this thesis.
One of the main problems is making a comprehensive analysis on chaos is the use of computing resource, altogether, this project has so far required approximately 4 million core hours.

1.2 Development of a cosmological hydrodynamical simulations archive

Cosmological simulations have been a crucial implement in exploration of the formation and evolution of the universe's large-scale structures, from galaxies to galaxy clusters. By testing theoretical models against observed phenomena, these simulations provide insights into fundamental questions about the nature of dark matter, dark energy, and the evolution of cosmic structures, making them indispensable for both theoretical and observational cosmology. Currently, cosmological simulations (i.e. the resulting raw and post-processed data products) are mainly hosted at HPC (High performance computing) centers, where access to the simulation data is not straightforward, considering restricted access, diverse data formats, a required technical know-how etc.
To address these issues, a consolidated infrastructure adhering to Virtual Observatory (VO) standards and FAIR (Findable, Accessible, Interoperable, and Reusable) data principles is proposed. This system integrates the IVOA's (International Virtual Observatory Alliance) simulation standards for detailed metadata descriptions, enabling comprehensive data discovery, access, and retrieval. By adhering to these standards, this work provides a foundation for regularizing the access to simulations. It also aims to identify existing limitations and refine standard implementation to maximize the discoverability and reusability of published simulations, ensuring they are widely accessible and usable by the scientific community. Collaborative efforts with the IVOA play a crucial role in driving these improvements for the future of this project.

1.3 An outline of the thesis

Chapter 1, considering that the PhD is comprised of 2 distinct topics, sets the context with a broad introduction to the topics. The first topic, i.e study of chaotic effects in cosmological hydrodynamical simulations forms the first part of the thesis structure. Thus, for **Part 1**, I begin with introductory chapters, which are **Chapter 2** and **Chapter 3**. **Chapter 2** provides a background on cosmological simulations, their role and key scientific findings in the literature. Essential components required for the set up and execution of cosmological simulations, including initial conditions, the code, and the subgrid physics models used as well as the astrophysical objects studied within the context of this project, i.e. galaxy clusters, are detailed. **Chapter 3** provides a historical overview of the study of chaotic effects, particularly in the realm of astrophysics, and highlights

key findings, placing an emphasis on recent advancements and their implications. Methodology and main results of **Part 1** are detailed in **Chapter 4**, which presents an in-depth discussion of the simulation setup, including the types of tests conducted and the methods of analysis used to derive the results. The primary results of this work are showcased, covering a range of observables such as the galaxy stellar mass function, the star formation history on large scales, and the variations in galaxy properties like dark matter mass, stellar mass, taken for a sample of galaxies.
In **Part 2**, I cover the work related to the development of cosmological hydrodynamical simulations archive all within a single chapter: **Chapter 5**. This chapter begins with introductory sections. Firstly, outlining the key requirements related to simulation data storage, curation, and access within the broader context of scientific research. Next, the introduction delves into the relevant standards that underpin this work, offering an exploration of protocols and frameworks essential for ensuring interoperability and long-term accessibility of the data. In the subsequent sections specifics of the implementation details of the archive are provided. Starting with a section on the design of custom implementation of a database and interface, showcasing the elements implemented thus far, accompanied by visual representations of the various features. In the last part of **Part 2**, I cover the particulars of standard implementations (Simulation Data Model and Simulation Data Access Layer), including the details of the evolving requirements that emerged.
In the last chapter, **Chapter 6**, I summarize and conclude my work within the context of both projects.

Chapter 2

Numerical Simulation

This chapter explores the development and application of numerical simulations in astrophysics and cosmology, from their theoretical foundations to modern methods discussing the essentials of N-body and hydrodynamical simulations. A particular focus is given to GADGET-3, which is relevant for current work. Key aspects like initial conditions and sub-grid physics (e.g., cooling, star formation, and black hole feedback) are highlighted, along with clustering algorithms used for galaxy identification. The chapter culminates with an overview on galaxy clusters which is the primary object simulated in my work.

2.1 Historical overview

The origins of cosmological simulations trace back to the pioneering work of Jim Peebles and Yakov Zeldovich in the 1960s and 1970s (see [105] and [157]). Peebles conducted pioneering numerical experiments on gravitational collapse in an expanding universe, utilizing simple models with a limited number of particles to explore the formation of cosmic structures. In 1970, Zeldovich introduced the Zeldovich approximation, an analytical model that approximates matter behavior under gravitational collapse, laying a critical foundation for more complex simulations.

The 1980s marked significant advancements in computational power and algorithm development, enabling more detailed and larger-scale simulations. This period saw the introduction of particle-mesh (PM) and particle-particle particle-mesh (P^3M) methods, allowing for more accurate modeling of large-scale structures. In 1985, researchers at the Center for Astrophysics (CfA), including Davis et al., conducted one of the first high-resolution N-body simulations, modeling large-scale structures with thousands of particles and visualizing filamentary structures resembling the cosmic web [31]. The emergence of the Cold Dark Matter (CDM) paradigm in this decade provided a robust framework for understanding cosmic structure formation, with simulations incorporating CDM accurately reproducing the observed distribution of galaxies.

By the 1990s and into the 2000s, advancements in hardware and software facilitated the transition to simulations involving millions of particles. The "Hubble Volume Simulation", conducted around 2002 by the Virgo Consortium, simulated

a cubic volume of the universe spanning two billion light-years on each side, providing detailed predictions for the distribution of dark matter [41]. The "Millennium Simulation" in 2005, also by the Virgo Consortium, used over 10 billion particles to model cosmic structures in a CDM universe, becoming a benchmark for comparing theoretical predictions with observational data (see [138], [133]). The 2010s marked a significant advancements in hydrodynamical models that comprehensively integrated the physics of gas dynamics, star formation, and feedback processes alongside gravitational interactions. The "Illustris Project" in 2014 was a pioneering effort in this domain, offering detailed simulations of galaxy formation and evolution within a Cold Dark Matter framework. Illustris simulated both dark matter and baryonic matter, producing realistic galaxy populations and providing insights into how supernovae and active galactic nuclei impact galaxy development. Building on this foundation, the "IllustrisTNG" (The Next Generation, see [106]) simulation launched in 2018, improved physical models and achieved higher resolution, yielding more precise predictions for the properties of galaxies, galaxy clusters, and the intergalactic medium.
Looking ahead, the 2020s promise further advancements driven by the advent of exascale computing and increasingly sophisticated physical models. Exascale computing, capable of performing at least 10^{18} floating point operations per second, will enable simulations with unprecedented detail and accuracy (e.g., ExaNeSt H2020; see [74], [79]). Future simulations are expected to offer critical insights into fundamental cosmological questions, including the nature of dark matter and dark energy, as well as the formation of the first stars and galaxies. As computational power continues to grow, these cutting-edge simulations will play an essential role in deepening our understanding of the universe and interpreting data from next-generation observational instruments.

2.2 Theoretical Framework

Cosmological simulations rely on a combination of theoretical frameworks, mathematical and computational techniques to model the complex interplay of forces and particles over cosmic time scales. This subsection provides an outline of the theoretical framework.
The foundation of cosmological simulations is rooted in the ΛCDM (Lambda Cold Dark Matter) model, which is the prevailing cosmological paradigm describing the large-scale structure and evolution of the universe. It assumes a flat, homogeneous, and isotropic universe, consistent with observations of the cosmic microwave background (CMB) and large-scale structure surveys. The initial conditions for cosmological simulations are thus, typically set by the CMB observations (e.g., [27]), which provide a snapshot of the universe approximately 380,000 years after the Big Bang. These initial conditions are often represented by a Gaussian random field based on the primordial density fluctuations. In the ΛCDM model, dark matter and dark energy are critical components. Dark matter is the dominant form of matter in the universe, interacts through gravity and drives of formation structures like galaxies and clusters, while dark energy causes the accelerated expansion of the universe.

Friedmann-Lemaître-Robertson-Walker (FLRW) Metric provides a solution to Einstein's field equations of General Relativity (foundational theory for gravity, [40]) under the ΛCDM assumption of a homogeneous and isotropic universe [44]. It describes how the universe expands over time, characterized by the scale factor $a(t)$, which is determined by the Friedmann equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \tag{2.1}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \tag{2.2}$$

Here, ρ is the density, p is the pressure, G is the gravitational constant, k is the curvature parameter, Λ is the cosmological constant and c is the light speed. Basically this equation relates the universe's expansion to its energy content (matter density, dark energy) and its curvature.
The FLRW framework governs the large scale spatial evolution of the simulation, representative of the background cosmological model of the universe. On small scales, the universe is not homogeneous and Newtons laws of motion are valid. The content is modelled as a fluid made of collisionless particles which interacts only through gravity. The Vlasov-Boltzmann equation governs the phase space distribution function $f(\mathbf{x}, \mathbf{v}, t)$ of the collisionless particles, describing how this function evolves over time due to gravitational interactions. It can be written as:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_x f - \nabla\Phi \cdot \nabla_v f = 0 \tag{2.3}$$

where Φ is the gravitational potential. Gravity's influence on these particles is calculated using the Poisson equation, which relates the gravitational potential ϕ to the matter density ρ:

$$\nabla^2\phi = 4\pi G\rho \tag{2.4}$$

By solving the Poisson equation, one can determine the gravitational potential from the matter density distribution. This potential then influences the trajectories of particles, completing the cycle of gravitational interaction in collisionless matter dynamics on smaller scales.

Evolving the system: The described system needs to be evolved with time and the properties associated with the N-particles need to be computed. Euler's method is a straightforward numerical technique based on the idea of approximating the solution by taking small, discrete steps along the curve defined by the differential equation. Specifically, if we have a first-order ODE of the form $\frac{dy}{dt} = f(t, y)$ with an initial condition $y(t_0) = y_0$, where t is time. Euler's method approximates the solution at a point t_{n+1} by using the formula:

$$y_{n+1} = y_n + hf(t_n, y_n) \tag{2.5}$$

Here, h is the step size for time, which determines the interval between points t_n and t_{n+1}. The method essentially uses the slope $f(t_n, y_n)$ at the current point to estimate the value of the function at the next point. While Euler's

method is valuable for its simplicity and ease of implementation, it is often inadequate for problems requiring high precision or involving stiff equations. Its first-order accuracy leads to larger errors and may necessitate very small step sizes to achieve acceptable results, which increases computational time. In contrast, we have Runge-Kutta methods which offer a more sophisticated approach to solving ODEs, providing greater accuracy without significantly increasing computational complexity. The most widely used version is the fourth-order Runge-Kutta method (RK4), which improves upon Euler's method by considering intermediate points within each step. For an ODE, $\frac{dy}{dt} = f(t, y)$, RK4 calculates the solution at t_{n+1} using the following steps:

$$k_1 = hf(t_n, y_n) \tag{2.6}$$

$$k_2 = hf(t_n + \frac{h}{2}, y_n + \frac{k_1}{2}) \tag{2.7}$$

$$k_3 = hf(t_n + \frac{h}{2}, y_n + \frac{k_2}{2}) \tag{2.8}$$

$$k_4 = hf(t_n + h, y_n + k_3) \tag{2.9}$$

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{2.10}$$

These intermediate calculations k_1 through k_4 account for the slope at several points within the interval, resulting in a weighted average that provides a much more accurate estimate of y_{n+1}. The fourth-order accuracy means that the error per step is proportional to h^5, making RK4 significantly more precise than Euler's method for the same step size.

The leapfrog method is another numerical technique, particularly useful for solving second-order differential equations and systems involving Hamiltonian mechanics. It is known for its simplicity and symplectic nature, which conserves the geometric properties of the phase space over time. For an ODE of the form $\frac{d^2y}{dt^2} = f(t, y)$, the leapfrog method updates positions and velocities in a staggered fashion:

$$y_{n+\frac{1}{2}} = y_n + \frac{h}{2}v_n \tag{2.11}$$

$$v_{n+1} = v_n + hf(t_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}) \tag{2.12}$$

$$y_{n+1} = y_{n+\frac{1}{2}} + \frac{h}{2}v_{n+1} \tag{2.13}$$

Here, y represents the position, v the velocity, and h the step size. By updating positions and velocities at half-step intervals, the leapfrog method provides a second-order accurate solution, meaning the error per step is proportional to h^3. Its conservation properties make it particularly advantageous for long-term integrations in physical systems where energy conservation is crucial.

2.2.1 Hydrodynamics

Building on the principles of N-body simulations, where the dynamics of collisionless matter like dark matter are primarily governed by gravitational interactions, hydrodynamics extends these simulations to include the behavior of baryonic (normal) matter, which is typically in the form of gas. Hydrodynamics is essential for modeling various astrophysical phenomena such as star formation, supernova explosions, and the evolution of galaxies. This extension requires solving fluid dynamics equations that govern the motion and interaction of gas particles.
The fundamental equations of hydrodynamics are the Navier-Stokes equations, which describe the motion of fluid substances. In the context of astrophysical simulations, these equations can be simplified by ignoring viscous and external forces, resulting in:
1. Continuity Equation (for conservation of Mass):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{2.14}$$

where ρ is the fluid density and $\mathbf{v}$ is the velocity field.
2. Euler Equation (for conservation of Momentum):

$$\frac{\partial (\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v}\mathbf{v}) = -\nabla P - \rho \nabla \Phi \tag{2.15}$$

Here, P is the pressure and Φ is the gravitational potential. The term $-\rho \nabla \Phi$ represents the gravitational force per unit volume.
3. For energy conservation:

$$\frac{\partial E}{\partial t} + \nabla \cdot [(E + P)\mathbf{v}] = 0 \tag{2.16}$$

where E is the total energy density.
However, the system of equations described above is not closed, as it contains five unknowns (ρ, $\mathbf{v}$, P, E, and Φ), but only three equations. To close the system and make it solvable, we need additional relationships.

Poisson Equation: To connect gravity with the density distribution, we use the Poisson equation:

$$\nabla^2 \Phi = 4\pi G \rho \tag{2.17}$$

where G is the gravitational constant, ρ is the mass density, and Φ is the gravitational potential. This equation links the gravitational potential Φ to the density field ρ, allowing us to solve for the gravitational potential once the density is known.

Equation of State: Additionally, the equation of state (EOS) relates the pressure P to the density ρ.

$$P = K \rho^{\gamma} \tag{2.18}$$

where K is a constant and γ is the adiabatic index (typically 5/3 for a monatomic ideal gas).
The EOS provides a relationship between pressure and density, which is essential for closing the system of equations and determining the pressure from the density.

2.3 Numerical Implementation of Cosmological Simulations

This section delves into the practical aspects of implementing cosmological simulations, translating the theoretical framework into computational algorithms. I start with discussing the numerical techniques employed for both N-Body and Hydrodynamical simulations. Following that, domain decomposition, a strategy for distributing computational workloads efficiently across parallel processors, a typical aspect of cosmological simulations is discussed as well.

2.3.1 N-Body simulations

N-Body simulations model the formation of dark matter halos and thus the subsequent formation and evolution of large-scale structures. The particles follow the equations of motion derived from Newton's laws, and their interactions are computed using techniques such as the particle-mesh (PM) method , particle-particle/particle-mesh (P^3M) method which are discussed in the following texts. For a system of N particles, each with mass m_i and position $\mathbf{r}_i$, the gravitational force $\mathbf{F}_i$ on particle i due to all other particles is given by:

$$\mathbf{F}_i = -G \sum_{j \neq i} \frac{m_i m_j (\mathbf{r}_i - \mathbf{r}_j)}{(|\mathbf{r}_i - \mathbf{r}_j|^2 + \epsilon^2)^{3/2}} \tag{2.19}$$

where G is the gravitational constant. To prevent numerical divergences and artificial two-body relaxation effects, gravitational softening is employed. This involves modifying the force law at small separations by introducing a softening length ϵ. Thus this is a softened gravitational force between particles i and j. This ensures that the force remains finite even when particles are very close, thereby stabilizing the numerical integration and mimicking the finite spatial resolution of the simulation. This direct summation approach scales as $\mathcal{O}(N^2)$, making it computationally expensive for large N. Thus we look for effective gravity solvers to balance computational cost and accuracy particularly for simulations with large number of particles. Few of these alternatives are Particle-Mesh Method, Particle-Particle/Particle-Mesh, Tree Codes.

Particle-Mesh Method: The Particle-Mesh (PM) method addresses the inefficiency of direct summation by interpolating particle masses onto a regular grid to compute the density field ρ. The Poisson equation is then solved on this grid to obtain the gravitational potential ϕ:

$$\nabla^2 \phi = 4\pi G \rho \tag{2.20}$$

The potential is typically solved using Fast Fourier Transforms (FFTs, reference: [112]), making the method scale as $\mathcal{O}(N \log N)$. The gravitational force is derived from the potential via:

$$\mathbf{F} = -\nabla \phi \tag{2.21}$$

The PM method is efficient for large-scale simulations but can be less accurate for small-scale interactions due to grid resolution limitations.

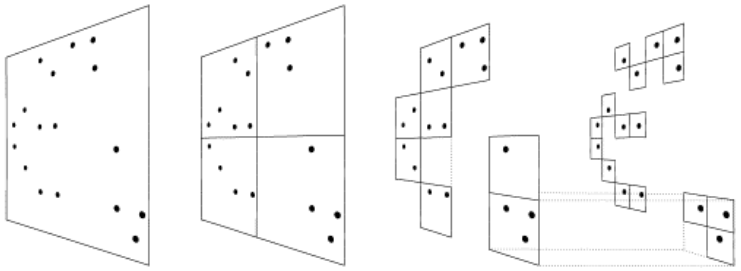

Figure 2.3.1: A 2D representation of a particle-based tree structure, where each region (square) recursively divides into four smaller regions until each subregion contains only one particle. When calculating forces, the algorithm traverses the tree structure and halts in any subregion where the angle subtended by that region falls below a predetermined threshold. Traversal of the tree also stops if a "leaf" has only one particle. Figure taken from [136].

Particle-Particle/Particle-Mesh Method: The Particle-Particle/Particle-Mesh (P^3M) method combines the PM method for long-range interactions with direct summation for short-range forces. This hybrid approach presented by [60] enhances accuracy while maintaining computational efficiency. The long-range forces are computed on a mesh, and the short-range forces are calculated using:

$$\mathbf{F}_i^{\text{short}} = -G \sum_{j \in \text{near}} \frac{m_i m_j (\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} \tag{2.22}$$

This method effectively captures small-scale dynamics while still benefiting from the efficiency of the PM approach for larger scales.

Tree Codes: Tree algorithms (examples include Barnes-Hut method [8]; k-dimensional tree KD-tree, see [139], [110] and [6]), use a hierarchical tree structure to approximate the gravitational forces. The simulation volume is recursively subdivided into a tree, and distant groups of particles are treated as single entities (multipole expansions). The force on a particle is computed by traversing the tree and summing contributions from individual particles or groups:

$$\mathbf{F}_i = -G \sum_{j \in \text{nodes}} \frac{m_j (\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} \tag{2.23}$$

Figure 2.3.1 shows a representation of the Barnes Hut structure taken from [136]. This approach reduces the computational complexity to $\mathcal{O}(N \log N)$, balancing accuracy and efficiency.

Tree-Particle-Mesh Codes: The Tree-Particle-Mesh (Tree-PM) algorithm is a hybrid approach used in cosmological simulations to efficiently compute gravitational interactions. It combines the PM method for long-range forces with a tree algorithm, like the Barnes-Hut method, for short-range interactions.
In the Tree-PM method, the simulation volume is partitioned using Orthogonal Recursive Bisection (ORB). ORB recursively divides the simulation space along

different spatial dimensions, splitting it at positions that balance the number of particles on both sides. This results in a hierarchical tree structure where each sub-volume contains approximately the same number of particles, ensuring a balanced computational load across processors in parallel computing environments. The large-scale nature of cosmological simulations and the complexity of the algorithms require a massive number of calculations. Cosmological simulations employ parallel computing techniques, where the computational workload is distributed across multiple processors. This parallelization allows for faster computation and the ability to handle larger simulations. The workload needs to be distributed well to support optimal balance between different computational nodes.
A significant challenge in parallelizing tree algorithms is the memory required to store the entire tree structure on each computing node. To overcome this, nodes store only a portion of the tree related to their local domain. When calculating forces, if a processor needs data from another domain, it requests a simplified or "pruned" version of the relevant tree branches from the corresponding node. This "pruned" tree contains only the essential information needed for the interaction calculations [38].
By exchanging these minimal tree structures, the Tree-PM algorithm reduces memory usage and communication overhead. This strategy allows for efficient parallel computation of gravitational forces, making it suitable for large-scale N-body simulations in cosmology, such as those performed with codes like GADGET.

2.3.2 Hydrodynamics

Numerically solving the equations of hydrodynamics involves discretizing both the fluid quantities and their governing equations to simulate the behavior of fluids in astrophysical systems. In practice, two primary computational approaches are employed:
1. Grid-Based Methods [26]: These methods discretize the simulation volume into a fixed or adaptive grid. The fluid equations are solved on this grid using finite difference, finite volume, or finite element techniques.

- Finite Volume Method: This grid-based technique divides the simulation domain into finite volumes (or cells) and solves conservation equations by integrating fluxes across the cell boundaries. This approach ensures the conservation laws are satisfied by integrating the fluid quantities over discrete control volumes.

- Adaptive Mesh Refinement (AMR, see [4], [25]): This technique dynamically adjusts the grid resolution, increasing it in regions requiring higher precision, such as shock fronts or regions of intense star formation.

2. Lagrangian Methods: Unlike Eulerian methods, which solve the equations of motion on a fixed grid, Lagrangian methods track the positions and velocities of discrete particles or fluid elements. One of the most prominent Lagrangian methods in cosmological simulations is Smoothed Particle Hydrodynamics (SPH,

e.g. [113]). This Lagrangian method represents the fluid with particles, each carrying mass, momentum, and energy. The fluid quantities are smoothed over a kernel function W, which dictates the influence radius of each particle.

- Density Estimation: The density at the position of particle i is calculated as:
$$\rho_i = \sum_j m_j W(\mathbf{r}_i - \mathbf{r}_j, h) \tag{2.24}$$
where m_j is the mass of particle j, $\mathbf{r}_i$ and $\mathbf{r}_j$ are the positions of particles i and j, respectively, and h is the smoothing length.

- Momentum Equation: The momentum equation in SPH form for particle i is:
$$\frac{d\mathbf{v}_i}{dt} = -\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} \right) \nabla W(\mathbf{r}_i - \mathbf{r}_j, h) \tag{2.25}$$
where P_i and P_j are the pressures at particles i and j, respectively

The use of smoothing kernels can introduce artifacts and reduce accuracy in certain scenarios. On the other hand, SPH methods can accurately capture the dynamics in regions of high density contrast, making them ideal for simulating astrophysical phenomena such as shock waves, turbulence, and star formation (for more, see [132] and [102]). They have been widely applied in cosmological simulations and are the framework for the simulation code used in the current work. However, SPH's reliance on artificial viscosity for shock handling can lead to excessive dissipation, and its particle-based approach may struggle to resolve small-scale turbulence and mixing processes effectively. These limitations highlight the need for careful implementation and high-resolution settings in shock-dominated or turbulent regimes.

2.3.3 Domain decomposition

The *domain decomposition* is a required technique in simulations used to distribute the computational workload across multiple processors. This method divides the simulation volume into smaller subdomains, each handled by a different processor, enabling parallel computation and thus significantly enhancing the efficiency and scalability of the simulations.
The entire simulation volume, which contains N particles, is divided into P subdomains, where P is the number of available processors. Each subdomain is assigned a subset of the particles, and the processor responsible for a subdomain computes the forces, updates the positions, and handles the interactions of particles within its domain. The primary goal of domain decomposition is to ensure load balancing and minimize communication overhead between processors.
There are several strategies for this:
1. Regular Grid Decomposition: The simulation volume is divided into a regular grid of equal-sized cells, with each cell assigned to a different processor. For example, in a 3D simulation, the volume could be divided into $P = n_x \times n_y \times n_z$ cells, where n_x, n_y, and n_z are the number of divisions along the x, y, and z axes,

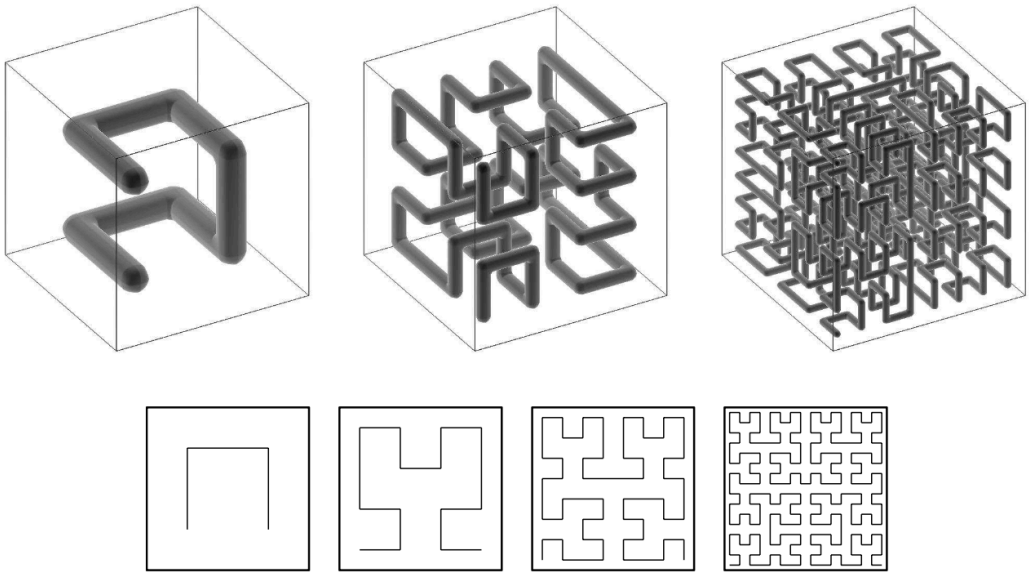

Figure 2.3.2: Space-filling Peano-Hilbert curve, 3D and 2D representation, taken from [133]

.

respectively. But this approach could result in assigning all dense region to one and the computation of low density regions to another, easily unbalancing the scales.
2. Octree and Quadtree Decomposition: The simulation volume is recursively subdivided into smaller regions based on the distribution of particles. This hierarchical approach, using octrees in 3D and quadtrees in 2D, adapts to non-uniform particle distributions, ensuring a more balanced load across processors. A disadvantage would be a higher communication overhead.
3. Space-Filling Curves: This method maps the 3D simulation volume onto a 1D space-filling curve (e.g., Hilbert or Peano curve). The curve is then divided into segments, with each segment assigned to a processor. This approach maintains spatial locality and reduces communication overhead. Most contemporary simulation codes use Peano-Hilbert space-filling (e.g. GADGET, [133]). Figure 2.3.2 show a sketch of the Peano curve taken by [133].
The implementation of space-filling curves involve:

- Mapping Particles to the Curve: Each particle in the 3D simulation volume is assigned a position on the space-filling curve. This is done by converting the particle's 3D coordinates into a single parameter t on the 1D curve.

- Partitioning the Curve: The 1D curve is divided into P contiguous segments, where P is the number of processors. Each processor is assigned a segment, ensuring that each processor handles a contiguous region of the 3D space.

Since the particle distribution may be non-uniform, the length of each segment on the curve can be adjusted to balance the computational load among processors.

This dynamic adjustment ensures that each processor has an approximately equal number of particles to process.
Domain decomposition in large-scale cosmological simulations presents several challenges such as communication overhead due to the need to share particle information across subdomain boundaries, load balancing issues arising from uneven particle distribution, and the difficulty of maintaining an efficient tree structure for long-range gravitational force calculations. Additionally, dynamic changes in particle distribution require frequent repartitioning of domains, increasing computational overhead. The accuracy of interactions near subdomain boundaries can introduce small errors, and memory management becomes more complex as the number of particles increases.

2.4 Cosmological simulations

Cosmological simulations are very complex due to the need to model a vast array of physical processes across multiple scales and regimes.
Beyond gravity and hydrodynamics, simulations must model a wide range of astrophysical processes as:

- Radiative Cooling: Implementing cooling functions that depend on gas temperature, density, and chemical composition to simulate how gas loses energy and condenses.
- Star Formation: Using subgrid models to convert cold, dense gas into stars based on criteria like density thresholds and temperature conditions, often following empirical laws like the Schmidt law.
- Stellar Feedback: Incorporating the energy and momentum input from processes like supernova explosions and stellar winds, which heat the surrounding gas and can trigger or suppress further star formation.
- Dust Processes: Including the formation and evolution of dust grains, which affect the opacity of the interstellar medium and influence cooling rates and chemical reactions.
- Black Hole Dynamics and Feedback: Modeling the growth of supermassive black holes through accretion (using formulas like the Bondi-Hoyle-Lyttleton accretion rate) and their feedback mechanisms, which inject vast amounts of energy into their host galaxies and can regulate star formation on galactic scales.

Each of these processes require specialized algorithms and numerical methods, often with their own sets of equations and computational challenges. The interplay between different physical phenomena adds layers of complexity, as processes like star formation and black hole feedback are tightly coupled and can have cascading effects on the simulation.
Several specialized simulation codes have been developed for cosmological studies. Examples include GADGET (refer [137], [133]), RAMSES (refer [145]), AREPO

(refer [132]) and ChaNGa (refer [64], [65] and [98]). These codes implement the numerical methods described above and are optimized for high-performance computing environments.
High-Performance Computing (HPC) refers to the use of powerful computational resources to solve complex problems requiring substantial processing power, memory, and storage. It involves clusters of interconnected computers working in parallel. MPI (Message Passing Interface) is a standardized protocol that facilitates communication between nodes in distributed computing environments, enabling tasks to be divided efficiently. Job efficiency can be determined with respect to how effectively computational resources are used or in terms of time or energy, while scalability describes how well a system performs as resources increase. HPC systems leverage CPUs for general-purpose computations and increasingly rely on GPUs for highly parallelizable tasks due to their speed in processing simultaneous operations.

2.4.1 Cosmology with GADGET-3

In this section, an overview of GADGET code is provided specifically, as this is code used to perform all the simulations related to my project.
GADGET-3 is an advanced version of the widely-used cosmological simulation code, GADGET (GAlaxies with Dark matter and Gas intEracT), developed by Volker Springel. It is a highly efficient, parallel code designed for large-scale cosmological simulations, focusing on both collisionless dark matter and hydrodynamic gas.
TreePM for gravity and SPH is employed to discretize and solve hydrodynamical equations. To efficiently handle the vast range of dynamical timescales in cosmological simulations, GADGET-3 employs adaptive time stepping for particles. The time step Δt for a particle is determined based on its local conditions, such as the acceleration a and the Courant condition for SPH particles. The code is designed to run on massively parallel supercomputers, using the MPI for communication between processors. It can scale efficiently to handle billions of particles across thousands of processors as shown in Figure 2.4.3. Strong scalability specifically evaluates performance improvements when the problem size remains fixed but the number of processors increases. Speedup refers to the improvement in computational performance when a task is parallelized or executed on more powerful hardware, typically measured as the ratio of the time taken to complete a task on a single processor to the time taken on multiple processors or nodes.
In the current work I use an older, yet relatively stable version of OPENGADGET3 (version used in [52] and configured similar to that of [10]). The main upgrades in the OPENGADGET3 code version is the refactoring in tree algorithm for improved computational efficiency and also, the code optimizations made with OpenMPI (an open-source implementation of MPI, widely used for its flexibility and performance), making it a scalable choice for large scale simulations considering its enhanced parallel processing capabilities.
The subgrid physics are following the implementations detailed in [10].
In the following subsections, I present algorithms and numerical methods, imple-

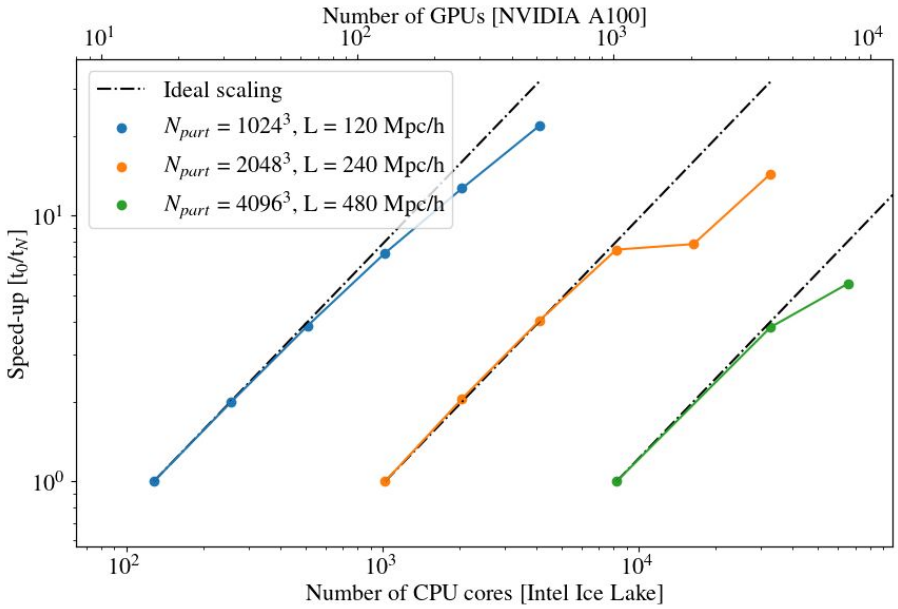

Figure 2.4.3: Strong scaling scalability of the OPENGADGET code on LEONARDO supercomputer at CINECA. The figure presents the scalability in terms of number of CPU Cores and GPUs for cosmological box simulations of different box size and number of particles. Credits: Dr. Luca Tornatore (INAF, Trieste).

mented in GADGET code to model the main astrophysical processes.

2.4.2 Initial conditions

Initial conditions (ICs) in simulations typically reflect the state of the universe shortly after the Big Bang, during the epoch of cosmic inflation, when quantum fluctuations were stretched to macroscopic scales, seeding the formation of large-scale structures. These fluctuations are characterized by a nearly scale-invariant power spectrum, consistent with the predictions of inflationary cosmology (refer [55], [88]). The ICs are drawn based on perturbation theory (see [24], [15]) which describes the early universe's density fluctuations .
The density field, $\delta(\mathbf{x})$, representing the fractional overdensity at position $\mathbf{x}$, is typically expressed in Fourier space:

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3} \int \delta(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d^3\mathbf{k}, \tag{2.26}$$

where $\delta(\mathbf{k})$ is the Fourier transform of the density field. The power spectrum, $P(k)$, of these fluctuations is defined as:

$$\langle \delta(\mathbf{k}) \delta^*(\mathbf{k}') \rangle = (2\pi)^3 P(k) \delta_D(\mathbf{k} - \mathbf{k}'), \tag{2.27}$$

where δ_D is the Dirac delta function, ensuring that different Fourier modes are uncorrelated. The primordial power spectrum is often parameterized as:

$$P(k) = A_s \left(\frac{k}{k_0} \right)^{n_s} T^2(k), \tag{2.28}$$

where A_s is the amplitude, n_s is the spectral index, and $T(k)$ is the transfer function that accounts for the evolution of perturbations through the radiation-dominated and matter-dominated eras [7].
As the universe expands, regions with slightly higher density than their surroundings start to collapse under their own gravity. This process is described by the Jeans instability criterion [63], which indicates that perturbations on scales larger than the Jeans length will grow over time. The formation of bound structures, or halos, occurs as particles collapse into overdense regions.
To create the initial condition for numerical simulations, a Gaussian random field is generated in Fourier space for a given power spectrum. The real and imaginary parts of $\delta(\mathbf{k})$ are drawn from a normal distribution with zero mean and variance proportional to $P(k)$. The inverse Fourier transform of $\delta(\mathbf{k})$ is computed to obtain the density field $\delta(\mathbf{x})$ in real space. The particles' initial positions and velocities are determined using the Zeldovich approximation [157], which provides a first-order Lagrangian perturbation theory solution. The displacement field, $\mathbf{\Psi}(\mathbf{q})$, is given by:

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + D(t)\mathbf{\Psi}(\mathbf{q}), \tag{2.29}$$

where $\mathbf{q}$ is the initial Lagrangian position, $\mathbf{x}$ is the Eulerian position, and $D(t)$ is the linear growth factor. The velocity field is then:

$$\mathbf{v}(\mathbf{q}, t) = \dot{D}(t)\mathbf{\Psi}(\mathbf{q}), \tag{2.30}$$

where $\dot{D}(t)$ is the time derivative of the growth factor.
The ICs are generated at a high enough redshift well within the linear regime, usually above z=50
Two common approaches to setting up ICs are the "box" or "periodic box" simulations and the "zoom-in" simulations.

Box simulations involve modeling a representative cubic volume of the universe with periodic boundary conditions. This approach is widely used to study the statistical properties of large-scale structures, such as the cosmic web, galaxy clustering, and the formation of dark matter halos. For box simulations, the size of the cubic volume, L, typically ranging from tens to hundreds of megaparsecs (Mpc) on each side. The choice of volume depends on the scale of structures of interest and the computational resources available. The ICs are generated as described earlier. Following which a periodic boundary conditions is implemented to mimic an infinite universe, ensuring that particles leaving one side of the box re-enter from the opposite side.

Zoom-in simulations focus on a specific region within a larger cosmological volume. This technique allows for high-resolution simulations of individual objects, such as galaxies or clusters, embedded within their larger cosmic environment. A region or object of interest is identified within a larger low-resolution simulation. This region is typically selected based on its mass, environment, or specific structural properties. Then, the initial conditions are created with varying resolution:

- High-Resolution Region: The region of interest is sampled with a much higher particle density to resolve fine structures.

- Low-Resolution Surrounding: The surrounding volume is sampled with lower particle density to reduce computational costs.

Using the high-resolution ICs, the selected region is re-simulated with enhanced accuracy, capturing detailed physical processes such as gas dynamics, star formation, and feedback mechanisms.

2.4.3 Cooling

Cooling processes are crucial for accurately simulating the thermal evolution of the gas in cosmological simulations. These processes include radiative cooling, which is essential for modeling phenomena such as the formation of stars, the behavior of interstellar and intergalactic media, and the interactions of galaxies. The cooling of the gas in GADGET-3 is primarily governed by radiative processes. The rate at which a gas particle loses energy due to radiative cooling is described by the cooling function $\Lambda(T)$, which depends on the gas temperature T and its density ρ. The cooling function $\Lambda(T)$ represents the amount of energy radiated away per unit volume and per unit time as a function of temperature. In GADGET-3, this function is typically obtained from theoretical models or empirical fits. For a gas particle i, the rate of energy loss due to cooling is given by:

$$\frac{dE_i}{dt} = -\rho_i \Lambda(T_i) \tag{2.31}$$

where: - E_i is the internal energy of particle i, - ρ_i is the density of particle i, - $\Lambda(T_i)$ is the cooling function evaluated at the temperature T_i of the particle. Other processes like stellar and blackhole feedback affects the thermal state of gas particles by counteracting cooling.
The cooling function $\Lambda(T)$ is precomputed and stored in a lookup table for efficiency (see [151] and [150]). Mainly the effects of hydrogen and helium with their different ionization states are included. During each time step, GADGET-3 interpolates this table to find $\Lambda(T_i)$ for each gas particle based on its current temperature. The updated internal energy is then used to recalculate the thermodynamic properties of the gas, influencing the subsequent evolution of the simulation. The energy equation is solved with the cooling terms integrated implicitly, ensuring stability even when cooling rates change rapidly, while adiabatic terms (e.g., energy changes due to gas compression or expansion) are integrated explicitly. The time step for these calculations is set to half the shortest time step of any SPH particle in the simulation, ensuring stability across the system. Furthermore, to avoid numerical instabilities, the cooling rate is dampened so that a gas particle cannot radiate more than half its thermal energy during a single time step.

2.4.4 Thermal Conduction

Thermal conduction is an important physical process, governing the transfer of heat within the interstellar and intergalactic media. It can significantly impact the temperature distribution in simulations. In galaxy clusters, for example, it can lead to the redistribution of thermal energy, affecting the cooling flow and the thermal structure of the intracluster medium.
Thermal conduction in a plasma is described by the heat flux $\mathbf{q}$, which is proportional to the temperature gradient ∇T. The heat flux is given by:

$$\mathbf{q} = -\kappa \nabla T \tag{2.32}$$

where κ is the thermal conductivity.
Then, in GADGET-3, thermal conduction is implemented by solving the conduction equation, which in its general form is:

$$\frac{\partial u}{\partial t} = \frac{1}{\rho} \nabla \cdot (\kappa \nabla T) \tag{2.33}$$

where:

- u is the internal energy per unit mass,
- ρ is the density,
- κ is the thermal conductivity,
- T is the temperature.

This equation describes the time evolution of the internal energy due to thermal conduction.
The numerical implementation of thermal conduction in GADGET follows the approach outlined in [66]. The conduction equation is reformulated in the SPH framework to account for energy conservation and numerical stability. The Laplacian operator is discretized using kernel interpolation methods, ensuring that the calculation involves only first-order derivatives of the smoothing kernel to minimize noise. The discretized form of the conduction equation is:

$$\frac{du_i}{dt} = \sum_j m_j \frac{\kappa_i + \kappa_j}{\rho_i \rho_j} \frac{T_j - T_i}{|x_{ij}|^2} x_{ij} \cdot \nabla_i W_{ij}, \tag{2.34}$$

where:

- T_i and T_j are the temperatures of particles i and j,
- κ is the conductivity,
- W_{ij} is the smoothing kernel,
- x_{ij} is the distance between particles i and j.

This is solved for each timestep using an explicit time-stepping scheme. For a given particle i, the total change in internal energy is calculated as the sum of contributions from all neighboring particles j within its smoothing radius. To ensure stability, the timestep is adaptively shortened in regions with steep temperature gradients, preventing unphysical results or numerical instabilities.

2.4.5 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 of stellar feedback on the surrounding gas. For reference model of star formation and feedback implemented, please check [135]. Here I only provide 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 $\rho_{\rm th}$. 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 $T_{\rm th}$. 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 ($M_\odot\,{\rm yr}^{-1}$). 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 for a gas particle i is given by:

$$\dot{M}_\star = \dot{\rho}_{\star,i} = (1-\beta)\frac{\rho_i}{t_\star} \tag{2.35}$$

where:

- $\dot{\rho}_{\star,i}$ is the star formation rate density,

- β is the mass fraction of massive stars,

- ρ_i is the gas density of particle i,

- $t_\star = \frac{t_{\rm dyn}}{c_\star}$ is the star formation timescale configured according to Schmidt's law [127].

The actual process of star formation is implemented stochastically, converting gas particles into star particles according to the computed SFR. β is considered for stars that are more massive than M_{sun} for a given initial population. This, is thus specific to the initial mass function (IMF) that is used.

Stellar Feedback Stellar feedback in GADGET-3 encompasses several processes, including supernova (SN) explosions, stellar winds, and radiation pressure. 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 cold phase via thermal conduction. This recycles in the dense gas clouds back to ambient phase. The evaporated cold gas mass follows,

$$\dot{\rho}_c = -A\beta\frac{\rho_c}{t_\star} \tag{2.36}$$

where: - A is the evaporation efficiency and assumed to follow a theoretically modelled dependence on density of the ISM $A \propto \rho^{-\frac{4}{5}}$
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. At a given timestep, gas particles are probabilistically converted to star particles based on the computed star formation rate. The model evolves gas particles probabilistically into star particles based on a star formation rate. For each timestep, a new star particle is spawned if a random number drawn uniformly falls below the probability:

$$p_\star = \frac{m}{m_\star}\left(1 - \exp\left[-\frac{(1-\beta)x\Delta t}{t_\star}\right]\right), \tag{2.37}$$

where $m_\star$ is the mass of a single star particle.
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 [135]. We have, hot gas associated with SN, converted into wind particles following a mass loss proportional to SFR :

$$\dot{M}_w = \eta\dot{M}_\star \tag{2.38}$$

where: - η is the efficiency parameter, assuming a conversion factor η =2. This conversion depends on probability of a generated random number to be less than a limit of:

$$p_w = 1 - \exp\left[-\frac{\eta(1-\beta)x\Delta t}{t_\star}\right] \tag{2.39}$$

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 set by hand, a $v_w = 350$ km s^{-1} corresponds to a fixed fraction of 50% of the supernovae energy converted into the kinetic energy carried by the wind.
Supernovae feedback is implemented together with metal enrichment following the prescriptions presented in [150]. A metal dependent cooling function is used to compute the respective energy radiated due to the quantity of metals produced as mentioned in section 2.4.3. 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 modelling respective to type-Ia, II and asymptotic giant branch (AGB) stars; see [96], [147], [155], [121] and [73].

2.4.6 Blackhole and associated feedback

The base reference for blackhole implementation is [134]. 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 [21]), modified for a realistic astrophysical context:

$$\dot{M}_{\mathrm{BH}} = \alpha \frac{4\pi G^2 M_{\mathrm{BH}}^2 \rho}{(c_s^2 + v^2)^{3/2}} \tag{2.40}$$

where:

- $\dot{M}_{\mathrm{BH}}$ 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:

$$\dot{M}_{\mathrm{Edd}} = \frac{4\pi G M_{\mathrm{BH}} m_p}{\epsilon_r \sigma_T c} \tag{2.41}$$

where:

- $\dot{M}_{\mathrm{Edd}}$ 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:

$$\dot{E}_{\text{feed}} = \epsilon_f \epsilon_r \dot{M}_{\text{BH}} c^2 \tag{2.42}$$

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 M_{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:

$$p_j = w_j \frac{\dot{M}_{\text{BH}} \Delta t}{\rho}, \tag{2.43}$$

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.

2.4.7 Clustering Algorithms

Clustering algorithms are essential in identifying and characterizing structures such as dark matter halos. These algorithms analyze the spatial distribution of particles to find regions of high density that correspond to astrophysical objects of interest.

The Friends-of-Friends (FoF) algorithm is a widely used method for identifying dark matter halos in cosmological simulations. The FoF algorithm was first introduced by [31] and has since become a standard tool in cosmological simulations due to its simplicity and efficiency. The algorithm operates by linking together particles that are within a specified distance, known as the linking length, thereby forming groups or "friends-of-friends".
The linking length is a critical parameter in the FoF algorithm. It is usually set as a fraction of the mean inter-particle separation. The fraction is typically set to 0.2 for identifying dark matter halos. For each newly linked particle, the process is repeated until the given structure is complete of linked particles and no more can be linked.

Subfind Algorithm is an advanced algorithm used to identify substructures within dark matter halos, which are themselves identified by the FoF algorithm. This algorithm helps to distinguish bound substructures, known as subhalos, within a parent halo. The Subfind algorithm was developed by [131] and has been widely used in the analysis of high-resolution cosmological simulations.
Starting with the halos identified by the FoF algorithm, Subfind estimates the local density around each particle using a kernel interpolation technique, typically employing a smoothing length that encompasses a fixed number of nearest neighbors. The gravitational potential is calculated for all particles in the FoF group. At this point, substructures are identified by finding density peaks and tracing the boundaries where the density gradient points towards the peak. In Gadget3, an updated version of Subfind is used [34], which accounts for baryons as well; gas, dark matter, stars and blackholes are all affixed to a given halo and the substructures within are distinguished.
Thus a two-step process is used, with FoF (quick, executed at runtime) and Subfind (computationally expensive, run as part of post-processing using FoF prefiltering), each algorithm handling different regimes: one based on distance-based linking and another identifying gravitationally bound regions. This approach provides a trade-off between efficiency and precision to capture the large scale and internal structures effectively.

Figure 2.5.4: James Webb Telescope image of the galaxy cluster SMACS 0723. In the image, the distortion of galaxy shapes due to gravitational lensing and the complexity of those structures are clearly visible. Credits: https://webbtelescope.org/.

2.5 Galaxy clusters

Chaotic variations on cosmological simulations are present at all scales, from single-object simulations of an isolated galaxy to cluster and large-scale structures. In my work, I am focusing on the impact of chaos in the simulated galaxy clusters.

Galaxy clusters are the largest gravitationally bound structures in the universe, containing hundreds to thousands of galaxies, as well as hot gas and dark matter. They are an overdensity of galaxies surrounded by a diffuse plasma, not associated with individual galaxies; both in equilibrium within a common gravitational potential well dominated by dark matter. Dark Matter being the gravitational backbone of the cluster constitutes about 85% of the cluster's mass. Intra-Cluster Medium (ICM) i.e. hot, X-ray emitting gas accounts for about 10-15% of the total cluster mass. Galaxies are embedded within the dark matter halo and contribute only about 1-2% of the mass. Galaxy clusters form and evolve through hierarchical clustering, where smaller structures merge to create larger ones. This process is driven by gravitational collapse and influenced by the distribution of dark matter and baryonic matter. Observationally, clusters are characterized by their X-ray Emission. The hot gas in the ICM emits X-rays, providing the data on the temperature and density profiles. Distortions in the CMB caused by the

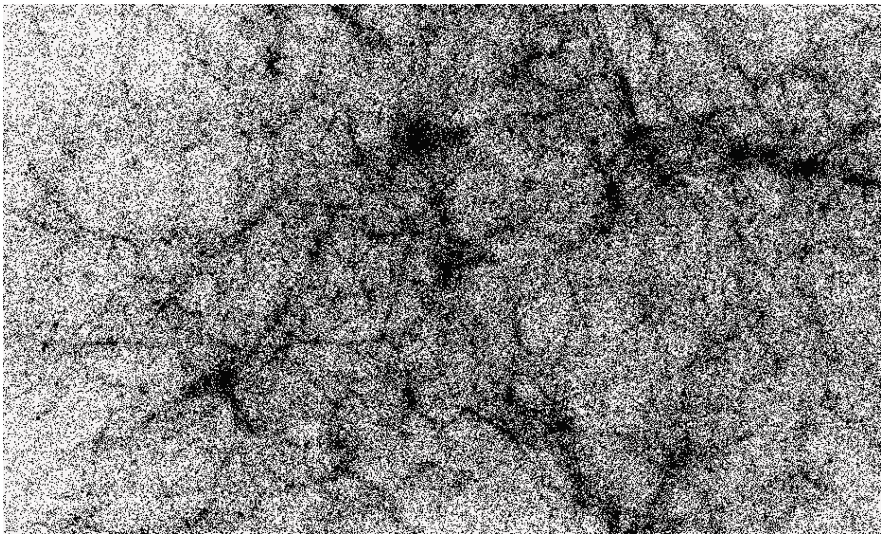

Figure 2.5.5: Projected density field for the dark matter only, high resolution Dianoga simulation. The densest regions are showed in black and corresponds to clusters and groups of galaxies.

interaction with the ICM (Sunyaev-Zel'dovich Effect), when combined with the x-ray data provides more details of the structure of galaxy clusters. Clusters also act as lenses, bending light from background objects, revealing the distribution of dark matter. In Figure 2.5.4 we can see an example of a cluster as observed by James Webb telescope.
Today, galaxy clusters serve as a valuable tool for both cosmological studies -particularly through analyses of their number counts (e.g., [2], [80]) and for models of galaxy formation. Positioned at the nodes of the cosmic web (see e.g., Figure 2.5.5), galaxy clusters reside in the densest regions of the universe, where galaxy evolution and resulting properties are shaped by a variety of environmental factors.
In cosmological simulations, galaxy clusters form as a result of the evolution of the primordial density field as discussed in Section 2.4.2. They form and evolve via hierarchical clustering, a process where smaller halos merge to form larger structures. This is a natural consequence of the CDM model, where small-scale fluctuations collapse first, and larger structures form from the merging and accretion of these smaller halos (see [111], [153]). The rate of mass accretion and the merger history of halos can be quantified using merger trees derived from simulations. These trees track the formation and growth of halos over time and are essential for understanding the assembly history of galaxy clusters (see [81]).
The concentration parameter describes the density profile of a cluster, often characterized by the Navarro-Frank-White (NFW) profile [103]:

$$\rho(r) = \frac{\rho_s}{\frac{r}{r_s}\left(1+\frac{r}{r_s}\right)^2} \tag{2.44}$$

where ρ_s is the characteristic density equal to twice the density at r_s and r_s is the scale radius. The concentration parameter c is defined as:

$$c = \frac{R_{200}}{r_s} \tag{2.45}$$

The clusters are bifurcated with a virial radius $R_{\rm vir}$ which is defined as the radius within which the mean density of the cluster is Δ_c times the critical density of the universe:

$$\Delta_c = \frac{M_{\rm vir}}{\frac{4}{3}\pi R_{\rm vir}^3 \rho_c} \tag{2.46}$$

A common choice for Δ_c is 200, denoted as R_{200}.
In simulations, the clusters are thus typically defined, through their gravitationally bound structures and then, specific parameters such as mass, temperature are computed. For instance, the mass of a galaxy cluster is typically computed within a certain radius, for e.g. with virial radius $R_{\rm vir}$, where the density is a specific multiple of the critical density ρ_c:

$$M_{\rm vir} = \int_0^{R_{\rm vir}} 4\pi r^2 \rho(r)\, dr \tag{2.47}$$

All of the generic quantities are computed in post-processing by Subfind; a catalog for clusters and galaxies are provided with all the computed properties such as mass, temperature, circular velocity etc.

2.5.1 Observations and Simulations

Galaxy stellar mass function (GSMF) is one of key statistical measures that represents how many galaxies exist per unit co-moving volume in each stellar mass bin and is often expressed as a function, $\phi(M_*)$. Figure 2.5.6 shows the Galaxy stellar mass function for Dianoga galaxy (10x better resolution than the cluster simulations run within the context of this work) at z=0. In the figure 2.5.6, the stellar mass from the cluster galaxies are plotted with 3 different measures: as found in Subfind catalog, as well as filtered within a 2D aperture of 2 different radius from the galaxy center, all of which do not significantly differ from each other. Bassini et al. (2020) GSMF obtains an overall close agreement with the observed data from [16], except for a higher number density for lower mass bins (for $M_* \sim 10^{10} M_\odot$) which is attributed to effects of stellar feedback.
Different implementations of the physical mechanisms impact the properties of galaxies at macroscopic scales, for instance, how feedback processes can affect the star formation history which can impact how galaxies of different stellar masses evolve. Figure 2.5.7 shows the evolution of cosmic star formation rate density (SFRD) of cosmological simulations SIMBA [125] with a $50h^{-1}cMpc$ volume. Effects of different feedback implementations are tested: stellar and 3 different AGN feedback implementations i.e. winds and jets and x-ray radiation. Tests include separate runs for each, an all inclusive fiducial run and one with no feedback. Results were tested for convergence with different volumes and resolution. As seen in the figure, no feedback results in an overproduction of stars throughout the evolution compared the with-feedback contemporaries. The results reflect the

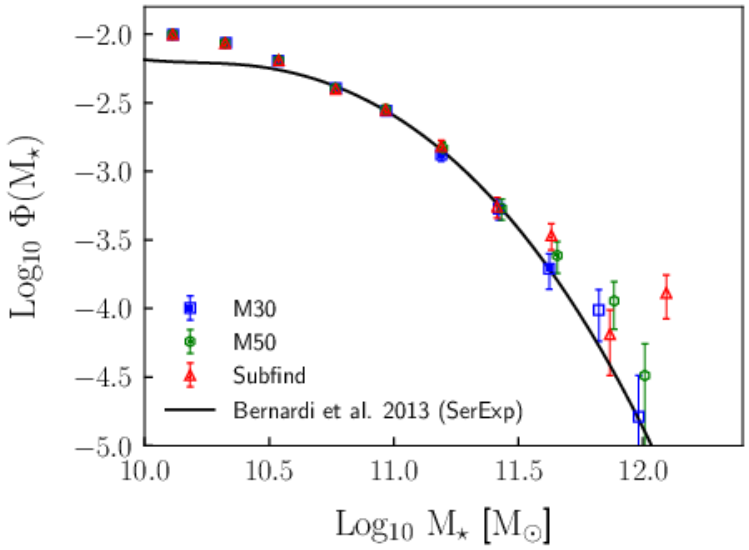

Figure 2.5.6: Galaxy stellar mass function (GSMF) for Dianoga 10x simulations compared to observational data at z=0, see text for more details. Credits: [10]

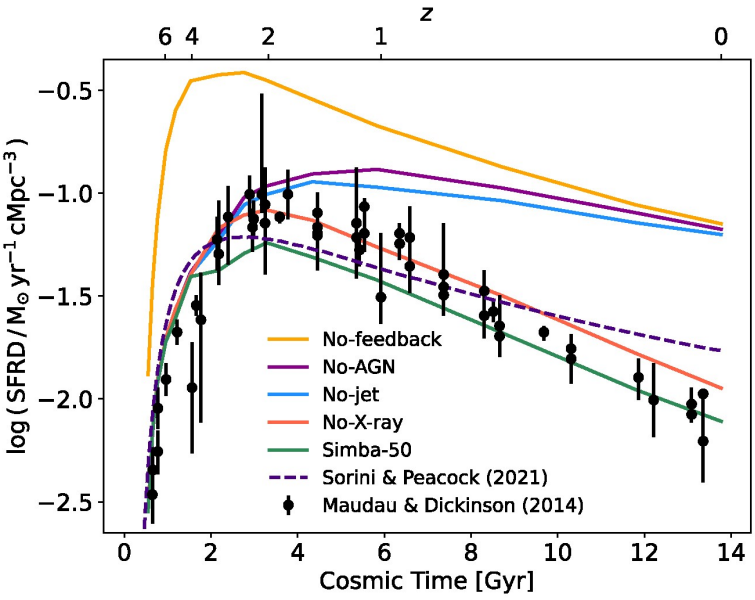

Figure 2.5.7: Cosmic star formation rate density (SFRD) with cosmological simulations SIMBA, with and without the effects of stellar and AGN feedback (X-ray, jet and winds) compared with survey results listed in [93] and [130] model for cosmic starformation. See [125] for more details.

important role of both stellar feedback and AGN feedback in regulating the star formation at different epochs. Stellar feedback is controlling SFRD at earlier time ($z \simeq 2$) and AGN driven feedback (particularly their jet mode) having a marked effect on overall star formation at later times. See [125] for more details.

Chapter 3

Chaos in Numerical Systems

The origins of chaos theory trace back to the late 19th century, with the pioneering work of the French mathematician Henri Poincaré, who first revealed the sensitive dependence on initial conditions within certain dynamical systems [109]. Poincaré's investigations into the three-body problem (analyzing the motion of three celestial bodies under their mutual gravitational attraction) led him to recognize that even small deviations in initial conditions could result in vastly different outcomes–a hallmark of what would later be known as chaotic behavior. Despite Poincaré's insights, the implications of chaos remained largely theoretical, overshadowed by the prevailing belief in deterministic predictability within classical physics.

The true advent of chaos theory as a field of study came much later, in the mid-20th century, with the advent of digital computers, which allowed for the extensive numerical exploration of complex dynamical systems. In 1963, meteorologist Edward Lorenz published his groundbreaking paper "Deterministic Nonperiodic Flow" [90], in which he described how a simple model of atmospheric convection could exhibit unpredictable behavior. Lorenz discovered that minute differences in initial conditions, such as those introduced by rounding errors in numerical computations, could lead to dramatically different outcomes over time. This phenomenon, famously illustrated by the metaphor of a butterfly flapping its wings and causing a tornado weeks later, became known as the "butterfly effect". Lorenz's equations, now known as the Lorenz system, demonstrated that small changes in initial conditions could lead to vastly different results. The Lorenz system is given by:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x), \\ \frac{dy}{dt} &= x(\rho - z) - y, \\ \frac{dz}{dt} &= xy - \beta z, \end{aligned} \tag{3.1}$$

where σ, ρ, and β are parameters that lead to chaotic behavior for certain values. Lorenz's work was crucial, as it demonstrated that deterministic systems could nonetheless behave unpredictably, challenging the classical notion that future states of a system could always be predicted if its initial conditions were known

with sufficient precision.

Around the same time, in 1961, Benoît Mandelbrot introduced the concept of fractals, which are intricate geometric shapes that exhibit self-similarity at different scales. Mandelbrot's work on the statistical properties of natural landscapes and other irregular structures provided a visual and mathematical framework to understand the complex patterns often associated with chaotic systems [95].

Following this, in the 1970s, Robert May's research on population dynamics highlighted the ubiquity of chaos in biological systems. His work on the logistic map demonstrated how simple nonlinear equations could produce chaotic behavior [97]. The logistic map is defined as:

$$x_{n+1} = r x_n (1 - x_n), \tag{3.2}$$

where x_n represents the population at generation n and r is a parameter controlling the growth rate. The logistic map is often studied using a bifurcation diagram, which shows the different long-term behaviors of the system as the parameter r varies. For varying values of r, the system switches from steady state to periodic behavior. As r increases beyond a critical threshold, the logistic map begins to show chaotic behavior. This means that the system becomes highly sensitive to initial conditions, leading to seemingly random and unpredictable fluctuations in the population.

Around the same time as May's research, Mitchell Feigenbaum made significant contributions to the field by establishing the connection between chaos and phase transitions. Feigenbaum discovered universal constants that characterize the onset of chaos in a wide variety of systems. His work on period-doubling bifurcations showed that the transition to chaos follows a predictable pattern, governed by the Feigenbaum constants [43]. These works later led to experimental studies of chaotic effects in various other fields.

As digital computing power increased, scientists began to explore a wider range of nonlinear dynamical systems, uncovering chaotic behavior in contexts ranging from fluid dynamics to electrical circuits. However, numerical simulations of these systems revealed a new and unexpected challenge: the onset of numerical chaos. Numerical chaos occurs when small computational errors, such as those from round-off errors or discretization, grow exponentially over time, leading to results that deviate significantly from the true solution. Even in systems that are theoretically well-behaved, the limitations of numerical precision can cause simulations to produce chaotic outcomes, effectively masking the true dynamics of the system under study.

In the 1970s and 1980s, the study of numerical chaos gained momentum as researchers sought to understand and mitigate these instabilities. The realization that chaos could arise not only from the intrinsic dynamics of a system but also from the numerical methods used to simulate it prompted a re-evaluation of computational practices. This led to the development of more robust algorithms designed to minimize numerical errors, such as adaptive step-size control and high-order integration methods.

3.1 Emergence of Chaos in Simulations

One significant aspect of numerical chaos is its impact on long-term simulations. In fields like weather forecasting, climate modeling, and astrophysics—where predictions over extended periods are crucial—the accumulation of small numerical errors can lead to drastically incorrect predictions. This challenge was famously highlighted during the development of the first digital weather models, where forecasts would diverge rapidly from observed conditions despite seemingly accurate initial data. As a result, ensemble forecasting techniques, where multiple simulations with slightly varied initial conditions are run to estimate the range of possible outcomes, became a standard approach to account for the inherent uncertainty introduced by numerical chaos.
Numerical chaos in simulations is typically caused by several factors that can significantly affect the accuracy and reliability of the results. The primary causes of variation include:

1. **Discretization Errors**: When continuous equations are approximated by discrete steps in numerical simulations, the choice of time step or grid size can lead to errors. If the time step is too large, the simulation may not accurately capture the dynamics of the system, leading to instability.

2. **Round-off Errors**: Due to the finite precision of computer arithmetic, small errors introduced at each step of a simulation can accumulate over time, especially in long-term simulations or those involving chaotic systems. This accumulation can cause significant deviations from the expected outcomes.

3. **Truncation Errors**: These errors arise from approximating an infinite series or integral by a finite sum. If the higher-order terms are significant and neglected, it can result in instability.

4. **Inadequate Resolution**: Insufficient spatial or temporal resolution can lead to an inaccurate representation of the physical system being modeled, resulting in numerical instability. For instance, coarse grid sizes in fluid dynamics simulations may fail to resolve important features like shock waves, leading to unphysical results.

5. **Improper Boundary Conditions**: Incorrectly specified boundary conditions can introduce spurious reflections or other artifacts that destabilize the simulation.

6. **Nonlinearity and Chaos**: In systems that exhibit chaotic behavior, small differences in initial conditions or numerical parameters can lead to vastly different outcomes, making the simulation highly sensitive and prone to instability.

7. **Parallel Communication**: Large-scale simulations often distribute computational tasks across multiple processors. When results from different processors are combined, each intermediate result may have different round-off

errors due to the non-associative nature of floating-point arithmetic, leading to divergence in the simulation outcomes.

8. **Discreteness Noise or Finite-N Sampling**: In simulations that model continuous systems with discrete particles or elements, statistical fluctuations due to finite sampling can introduce noise that affects the dynamics, especially in gravitational N-body simulations.

9. **Force Softening**: The choice of the softening length ϵ is crucial in N-body simulations to prevent numerical divergences when particles come very close to each other. If ϵ is too large, important small-scale interactions are overly smoothed, leading to a loss of resolution and accuracy. If ϵ is too small, the simulation risks reintroducing numerical instability and noise. Therefore, ϵ must be chosen carefully, often through a convergence study, to balance the trade-off between accuracy and stability.

10. **Truncation of Multipole Expansion in Tree Codes**: While the truncation of the multipole expansion in tree codes introduces errors, these are generally well-controlled and negligible compared to other sources of numerical chaos.

Chaotic behavior is an intrinsic property of certain dynamical systems, characterized by sensitivity to initial conditions. This hallmark feature means that even minimal changes in the starting configuration of such systems can lead to drastically different outcomes over time. While sources of variation such as changes in MPI tasks, truncation errors, random number generators, or numerical precision (as presented in the above list) can seed chaotic behavior, the phenomenon itself arises inherently from the nonlinear and coupled interactions within the system. Even under conditions where external sources of variability are eliminated— by using fixed MPI tasks, maximum numerical precision, and identical RNG seeds, yielding byte-identical simulations, sensitivity to small perturbations in initial conditions persists. This sensitivity underscores the unavoidable chaotic nature of such systems. In astrophysical simulations, this means that even when noise is minimized, slightly varying the initial conditions leads to divergent evolutionary outcomes, reflecting the dynamical system's underlying complexity.
Thus, while these "seeds" may amplify variability, chaos is fundamentally a property of the equations governing these systems, requiring careful consideration when interpreting simulation results. Understanding and quantifying the impact of chaos is essential to establishing reproducibility and defining error bars in computational astrophysics.
The exploration of numerical chaos also intersected with advancements in computer science, particularly in the development of numerical methods and error analysis. Researchers recognized that traditional methods might be insufficient for accurately simulating chaotic systems, leading to the adoption of high-precision arithmetic and the study of stability in numerical algorithms.
Today, the study of numerical chaos is a critical component of computational science, with applications across a wide range of disciplines. From predicting the behavior of complex biological systems to understanding the dynamics of financial markets, the ability to accurately simulate chaotic systems is essential. However,

despite significant progress, challenges remain. Numerical chaos continues to pose a fundamental obstacle to long-term predictions, and the search for more reliable and efficient computational methods is ongoing.
In the context of astrophysics, the chaotic nature of simulations with gravitational systems was first systematically explored in the mid-20th century through numerical simulations of star clusters and galaxies. Early computational work by Aarseth in 1963 on the dynamical evolution of star clusters was among the first to hint at the complex and chaotic behavior inherent in such systems [1].
One of the first numerical studies to directly address chaos in an astrophysical context was conducted by Miller in 1964 [101]. In his work on the dynamics of N-body systems, Miller showed that cumulative contributions from integration errors or rounding-off in N-body computations lead to an exponential growth in the separation of trajectories of two phase points belonging to similar stellar systems. Miller's study was further corroborated by [140].
Around the same time, Hénon studied the orbits of systems with simplified galactic potentials [59]. Hénon observed that when the energy of the system was varied from a low value up to a certain critical level and beyond, the phase-space trajectories changed from predictable, smoothly repeating loops to unpredictable, scattered behavior.

3.1.1 Quantitative Measures of Chaos

The development of Lyapunov exponents in the 1970s provided a quantitative measure of chaos in dynamical systems. Lyapunov exponents (λ) quantify the rate at which nearby trajectories in phase space diverge and are defined by the relation:

$$\lambda = \lim_{t\to\infty} \lim_{\delta x(0)\to 0} \frac{1}{t} \ln \frac{|\delta x(t)|}{|\delta x(0)|}, \tag{3.3}$$

where $\delta x(0)$ and $\delta x(t)$ represent the initial and evolved separations of trajectories, respectively. A positive Lyapunov exponent indicates chaos, as it signifies exponential divergence of nearby trajectories over time [141].
Miller's study involved particle numbers ranging from 4 up-to 32. The development of efficient algorithms for N-body simulations, such as the tree code [8] and the fast multipole method [51], marked significant milestones in computational astrophysics. These advancements allowed for the study of systems with a much larger number of particles. Particularly, the study of dynamical chaos with respect to the characteristic timescale of growth of variations and its dependence on larger N (referred to as "N-dependence") was explored by Gurzadyan and Savvidy [53, 54], Heggie [57], and Kandrup [70, 69, 68].
It was found that the growth of variations from perturbations in initial conditions was exponential and that the growth rate increased with increasing N. However, further studies clarified that although there was a strong dependence of the Lyapunov timescales for smaller N, this dependence diminished for larger N (greater than 30), leading to an e-folding time that was approximately independent of N. The Lyapunov timescale for large N, e.g., a galaxy, was computed to be a fraction of $t_{\rm cr}$, the crossing timescale [50, 58]. The more recent paper by [158] computes

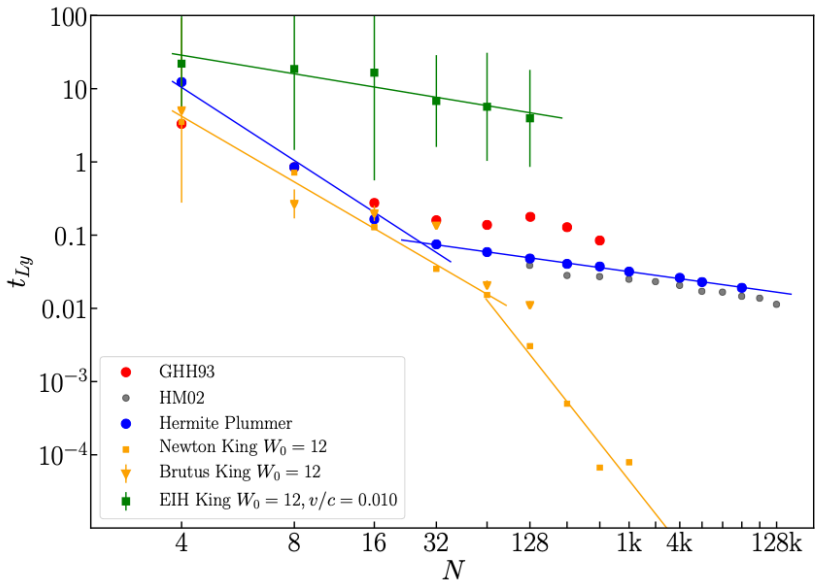

Figure 3.1.1: Lyapunov timescales computed for a Newtonian system case with particles in both in both a Plummer sphere distribution (blue) and King model (yellow), x-axis is $\ln(\ln(N))$. For comparision, the Lyapunov timescales for the King model are also shown using the Einstein-Infeld-Hoffmann (EIH) equations of motion (green) with a velocity ratio $v/c = 0.010$ (The EIH equations incorporate relativistic corrections in gravitational systems with significant interactions but moderate velocities). Credits: [158].

the Lyapunov timescales up to N = 128 000 (see Figure 3.1.1). For more specifics, see also the series of papers by Kandrup and collaborators [67, 72, 71]. Notably, with respect to the softening parameter (ϵ-dependence), it was found that it influences a decrease in the growth rate of variations. For more recent studies of N-dependence, see [14] and [156].
[156] clarified different phases of growth of errors in N-body systems. Before the exponential growth, the timescale of the growth rate of variation seems sensitive to N, being smaller for larger N. As N is further increased toward the collisionless limit, this saturates at spatial scales of $\frac{1}{\sqrt{N}}$ times the system size. Following the exponential growth, which is nearly independent of N, there is a third phase where the growth is multiplicative and progresses differently for phase-space variables and other quantities.
The study of chaos in N-body simulations became particularly significant in the context of planetary dynamics, where the long-term stability of planetary orbits is of great interest. The seminal work by [83], who used numerical methods to explore the chaotic nature of the Solar System, revealed that even the orbits of the planets (mostly the innermost planets) could be chaotic over sufficiently long timescales. For example, the chaotic implications for Mercury, in terms of an increase in its orbital eccentricity, were so significant that it could lead to the ejection of the planet from the Solar System due to an encounter with Venus, within an estimated timescale of less than 3.5 Gyr. These findings have profound implications for our understanding of planetary system evolution and stability [85, 99, 12, 84]. The Lyapunov timescale for the Solar System was found to be in the range of 5 Myr or a bit larger ([5] computed a maximum of 6.8 Myr; see also [39, 13] for more).

3.1.2 The role of chaos in modeling astrophysical processes

Apart from N-body systems, various processes within the astrophysical context have been studied for stochastic influences. Molecular clouds are turbulent and inherently chaotic environments where accretion, the onset of star formation, stellar dynamics, and the total stellar mass formed are influenced by chaotic processes [32, 11, 42, 46, 89]. The more recent work by [62], performed smoothed particle hydrodynamics (SPH) simulations by varying the initial turbulent velocity of otherwise identical molecular clouds and found nearly a 60% variation in star formation efficiency after a period of 5 Myr.
Different authors explored orbital chaos and its impacts on halo stars and stellar streams [49, 94, 114, 115]. The presence of a galactic bar induces chaotic orbits, as well as galactic disk dynamics, which are found to be widely stochastic [45, 129, 128]. [129] tested their simple N-body disk galaxy model to study dynamical evolution of models that form bars and other non-axisymmetric structures and reported stochasticity in terms of the bar amplitude (which could end up varying over a factor of 3 or more) and pattern (see Figure 3.1.2).
The impact of various physics, including the effects of discreteness in stellar feedback on galaxy evolution, was studied in [143] and [142]. They found that for galactic scales, among other physics, the main impacts on the interstellar medium

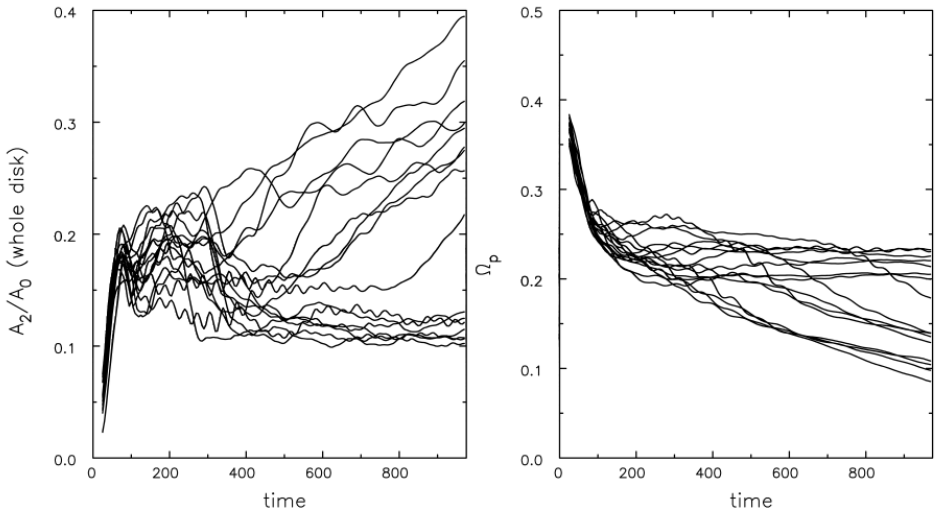

Figure 3.1.2: Variation in galactic bar amplitude and pattern speed with dynamical time in a simple N-body disk galaxy model run (16 runs) with identical particle distribution but random variation included in disc particle coordinates. Credits: [129].

(ISM) and star formation are from stellar feedback, turbulence, and self-gravity. Additionally, they found that the discreteness of stellar feedback was highly effective in regulating star formation compared to a continuous injection of feedback energy. More recently, [77] explored the impact of energy injection and timings (both the interval between star formation and first feedback and the duration of the energy injection) of supernova feedback. They found that the longer the delay of the feedback, the stronger the outflows and energy, and thus better the star formation regulation. Notably, they confirmed that these results held true against pure stochastic variations in their simulations.

Also a notable part is the baselines established to account for purely stochastic variation in discussed by [143, 142]. In [143], five cosmological simulations were made for a Milky Way-like disk galaxy ($M_{\rm halo} = 2 \times 10^{12} M_\odot$), but with small variations introduced in the feedback scheme (they shuffled the particles to which feedback energy and momentum were applied). They found the resulting stochastic variation to be small. By contrast, in [142], the baseline run for purely stochastic variation was made with two runs of identical physics but with differing seeds for the random number generator. The difference in random number seed introduces variation in star formation–i.e., different positions, ages, and also stellar feedback. They found significant variation in their star formation histories (order-of-magnitude differences in star formation rates and outflow rates, both instantaneous values and averaged over 100 Myr). The final stellar mass for their $M_{\rm vir} = 8.0 \times 10^9 M_\odot$ isolated dwarf halo varies by a factor of 2. Note that black

holes are not included in either of the studies.
Some authors [28, 30, 29] examined the role of assembly history on galactic evolution using the EAGLE model [126]. They used a technique called genetic modification [122, 119, 61] 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.

3.2 Chaos in Cosmological Simulations

One of the first examples of estimating chaos in Cosmological simulation is the work of Thiébaut et al. (2008) [146]. The authors investigated the onset of stochasticity and its imprints on larger scales with cosmological simulations within the context of the Λ Cold Dark Matter (LambdaCDM) model. They ran identical simulations (boxes of size $100\,h^{-1}$ Mpc; two sets to test with respect to resolution) with perturbed initial conditions (they added slight variations in initial particle positions and velocities) and quantified the sensitivity of the large-scale structure to these minute differences.
They showed that chaotic behavior arises on small scales where dynamics is highly nonlinear, while it diminishes on larger scales, enabling the identification of a critical transition scale around $\sim 3.5\,h^{-1}$ Mpc (see figure 3.2.3). This scale corresponds to substructures with a mass upper bound of $M_{\rm crit} = 2 \times 10^{13}\,M_{\odot}$, which are sensitive to minute changes in initial conditions. Some integrative quantities such as mass and substructure center, which average over a large number of particles, tend to smooth out differences, whereas other observables such as spin remain more sensitive to the chaotic dynamics of the local environment. More recently, [75] performed test simulations of small volumes ($10, h^{-1}$ Mpc) with perturbed initial conditions and found that stochastic variations manifested on larger scales by the epoch of reionization and affected the 21-cm signal.

3.2.1 Effect of chaos on simulated galaxy

Keller et al. (2019) [76] studied the effects of chaos on simulated galaxies, particularly how stochastic processes in galaxy formation lead to variability in large-scale galaxy properties over long timescales. They conducted a comprehensive experiment involving 128 identical simulations for an isolated galaxy using different codes. The simulations were conducted using GASOLINE2 (a smoothed-particle hydrodynamics code) and RAMSES (an Eulerian adaptive mesh refinement code).

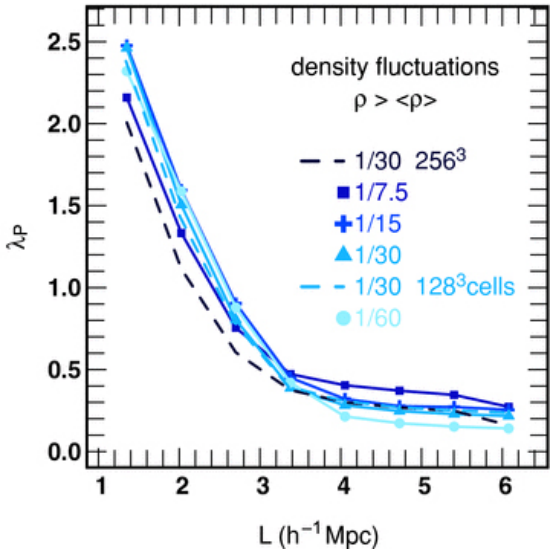

Figure 3.2.3: The figure shows the average Lyapunov exponent (λ_P, of pixel density fluctuations) calculated for different resolution Λ cold dark matter cosmological simulation sets (128^3 and 256^3, measured at scale factor $a = 1$ and 0.4, respectively) produced with GADGET code for different amplitudes of perturbation introduced ($A = \frac{1}{30}, \frac{1}{60}, \dots$) as a function of L, the smoothing length, considering only over-dense regions. Credits: [146].

These cosmological zoom-in simulations used MUGS2 cosmological initial conditions, with gas particle masses of approximately $2.2 \times 10^5, M_\odot$, allowing the formation of a Milky Way-like galaxy with a halo mass of $6.5 \times 10^{11}, M_\odot$. Additionally, the team ran isolated dwarf galaxy simulations to test the effects of chaos with respect to varied gas fractions.

To investigate the sources of chaotic variation, the authors study the evolution of perturbations in their simulations that can be introduced through several mechanisms, namely floating-point round-off errors, random number generators and algorithmic differences. Keller et al. (2019) found that small-scale stochasticity in galaxy formation models can grow into macroscopic differences in properties such as stellar mass and star formation rates. Feedback mechanisms, such as supernova-driven feedback largely affect chaotic behavior; they report a $\sigma_{M_*} \sim 5\%$ for simulations including feedback. Depending on the specifics of the implementation of the feedback, they find that feedback certainly contribute to chaotic fluctuations by amplifying small-scale perturbations in the interstellar medium (ISM) and star formation history. But, they also highlight the complex role of feedback mechanisms, in controlling star formation rates and mitigating chaotic variability(see Figure 3.2.5). Without regulation of star formation, the star formation rates increase dramatically, which diminishes the influence of random fluctuations, or shot noise, in the system (Shot noise can be approximated as $\sigma = 2N^{-1/2}$, where N is the number of stars). In contrast, when gas depletion times are longer, especially in galaxies with lower gas fractions, the chaotic variation rises far above what would be expected from shot noise alone. This occurs because in such environments, where gas is not rapidly exhausted, the chaotic

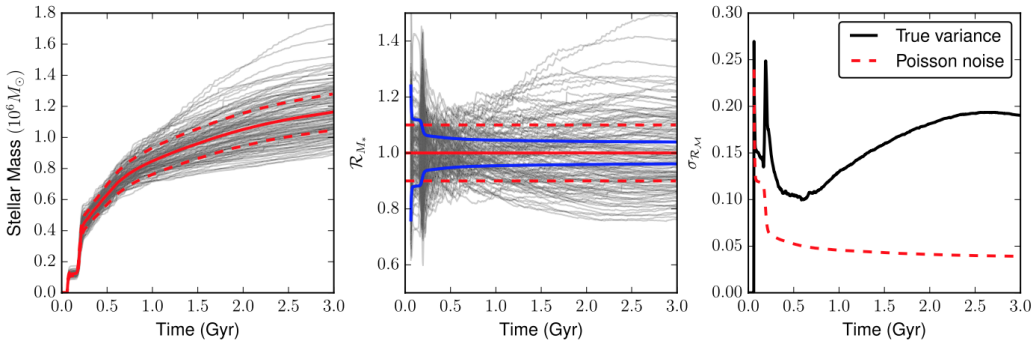

Figure 3.2.4: Keller at al.(2019) simulate merger scenario of two galaxies: for such specific samples they show, in the left panel the stellar masses produced, middle panel shows the deviation in stellar mass and the evolution of the variation is shown in the right panel. These results illustrate that despite feedback mechanism and gas depletion in place to dampen fluctuations, dynamic scenarios such as mergers has a significant impact on run-to-run variation that can persist over Gyr timeperiods. Credits: [76].

behavior is amplified, allowing stochastic processes to exert a stronger influence on the galaxy's evolution. This comparison between regulated and unregulated star formation demonstrates how delicate the balance is between feedback and gas supply in determining galaxy properties over long timescales. Feedback mechanisms like stellar feedback as well as gas exhaustion act as self-regulating processes that moderate chaotic variations.
Chaotic variations due to merger timing significantly affect galaxy properties, which is a notable result of Keller et al. (2019). They demonstrate that differences in merger histories can lead to variations in stellar mass by a factor of 2, with these effects persisting for over a Gyr (see Figure 3.2.4) In effect, Keller et al. (2019) conclude that chaotic variability introduces a significant constraint when comparing galaxy formation simulations. Even with identical initial conditions, divergent outcomes can arise, meaning that differences between simulations may not always reflect differences in physical models. Understanding chaos is thus crucial for interpreting numerical simulations accurately.

3.2.2 Effect of chaos on cosmological volumes

The their paper, Genel et al. (2019) [48] quantify how tiny initial perturbations in cosmological simulations lead to macroscopic differences in individual galaxies. The simulations were performed using the AREPO code, which combines a moving mesh hydrodynamics solver with a Tree-PM gravity solver. AREPO is well-suited for cosmological simulations because it handles complex gas dynamics and gravitational interactions while maintaining high accuracy across a large range of spatial scales. The study utilized simulations with cosmological volumes of 25 to 50 Mpc/h, allowing for the evolution of large-scale structures while capturing galaxy formation processes. The simulations were run at four different resolution levels, spanning a factor of 512 in mass resolution. The high-

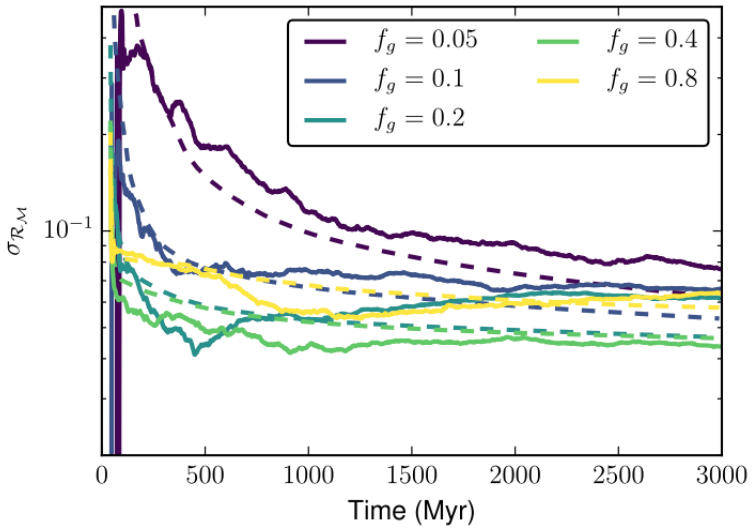

Figure 3.2.5: Keller et al. (2019) variations in stellar mass (computed for dwarf galaxy simulations, represented by solid lines) closely approach by Poisson noise (depicted as dashed lines). Feedback was incorporated in this set of runs, the impact of which is seen in reducing the scatter in galaxy stellar mass with respect to varying gas fractions (f_g). Credits: [76].

est resolution simulation had a particle mass of $1.8 \times 10^5\, M_\odot$ for dark matter particles and $9.2 \times 10^5\, M_\odot$ for gas cells, which is similar to the resolution used in the IllustrisTNG project.
Three main series of simulations were run:

- DM-only series: Pure N-body simulations with only dark matter
- No-feedback series: These included baryons, hydrodynamics, and star formation, but no feedback processes like supernovae or black hole activity.
- TNG series: These are full galaxy formation models with feedback mechanisms, including supernova feedback and black hole-driven winds. This series uses the physics models from the IllustrisTNG project [152], which is crucial in understanding the role of feedback in generating chaos

They run pairs of simulations with nearly identical initial conditions, except the tiny displacements applied to the particle positions. These displacements are on the order of $10^{-14} \times r$, where r is the particle position, resulting in changes of about 10^{-7} pc in physical terms. These tiny perturbations are applied only once at the start of the simulation (around redshift $z = 127$) and are small enough that they do not immediately affect the outcome of the simulation. However, over time, as gravitational and hydrodynamical interactions evolve, these initial differences amplify, leading to the chaotic divergence of galaxy properties.
The growth of differences in the identical runs were tracked over time by measuring the differences in quantities like stellar mass, star formation rates, and

rotational velocities for the clone galaxies. Note that the term clone/shadow signifies multiple realizations of a simulation and within the context of galaxies reflects, a one-to-one matching of individual objects across the identical runs. The authors use a logarithmic difference to quantify these divergences, fitting the distributions to Gaussian models to assess the typical divergence at different times and across different galaxy masses. Differences were computed for a given population of matched galaxies to obtain an averaged estimate of variation between simulations. This is a statistical approach, different from [76], who computed the differences between a single galaxy matched across a large number of identical simulations. Thus the study estimates a Lyapunov-like timescale for the divergence of galaxy properties, though not in a strict dynamical systems sense. The perturbations grow more rapidly during the early stages of the simulation, particularly in the first 2 Gyr, after which they tend to stabilize. This timescale is critical for understanding when chaotic divergence has the largest impact on galaxy evolution. They show that after billions of years, small initial displacements (on the order of 10^{-7} pc) lead to differences in galaxy properties such as stellar mass, circular velocity, and star formation rates. The divergence reaches about 2% to 25% depending on the property considered (For instance, the star formation rate can differ by as much as 25%, while the stellar mass and circular velocity show somewhat smaller but still significant variations of around 5% to 15%)

A crucial part of their methodology was to compare simulations with and without feedback mechanisms. The study finds that when feedback from supermassive black holes and star formation is included (in the TNG series), the chaotic divergence persists even at higher resolutions. Without feedback (in the No-feedback series), the chaos tends to decrease as the resolution improves. This suggests that feedback processes, particularly those involving black holes, play a key role in driving the chaotic behavior.

The study also quantifies how this chaos affects scaling relations. For example, the Tully-Fisher relation, which links galaxy luminosity and circular velocity, shows increased scatter due to chaotic effects (see Figure 3.2.6). The contribution of chaos to this scatter is non-negligible; for certain scaling relations like the star formation rate-mass relation, chaos can account for 15% to 30% of the overall scatter.

Borrow et al. (2023) [22] studied the effects of stochastic processes in galaxy formation using detailed cosmological simulations, focusing on how these processes introduce variability in individual galaxy properties, particularly in star formation and feedback mechanisms. The authors used SWIFT [124], a highly parallelized cosmological simulation code, to conduct zoom-in simulations of individual galaxies drawn from the larger EAGLE-25 cosmological initial conditions. These simulations provided a detailed look at the variability introduced by stochastic processes in galaxy formation models. For this study, all simulations were executed on a single node with 28 threads, taking full advantage of task-based parallelism without requiring MPI components. The SPHENIX SPH scheme was used for hydrodynamics, and the Fast Multiple Method (FMM) solver was employed to handle gravitational interactions. These methods allowed the authors

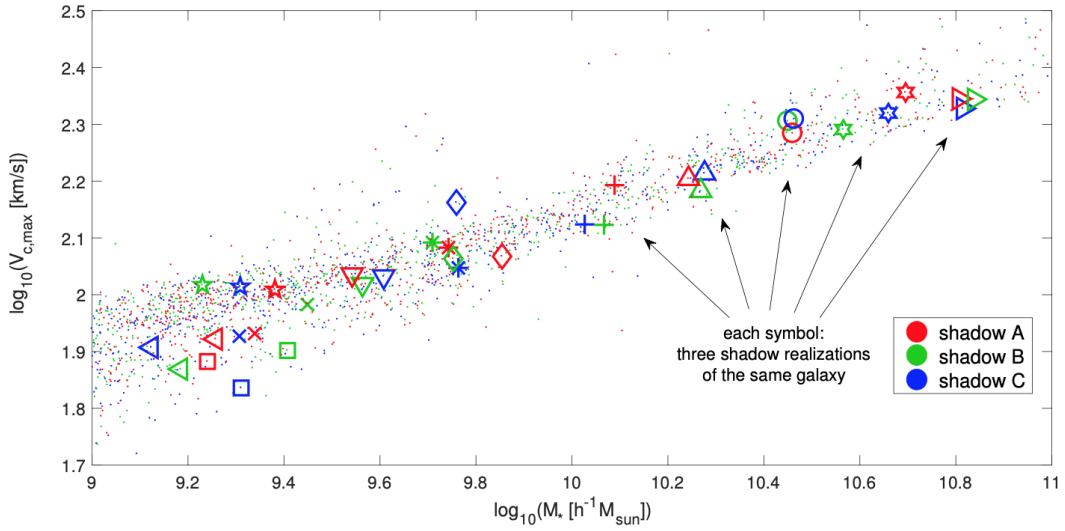

Figure 3.2.6: The $z = 0$ Tully-Fisher relation, defined here as $V_{c,max} - M_{\star}$, is shown from the TNG model simulations by Genel et al. (2019) at resolution level $\epsilon = 1$. Each point represent galaxies evolved in identical runs but with perturbed initial conditions at $z = 5$, the symbols represent matched galaxies across the simulations. The result depicts that the variation in each individual galaxy is significant to overall scatter in the relation. TNG model refers to the IllustrisTNG (The Next Generation) model, a suite of large-scale cosmological simulations aimed at studying galaxy formation and evolution. Credits: [48].

to simulate galaxy formation efficiently, with a high degree of accuracy for both gas dynamics and N-body gravitational forces.

The study focused on 16 identical 'clone' simulations, where the same initial conditions were used (from the EAGLE-25 dataset), but different random seeds were applied to the sub-grid models. This allowed the team to explore the impact of stochastic variations on individual galaxy properties, such as stellar mass, feedback efficiency, and other key observables.

Borrow et al. (2023) found that stochastic variations introduced by star formation, black hole accretion, and feedback mechanisms led to significant differences in the properties of individual galaxies, even in simulations where galaxies were resolved with tens of thousands of particles. These random processes caused the properties of individual galaxies, such as stellar mass, to vary by up to 25% across different clone simulations.

The authors also noted that while scaling relations averaged over many galaxies remain robust against stochasticity, the properties of individual galaxies —particularly during bursty events like black hole feedback or mergers—show significant variability. For instance, in the stellar mass-black hole mass relation, stochastic variations can cause changes of up to 1 dex during the rapid growth phases of black holes, significantly affecting the evolutionary pathways of individual galaxies. Figure 3.2.7 presents four key galaxy scaling relations, including the specific star formation rate (sSFR) and the stellar mass-black hole mass relation, showing the scatter between the different clone simulations. This figure clearly illustrates

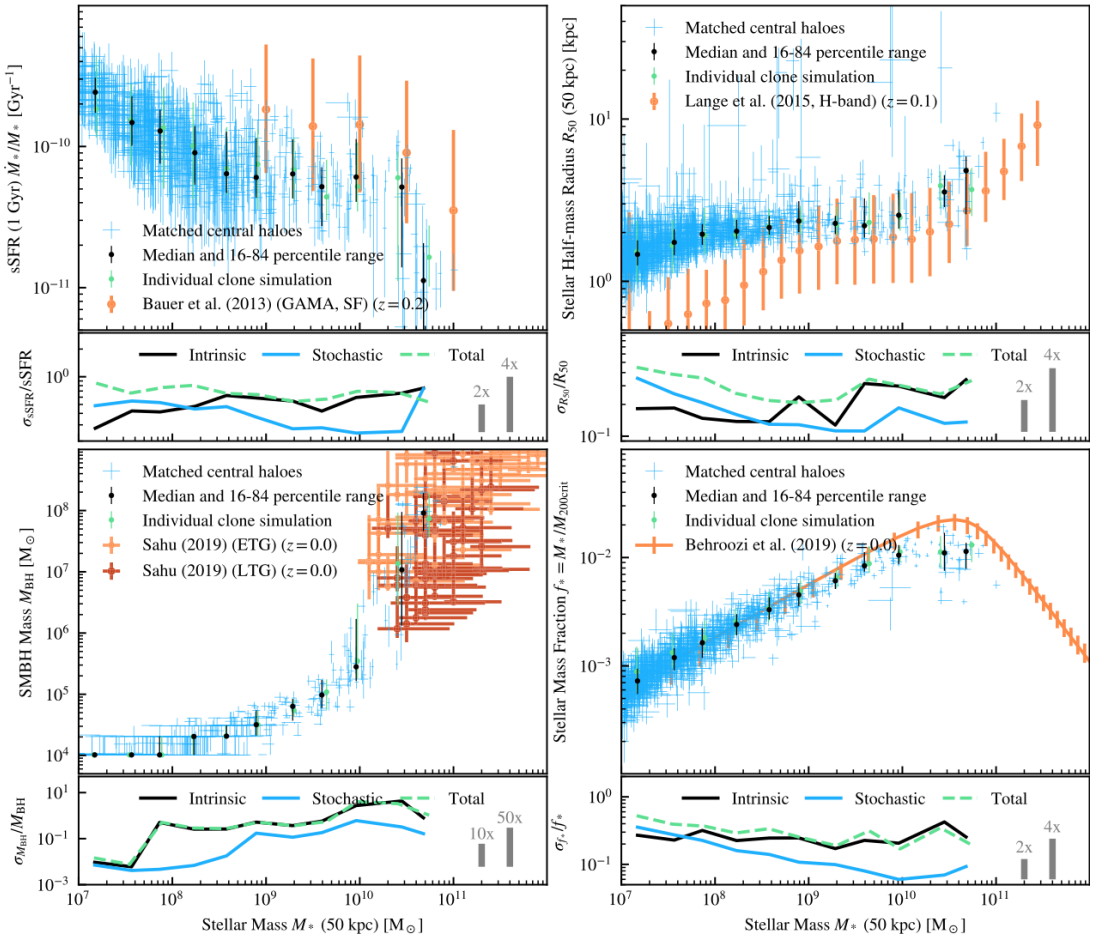

Figure 3.2.7: Figure shows, at z=0, specific star formation rate (sSFR, averaged value), stellar half-mass radius within 50kpc, blackhole mass, and stellar mass fraction plotted as function of galaxy stellar mass, along with observational data for comparison. Blue crosses represent indicate the scatter (16th to 84th percentile range for each clone galaxy) across all the multiple runs centered on the median values, black points are the medians and 16–84 percentile spread of these median values, binned by stellar mass. Green points reflect variation of a single clone. For each of these, the lower panels show variation in each of three components and are presented in same color. See [22] for more details.

how stochastic processes affect variability in galaxy properties, providing a visual representation of the core findings of the study.

Chapter 4

Study of chaotic effects in galaxy cluster simulations

The few studies (as listed in the introductory chapter 3) made with cosmological simulations reveal the significant role of chaos in galaxy formation and evolution. The divergence of galaxy properties due to minute initial perturbations suggests that chaos is a key driver of variability in galaxy scaling relations. These results have important implications for both theoretical models of galaxy formation and the interpretation of cosmological simulations, indicating that chaotic processes must be taken into account to fully understand the nature of galaxy evolution. The presence of chaos also poses challenges for the precision and reproducibility of cosmological simulations. Simulation codes are inherently chaotic in nature. The stochasticity can arise from sources such as floating-point round-off errors, random number generators, and slight algorithmic differences in the code.

Though the role of chaos in cosmological simulations is crucial, it is often an underappreciated factor. Evaluation of run-to-run variations in simulations are rarely performed. Without establishing a clear understanding of the inherent "error bars", the result thus obtained, often attributed to a physical phenomenon- might, in fact, stem from underlying chaotic effects.

Thus the primary goal of the my work is to follow the works of [76], [48] and [22], thereby establishing a baseline measure of chaotic variation that can be expected in simulations run with OpenGadget3 code. By quantifying the variation due to chaotic effects with identical runs, I aim to provide a benchmark to assess what magnitude of variation one can expect in simulations run with OpenGadget3 code.

The next pressing question is the interplay between chaos and feedback, which has led to conflicting findings in recent literature, as already summarized earlier in this chapter. Keller et al. (2019) find that feedback helps to stabilize the run-to-run variation. In contrast, Genel et al. (2019) find that feedback in general can amplify chaotic behavior. These contrasting conclusions highlight the complexity of subgrid models and their influence on the simulations results. This topic forms the second of the objectives for my thesis. The focus of this effort would be to systematically disentangle the effects of various subgrid models with respect to feedback and clarify if indeed chaotic processes are modulated or exacerbated by

feedback mechanisms.
Overall, this research is a foundational step that is intended to provide insights into the numerical noise and chaotic influences that can be expected with OpenGadget code and also contribute to the understanding of chaotic behavior in cosmological simulations in general.

4.1 Methodology

In this section, I outline all the steps of my study to obtain the results; starting with the initial conditions for the simulations, the simulation set-up, parameters tested are all described. Next, I detail how data required for quantification of variation is extracted, and finally the analytical methods used to interpret the results.

4.1.1 Simulation details

Initial conditions

The Zoom-in initial conditions (see Sect. 2.4.2) used in this project for all simulations are the ones described by Bonafede et al. (2011) [20] .
They are the so-called 'Dianoga-set' of simulations (see [117], [118], [107], [19], [18], [116], [9] and [10]) that were generated from a parent low-resolution dark matter only simulation of comoving cubic volume 1 (Gpc h^{-1})^3 executed with GADGET-2 code and adopting a ΛCDM cosmological model (see table 4.1.1). 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 FoF (see 2.4.7) 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 [149] and [20]). Each 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, blackholes 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.
In this set, the ICs of one galaxy cluster (termed D1 or g0016649) of mass M_{vir}= 1.6e15 $M_\odot$ is chosen for all the simulations run within the context of this project.
ΛCDM cosmology is used with the following parameters: Ω_M=0.24, Ω_B=0.04, h=0.72 and σ_8=0.8 ([78]; WMAP-7).

Set-up

All simulations in this study were conducted using the OpenGadget3 code, aligning with the primary goal of investigating chaotic effects specifically in simulations executed with this code. For the purpose of my study, an older version of the code (here after named as "Trieste version") was employed. This version of the code was chosen primarily for practical reasons. While the latest release includes useful enhancements, it faced occasional stability issues that limited its applicability for my study. Additionally, it was optimized for specific initial conditions, which provided a focused configuration but did not fully support the broader parameter exploration required for this study. The selected version, though an earlier release, offered reliable performance and flexibility, enabling extensive testing across multiple configurations essential for my analysis. The code is introduced in Section 2.4.1.

In studies of chaos, especially within computational astrophysics, it is critical to tightly constrain various numerical factors because even subtle variations in code setup, compiler optimizations, libraries, or hardware architecture can affect the results. Different versions of a simulation code may contain slight changes in algorithms, bug fixes, or optimizations that can alter the order of operations or the handling of numerical precision. These differences are crucial in chaotic systems, where even small changes in floating-point arithmetic can produce divergent results.

Numerical libraries, such as those handling matrix operations or Fourier transforms, may implement different methods or optimizations that impact computational precision, or may be optimized differently from a version to another, for this reason I fixed the libraries and libraries version that I employed. In particular, the code has been compiled always with the following set of libraries: FFTW 3.3.10, GSL 2.7.1, HDF5 1.12.2 and CFITSIO 4.0.0, for each set of tests.

Deliberate choice has been made with respect to compiler and compiler flags. Different versions of the compiler may introduce differences in the way the executable is created. Moreover, compiler optimizations can rearrange calculations to improve performance, sometimes at the expense of exact floating-point reproducibility. Flags like -ffast-math or architecture-specific optimizations (e.g., -march=native) may alter floating-point operations, introducing variability in simulations. Thus the code has been compiled using GCC 11.2.0 and the compiler flags has always been kept the same.

Parallelized simulations, especially those using MPI, may encounter variations in operation ordering based on load balancing and processor availability. The way reduction is treated by the MPI library may vary according to the library developer or implementation, for this reason, I adopted only one version of MPI: OpenMPI 4.1.3.

Differences in CPU architectures (such as Intel vs. AMD) can lead to slight differences in how floating-point operations are handled, due to varying precision levels, rounding behavior, or parallel processing schemes. To avoid architecture related issues the simulations were run on Hotcat cluster (refer [144] and [17]). The computing nodes are composed of Intel(R) Xeon(R) CPU E5-2697v4 with frequency of 2.30GHz. I ran the simulations on 4 computing nodes, each node

comprising of 36 cpus, and using 1 MPI task per CORE.

Table 4.1.1: Simulation set-up

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

To begin, I performed galaxy cluster simulations at the lowest resolution level (Dianoga 1x).
Each dark matter particle had a mass of $8.43 \times 10^8\, M_\odot\, h^{-1}$, and the initial gas particle mass was $1.56 \times 10^8\, M_\odot\, h^{-1}$. Gravitational forces were softened at specific scales for different particle types: 5.62 kpc h^{-1} for gas and dark matter particles, and 3.0 kpc h^{-1} for star and black hole particles. Additionally, for dark matter particles at redshifts $z > 2$, the comoving softening length was set to 16.87 kpc h^{-1}.
To establish a baseline, I ran a set of four identical simulations for the give set. These simulations are my *Fiducial* simulation set and they include hydrodynamics and a comprehensive set of physical processes, such as cooling, star formation, and associated feedback mechanisms like supernova feedback and stellar winds (see Section 2.4.1 for a detailed description of the different processes and their implementation). The first set of tests are made without blackholes. Thus, black holes were intentionally excluded by increasing the critical halo mass required for black hole seeding to an unphysically high value in the parameter file. This ensured that no black holes formed while keeping the executable code unchanged, which is essential for minimizing differences between simulation runs. This strategy allows for a fair comparison between simulations without black holes and those that will include black holes in future studies.
For the *Fiducial Model*, the velocity of the stellar winds is fixed to a value of v_w = 350 km s^{-1} in the configuration (see 2.4.5); this value corresponds to a 50% of the supernovae energy being converted into the kinetic energy carried by the wind.
To study the effect of supernova feedback on mitigating or enhancing numerical chaos, I conducted feedback tests by varying the stellar feedback parameters from the Fiducial model. Specifically, I adjusted the stellar wind velocity to create low, high, and zero feedback scenarios:

- LOW feedback: Approximately 0.57× the fiducial wind velocity.
- HIGH feedback: Approximately 1.42× the fiducial wind velocity.

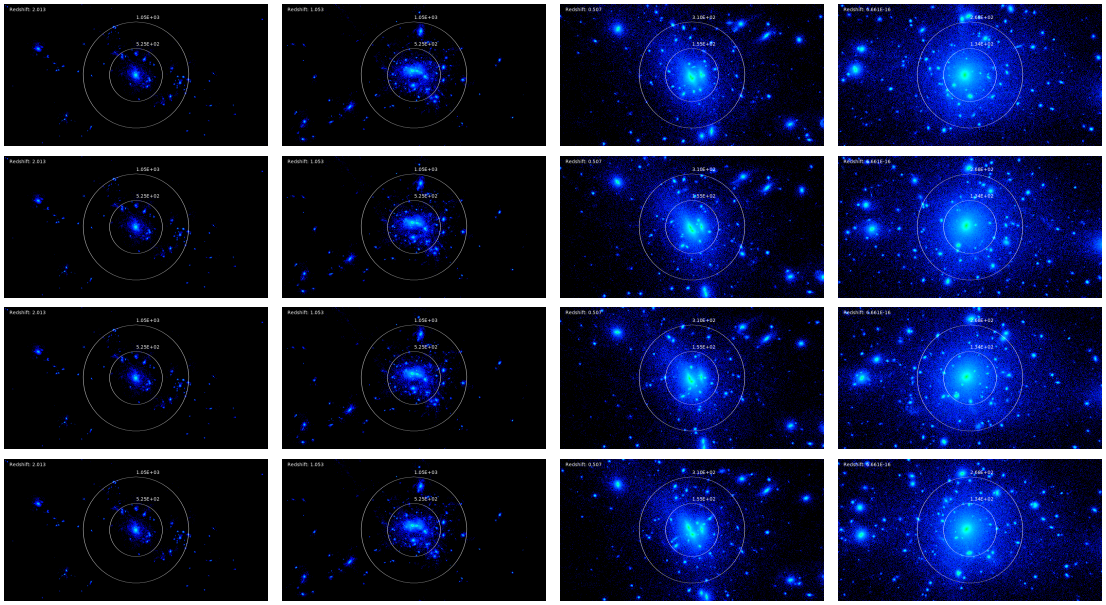

Figure 4.1.1: Visual representation of the Dianoga simulation for the Fiducial model. The figure comprises 16 plots arranged in a grid, where each row represents one realization of the same galaxy cluster simulated with identical initial conditions, code, and parameter settings as discussed in Section 4.1.1. From left to right, the columns correspond to snapshots at redshifts $z = 2$, $z = 1$, $z = 0.5$ and $z = 0$. The differences in substructures between the realizations at each redshift highlight the impact of variability in identical runs.

- ZERO feedback: Galactic wind velocity set to zero, effectively eliminating galactic winds.

While modifying the wind velocity requires recompiling the executable, all other aspects of these variants remain identical to the fiducial model. These parameter choices are consistent with those used in other simulations campaigns done with GADGET-3 and OpenGADGET–for example, the HIGH wind velocity setting aligns with that employed in [117]).
Figure 4.1.1 is a visual representation of the fiducial set of simulations for a single galaxy cluster evolving in the identical runs from z = 2, z = 1, z = 0.5 and z = 0. At the outset, the overall structure seems to align across the images, but variations are apparent in the substructures particularly at lower redshifts. Figure 4.1.2 is same as that of Figure 4.1.1, except for visual clarity the cluster is only shown at z=0. Here we can clearly see, the main object in the inner circle displays evident differences. Additional substructures, encircled in red, also show notable variation.
Figure 4.1.3 is same as Figure 4.1.1, but representation of the matched galaxy cluster in each of the feedback models: Fiducial, HIGH, LOW and ZERO (from left column to right) only at z = 0. Each cluster evolves differently considering the respective feedback; for lower stellar feedback, higher stellar mass galaxies (effect is visually apparent in satellite galaxies as well) are formed which can be seen in 4.1.3. In the next section, I will detail the methodology to quantify this

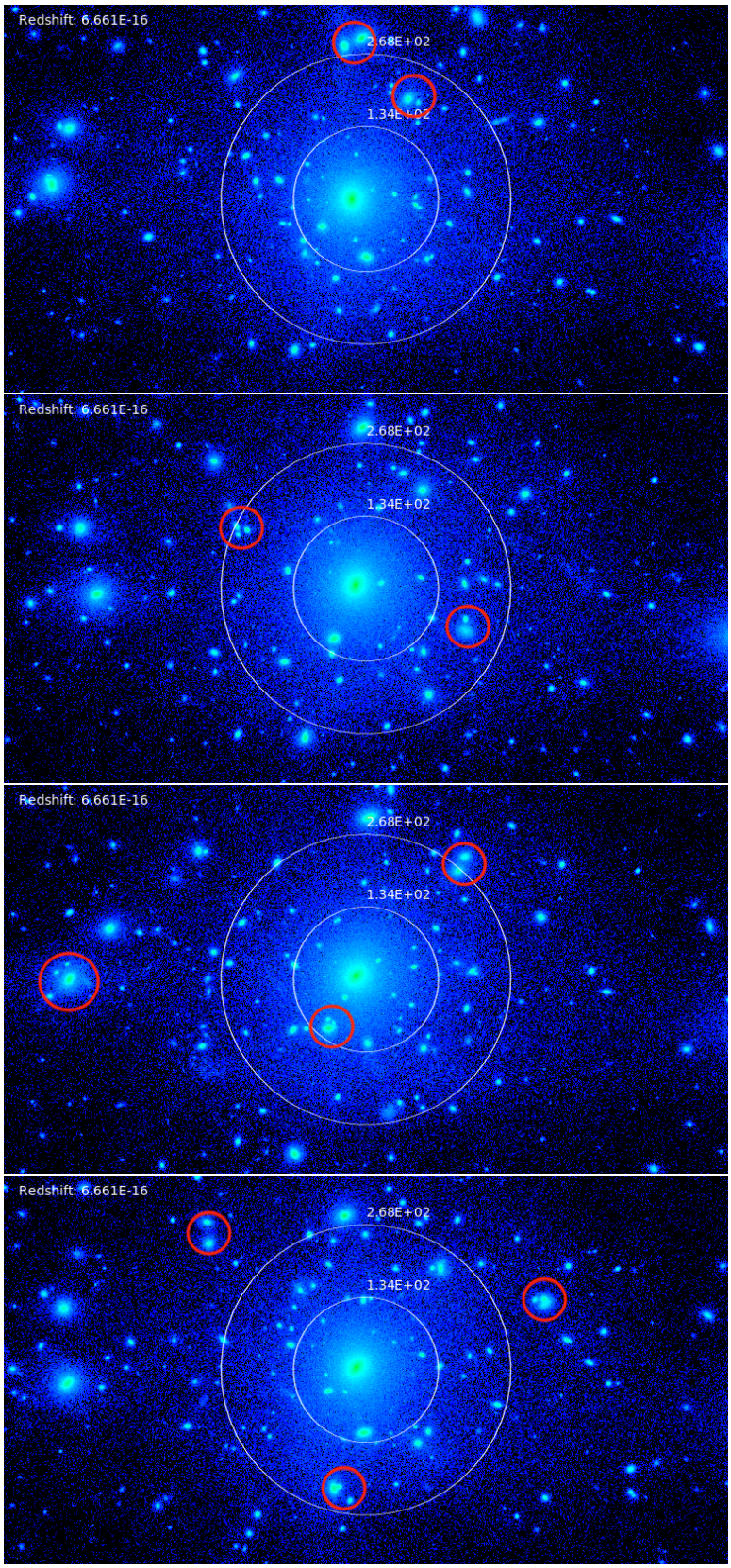

Figure 4.1.2: Same as Figure 4.1.1 but only showing the cluster at $z = 0$. Substructure differences encircled in red.

Type	SFB -v_w(km s^{-1})
FID	350
LOW	200
HIGH	500
ZERO	0

Table 4.1.2: Feedback test parameter

variation for each case.

4.1.2 Methods for matching clone BCGs

For the purpose of my study, the chaotic effects is quantified with respect to changes in galactic properties rather than calculating individual particle trajectory divergence through Lyapunov exponents. By examining how these macroscopic galaxy attributes diverge between simulations with identical initial conditions, the practical impact of chaos and its particular relevance in observable phenomena in cosmology can be captured. Thus the results of this work, i.e the overall variation assessed for the identical runs are in terms of the variation in the galactic properties matched across the identical runs. Galaxies between simulation runs can be matched by tracking particle IDs which are unique and bound to a given object, allowing direct correspondence between the same object of interest across different runs. This technique identifies "clones" or "shadow galaxies" since each match represents the same physical galaxy, but with potentially varied properties due to the simulation's variability.
The most massive and luminous galaxy of each galaxy cluster, depending on the halo mass, can be considered as the BCG: brightest cluster galaxy. The BCG is typically located near the center of the cluster's gravitational potential well. Each galaxy cluster in the simulation has one BCG, but given that multiple galaxy clusters exist within the simulated volume, there are multiple BCGs present across the dataset. In simulations, satellite galaxies typically exhibit higher variation compared to BCGs. This is because satellites are less massive and more susceptible to gravitational interactions, experience increased variability from tidal forces and dynamical friction within the host halo (the effect is visually apparent in Figures 4.1.1 and 4.1.3). Additionally, considering that the simulations are relatively of lower-resolution, BCGs are better resolved, allowing for somewhat reduced sensitivity to simulation noise and ensuring more stable tracking across runs, even in low-resolution simulations. I take a more conservative approach by limiting my analysis solely on BCGs, which are expected to exhibit less scatter than satellites.
The identification of BCGs is achieved through post-processing of simulation outputs using the Friends-of-Friends (FOF) and Subfind algorithms (see Section 2.4.7). These algorithms produce catalogs of halos and their associated galaxies at various redshifts. Thus, starting from one identical run as reference (determination of one as reference has no impact on the results; this has been tested by choosing different clones as reference) a match for the galaxies (only

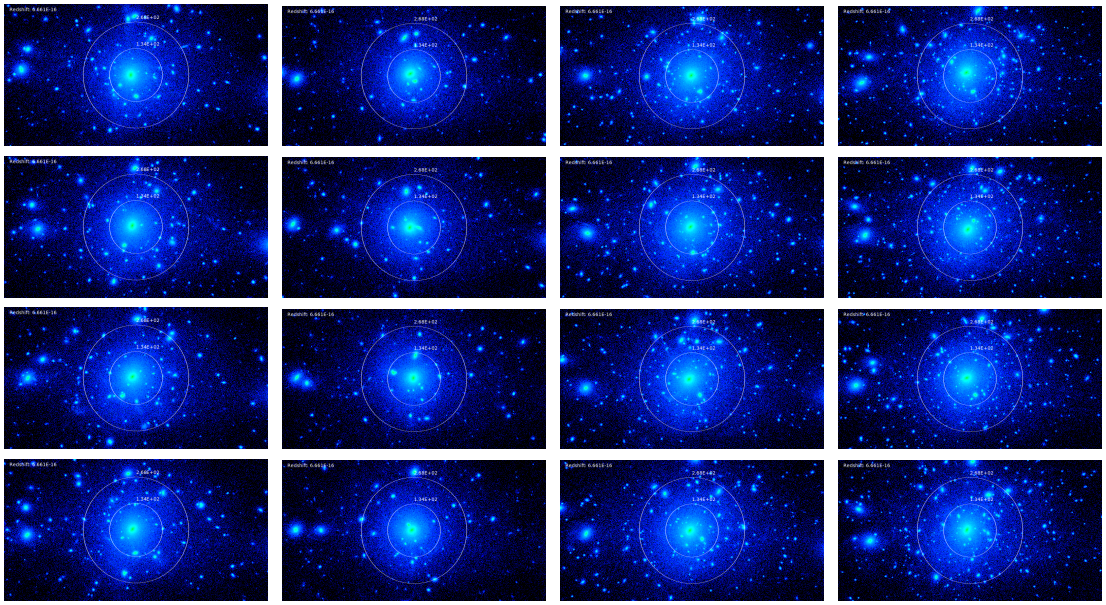

Figure 4.1.3: Visual representation of the Dianoga simulation at $z = 0$. Each row represents one realization of the same galaxy cluster simulated with identical initial conditions, code, and parameter settings as discussed in Section 4.1.1. From left to right, the columns correspond to different feedback models: Fiducial model, HIGH, LOW and ZERO.

BCGs; choosing the most massive central galaxy within the parent halo in the catalog) is found in other identical runs. To start with, I kept an upper limit of only comparing objects with total mass larger than $10^{11} M_{\odot} h^{-1}$, which must result in reasonable comparison despite the shot noise. A match for a given BCG is obtained by firstly, obtaining a match for the parent halo. A halo is considered a clone if it matches ($> 60\%$ of the same particle ids as well as match more than 50% of mass. Further, a similar comparison is made for the central galaxy in each of the matched-halos to obtain the clone BCG. To read gadget snapshot and Subfind files I used the following libraries: g3read and g3matcha (refer the library here: https://github.com/aragagnin/g3read)

While comprehensive matching methods such as [3], [123], [92], [23], [48] and [22] consider complex dynamics such as merger scenarios, this adapted approach provides a practical balance between accuracy and computational efficiency. Considering that few of the galactic properties, which are not available in the Subfind catalog, such as cold gas mass, star formation rate, these needed to be read from snapshot files, which makes for a computationally expensive post-processing for a large catalog of galaxies in the D1 cluster. $95\% - 98\%$ of BCGs are matched with the adapted technique; a higher refinement in 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.

4.1.3 Methods of statistical analysis

At the end of post-processing the identical runs discussed in Section 4.1.2, I obtained a catalog of matched clone BCGs with all the galaxy properties. The main objective of this study is to, then, quantify the variation between the properties of the galaxies from this catalog. For this, two distinct methods are employed, each computing the differences which are estimated basically with standard deviation for each dataset for given time. In this section, I will elaborate on these two methodologies.

Method 1: Computation of variation with Mixed linear model

The variation in the galaxy properties are affected not just by the run-to-run variation but also noise in a given simulation. While run-to-run variation reflects systematic differences between simulations that would otherwise be identical, noise is more unpredictable and random, originating from both inherent stochastic processes and computational limitations. I dedicated some effort in this direction with Method 1, with an intention of gaining some understanding as to how much of the total variation in my simulations are purely noise and how much can be attributed to chaotic effects.
Thus, the first method (hereafter referred to as **Method 1**), is a statistical approach, specifically aimed to analyze these components of variation, using: Mixed Linear Model (MLM ; refer [47]). MLM is a class of statistical models that combine both fixed and random effects to analyze data with complex dependencies. These models are particularly useful when data are hierarchical or involve repeated measures, where observations are nested within groups (accounting for the hierarchical structure of the data; where each BCG is common across different simulations i.e., repeated measures, wherein a common subject is measured multiple times under different conditions). In an MLM, fixed effects capture the systematic part of the variability in the data, which is typically related to observed factors or experimental conditions. Random effects, on the other hand, account for the unexplained variability or noise that arises from unmeasured factors, individual differences, or random fluctuations. The residual error is the noise in a given simulation and is assumed to be normally distributed.
As a first step, a straightforward approach is taken with only this simple target: separating the noise or residual error or within-simulation variation from that of random effects which for current analysis captures the run-to-run variation (or between simulation variation) which is related to the chaotic behavior of system. Further enhancements to account for other dependencies in the relation are left to future work.
The model is thus composed of the two components of variation and can be expressed as

$$y_{e,i} = u_e + v_{e,i} \tag{4.1}$$

where:

- $y_{e,i}$ represents a given property for matched galaxy i in each simulation run e.

- $u_e \sim N(0, \sigma_u^2)$ is the random effect capturing the variation between simulation runs. This term represents the differences in the overall mean galaxy property between simulations due to different realizations of chaotic processes.

- $v_{e,i} \sim N(0, \sigma_v^2)$ is the residual error that capture the variation within each simulation run. This represents the stochastic noise, the measurement variability within each run.

Data Preparation: To implement the statistical analysis it is necessary to extract a proper data set from the simulation output. The dataset consists of a catalog of matched galaxies from four identical simulations for a given fixed simulation set-up. To ensure consistent measurements across the simulations, the galaxy properties are normalized using the within-subject (i.e., matched-BCG) mean, thus accounting for differences in scaling between galaxies while preserving relative differences between the runs. Outliers were filtered out by considering only data within 3 standard deviation of the mean.
The MLM is fit using Restricted Maximum Likelihood Estimation (REML), where first, the fixed effects (such as group means) are fit, and then variance components are estimated based on the residual data left after accounting for fixed effects. The total variation in a given galaxy property is partitioned into the run-to-run variation and residual noise: both (σ_u^2 and σ_v^2) are estimated at the end. These components of variation is then used to assess the relative contributions of systematic differences in identical simulations and random noise to the total variability in galaxy properties. Time evolution was not incorporated into the analysis for current work; instead, the model is fitted independently for each unique timestep.
The analysis was performed always with a sample size (the term sample size refers to the number of observations included in a given study) greater than 100 matched galaxies at each given redshift to ensure statistical robustness : For each galaxy property at each redshift, the analysis involves explicitly examining the distribution of galaxy properties, residuals, and pairwise differences (described next). This provides a clear visual representation of the data and allows for the identification of trends and any deviations.

Method 2: Pairwise Differences for Estimating Total Variation

In addition to Method 1, an alternative method was employed to validate the results (here after referred as **Method 2**). This is following the methods described in [48] who also compute the variation for a population of galaxies. The data was prepared as described in 4.1.3, except, for cases where analysis of variation is specific to ranges of data, such as a particular mass bin, the analysis is restricted to only the data within that respective bin. For each galaxy property, pairwise differences was computed between pairs from the matched BCGs from the four repeated simulations. By computing the differences between all possible pairs of simulations, I obtained a distribution of differences for each galaxy property for given time. This distribution of differences allows one to quantify how much

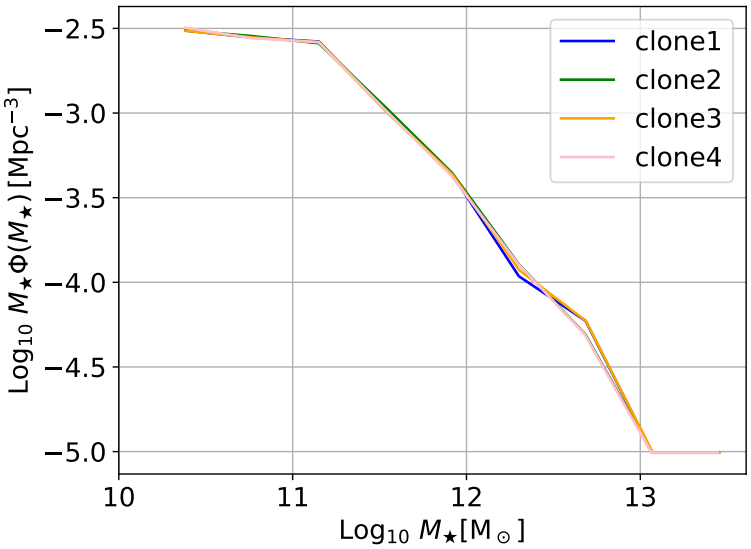

Figure 4.2.4: Galaxy stellar mass function at z≈0 for fiducial simulations.

galaxy properties fluctuate between simulations. The total variation then, is simply the standard deviation of the distribution, corrected, by a small factor (computed to be $\approx \sqrt{2}$) to reflect the total variation in real data rather than a variation in differences.
Since, with pairwise comparisons, the number of data points increases, for example, with four repeated experiments, one can generate six unique pairwise comparisons for each matched galaxy. This increase in sample size improves the robustness of the variability estimates. By comparing the total variation from pairwise differences with the variance components obtained from mixed modeling, I could cross-check the consistency of results. And also Method 2 serves as a relatively, more standard representation of variation to the reader.
Computation of variation is thus made with these two methods for galaxy properties: Dark matter mass M_{dm} and Stellar mass M_*, which are available from Subfind catalog and also some additional values such as cold gas mass (also M_*) are extracted from an aperture of 50pkpc centered on the matched BCG from the snapshots.

4.2 Results

In this section, I present the results of the simulations using the specific tests and analysis methods detailed earlier in this chapter. I begin by quantifying the expected variations in the Fiducial model, and then examine how these variations are influenced by stellar feedback.

4.2.1 Ensemble averaged properties in the simulation volume

Galaxy stellar mass function (GSMF) for fiducial simulation (see table 4.1.2) set is shown in figure 4.2.4. The GSMF is constructed for all galaxies identified by

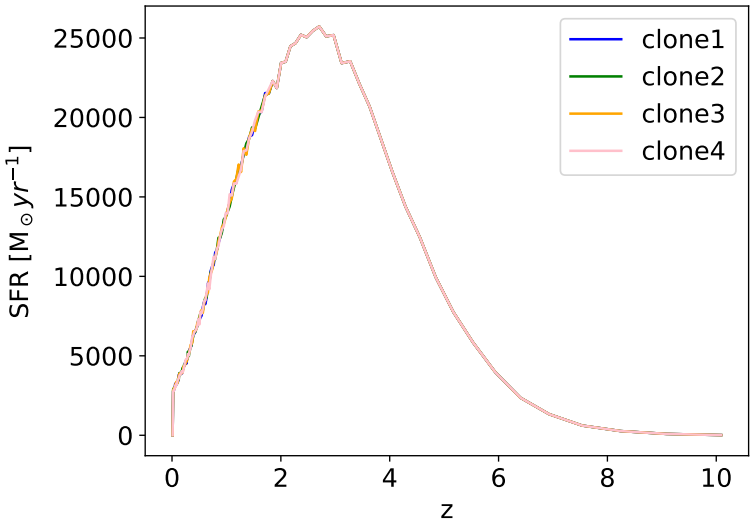

Figure 4.2.5: Star formation history for fiducial simulations.

Subfind at z=0, considering all star particles within 50pkpc of a given galaxy. Figure 4.2.5 is 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, considering all star particles within a sphere of radius 100Mpc at z=0.

These results —including galaxy stellar mass and star formation rate —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 clones, across a large population and on large scales, demonstrate that the ensemble averages of galaxy properties are reproduced by the simulations despite underlying chaotic dynamics. 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.

4.2.2 Variation for fiducial simulations without blackhole

Figures 4.2.7 and 4.2.8 illustrate the variations in dark matter mass and stellar mass, respectively, between matched BCGs in the identical simulations. These results pertain to the fiducial simulation set run without black holes. Note that the stellar mass presented is measured within a 50 physical kpc (pkpc) aperture around the BCG center, rather than using the stellar mass reported in the Subfind catalog. However, for the given set of simulations, the differences between these two measurements are not significant.

In both figures, the variations are quantified using the two methods: pairwise differences (**Method 2**, labeled "pairdiff") and regression analysis (**Method 1**, starting with the label "total"), as described in Section 4.1.3. The total standard deviation is calculated by summing both the residual error (i.e., variation within

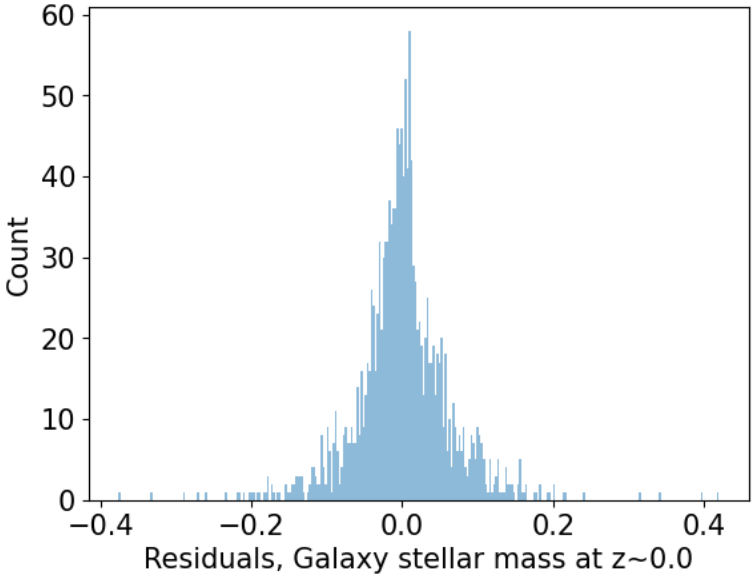

Figure 4.2.6: Residuals for galaxy stellar matter mass for fiducial simulations at z=0

a single simulation) and the run-to-run variation obtained using Method 1. The run-to-run variation is also plotted in the figures, along with a rough estimate of the shot noise.

In these simulations, discrete particles are used to represent continuous fields like dark matter and stars, which introduces shot noise. For a given BCG property, shot noise represents the statistical fluctuations due to the finite number of particles composing that property, leading to uncertainties in its value. The larger the number of particles per galaxy, the lower the shot noise; this is particularly relevant for smaller galaxies with fewer particles, where shot noise can become a significant source of error. An approximate lower limit for shot noise in measurements of galaxy dark matter mass and stellar mass is given by $\approx 1/\sqrt{N}$, where N is the number of stellar or dark matter particles in the galaxy.

The data evolution over time is plotted with respect to lookback time in gigayears (Gyr). Figure 4.2.6 shows that the distribution of errors for stellar mass at z = 0 is normally distributed.

From Figures 4.2.7 and 4.2.8, we observe that the variations computed using two separate methods —pairwise differences and total variation from Method 1 —are consistent, although the data exhibit a scattered rather than smooth evolution. The variation grows from around 9 Gyrs and stays within a range of $\sigma_{M_{dm}}$=0.03-0.06 and σ_{M_*}=0.04-0.08 with some scatter. From each of the respective plots we can see that the total variation grows to and closely approaches that of the lower limit set from poission noise alone. Thus shot noise seems to be the most dominant component for simulations for the respective low resolution runs. This is more evident in $\sigma_{M_{dm}}$ where as for σ_{M_*} the total variation a factor of 1-5 higher than that of the shot noise estimate. See figure 4.2.9 for the residual noise and figure 4.2.10 for run-to-run variation, comparing $\sigma_{M_{dm}}$ and σ_{M_*}. σ_{M_*} is a factor of 1-4 (even higher at earlier times) higher than $\sigma_{M_{dm}}$.

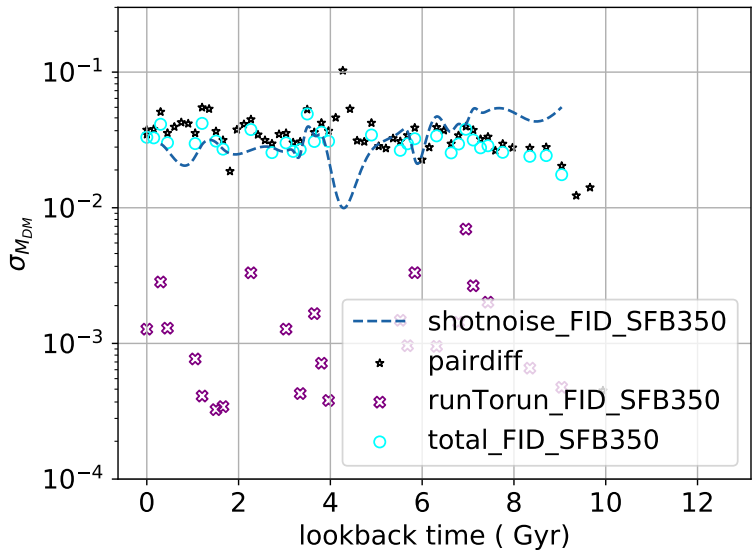

Figure 4.2.7: Variation in galaxy dark matter mass for fiducial simulations. Variation as computed from Method 1 (total and run-to-run) and Method 2 (pairdiff) presented along with an approximate, lower limit estimate of shot noise (dashed line).

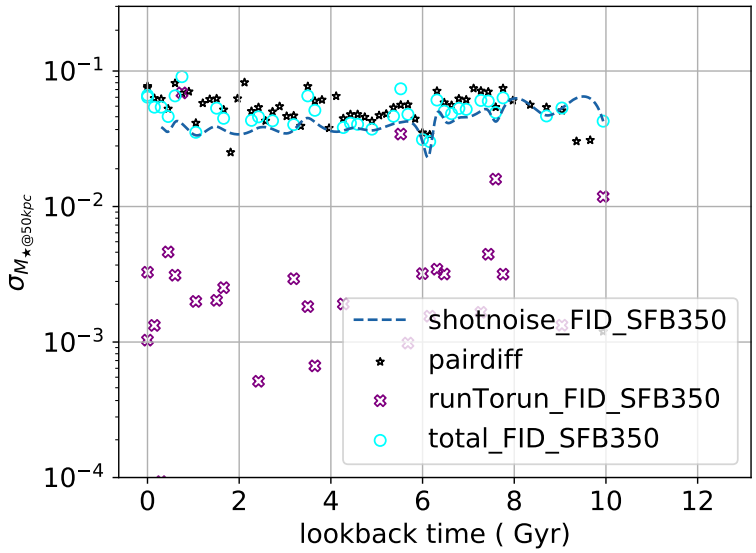

Figure 4.2.8: Same as 4.2.7, but for variation in galaxy stellar matter mass, stellar mass comprised of all stellar particles within an aperture of 50pkpc from BCG center.

This discrepancy can be attributed to two main factors related to increased noise within the simulation:

1. Shot Noise Dependence on Particle Number: Shot noise depends on the number of particles and their spatial density. Dark matter particles are more numerous and densely packed compared to stellar particles, resulting in lower shot noise for dark matter mass measurements.

2. Influence of Non-Linear Processes: Star formation peaks before 10 Gyr, but the OpenGADGET code truncates small-scale variations, leading to byte-identical snapshots. Around 10 Gyr, it is likely that the variations build up beyond this truncation threshold, marking the onset of measurable variation. The growth in variation, particularly in case of stellar mass reflects cumulative effects of highly non-linear and localized processes such as star formation, cooling, gas dynamics, and in particular, feedback mechanisms (as we will see in the succeeding section with feedback related tests). These processes introduce additional noise into the simulation (see Figure 4.2.9 for comparison). At late times, the variation plateaus, likely driven by gas exhaustion, which suppresses star formation, limiting further deviations. This physical constraint explains the observed saturation. Additional tests were performed by varying the number of MPI tasks used to run the clones, which led to greater variation, particularly at the peak of star formation which is not visible in the identical-MPI-tasks set. However, the system appears to plateau similarly to the unperturbed case, possibly due to gas exhaustion regulating the growth of variations. In comparison to σ_{M_*}, the plateauing of dark matter mass aligns closely with the shot noise floor, reflecting the inherently stable nature of collisionless dark matter dynamics.

The run to run variation for each of these properties are similar and less than σ=0.004 (as in figure 4.2.10). This is expected that the noise within each set of simulation is very high, considering the simulations are of very low resolution; we do not see any interesting increase in scatter between runs. Increasing the resolution would reduce noise and potentially reveal chaotic effects in galaxy evolution. These results emphasize the need for systematic convergence tests, as higher-resolution simulations could refine insights into physical processes.

4.2.3 Feedback tests

Figure 4.2.11 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 4.1.2. The variation is computed with Method 2, i.e pairwise comparisons. From figure 4.2.11 we can see that the variation is not very different for each feedback case. 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 (<2Gyr) for HIGH feedback than others, all the cases remain more or less the similar, with variation growing from 11-10Gyr and plateauing to a range of $\sigma_{M_{dm}}$=0.03-0.08 with some scatter.

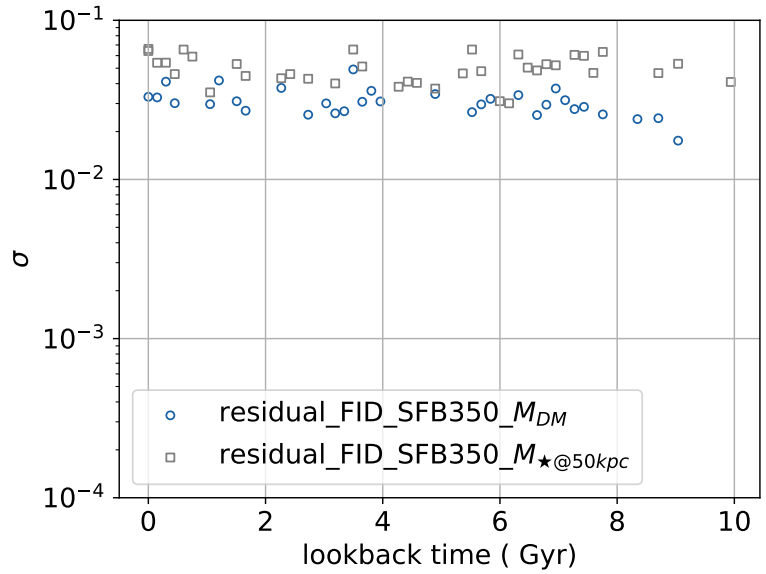

Figure 4.2.9: Comparing noise (computed with Method 1) in galaxy dark matter and stellar mass for fiducial simulations.

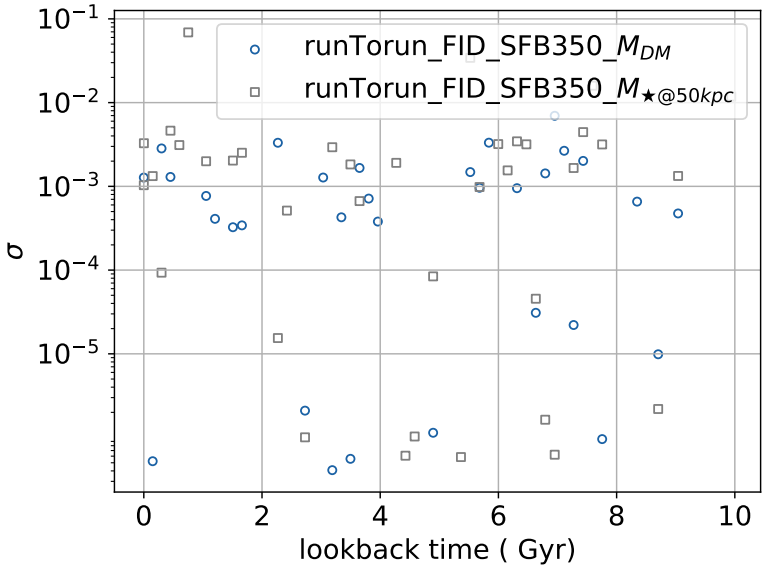

Figure 4.2.10: Same as 4.2.9 but comparing run-to-run variations.

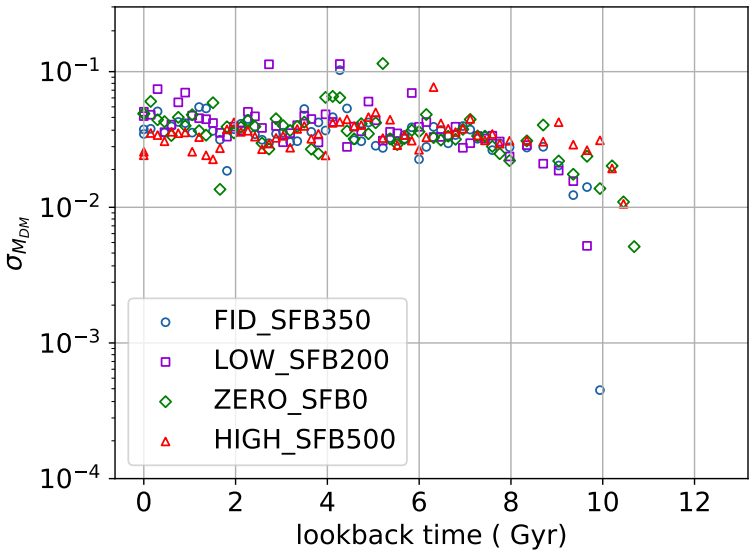

Figure 4.2.11: $\sigma_{M_{dm}}$(only Method 2) in dark matter mass for the 4 feedback variations: Fiducial, LOW, ZERO, HIGH stellar feedback.

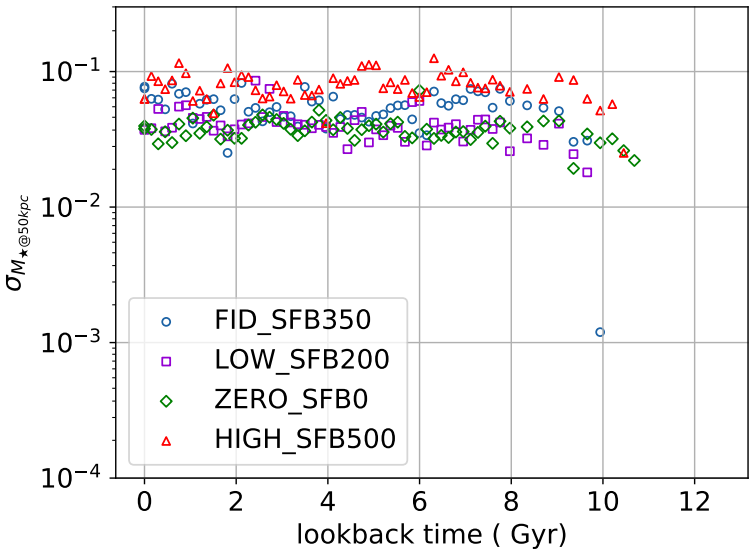

Figure 4.2.12: Same as 4.2.11 but for σ_{M_*}

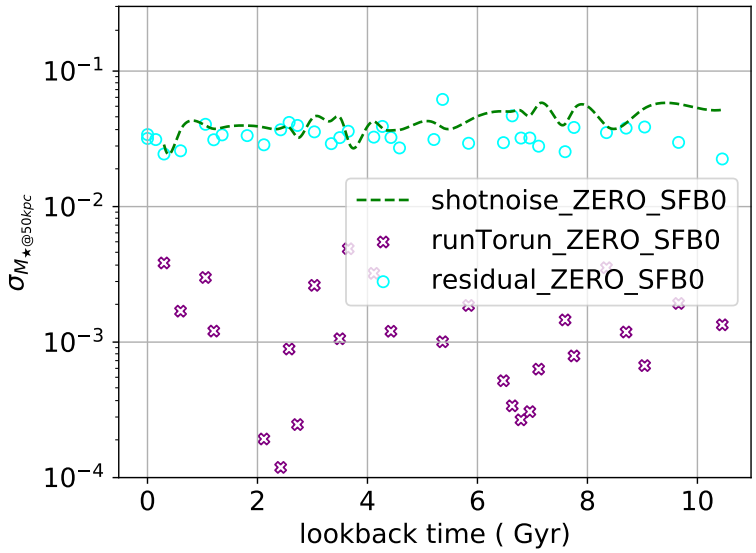

Figure 4.2.13: Noise and run-to-run variation in stellar mass (with Method 1) for only ZERO stellar feedback. Approximate, lower limit estimate of shot noise included as dashed line.

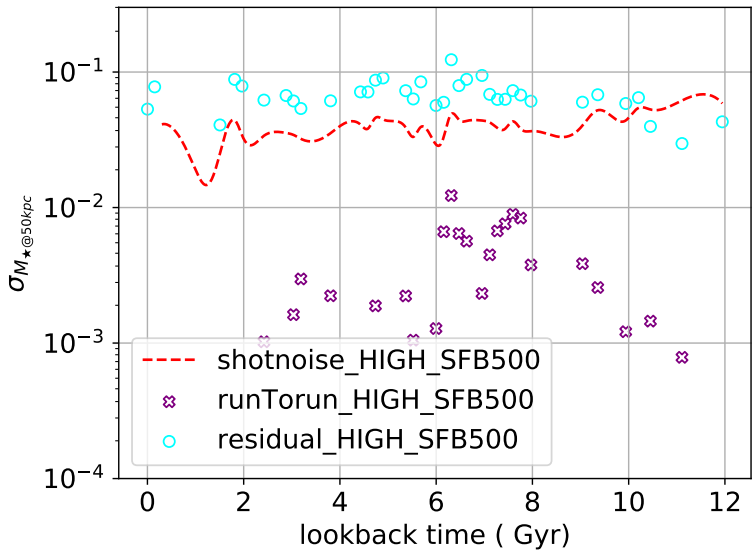

Figure 4.2.14: Same as 4.2.13 but for HIGH stellar feedback.

Figure 4.2.12 shows the variation in stellar mass for each feedback case, where the total variation is computed with Method 2. Clearly, the scatter for LOW and ZERO are lower than that of fiducial (already discussed) by factor of 1-2, but the scatter is even higher than the fiducial for HIGH 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 0.004. Thus, the increase in σ_{M_*} is not due to higher feedback driving an increase in run-to-run variation. This means there is an increase in noise within each simulation. This increase in noise can arise from: either an increase in shot noise or an increase in stochastic variation in each simulation in relation to the feedback. Now, I use Method 1 to understand how the residual noise varies with the shotnoise estimate for 2 cases with the most extreme differences: ZERO and HIGH feedback. In figure 4.2.13, for the feedback case: ZERO, the 2 components of variation (from Method 1) i.e residual noise and run-to-run variation are plotted. The residual variation closely approaches the limit of the shot noise estimate showing that for ZERO feedback case, there are nearly no additional contributions to noise. Whereas for the already discussed Fiducial case in Figure 4.2.8, there seemed a slight increase in total scatter compared to the shotnoise estimate. Now, proceeding to an even more effective stellar feedback, see figure 4.2.14 (same as figure 4.2.13 for HIGH feedback case), there is a markedly higher noise for high stellar feedback. Thus, the increase in σ_{M_*} for increased stellar winds is more likely an increase in stochastic noise within each simulation in response to the respective feedback process (observed increase in stochastic variation can be directly attributed to the nature of the stellar feedback model, which operates stochastically as presented in the introduction: 2.4.5).
Stellar feedback tends to have a localized effect, influencing regions close to star-forming areas. This can lead to patchy or bursty star formation because feedback from supernovae or stellar 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 higher scatter in the stellar mass across different regions of the galaxy. The effect can be further assessed by checking the impact of stellar feedback driving high variation in the cold gas mass as shown in figure 4.2.15 Note that the variation in quantities such as cold gas mass or instantaneous star formation rate, blackhole accretion rate etc, are highly variable (as seen, is higher than 0.1 σ_M for cold gass mass), considering they are highly susceptible to short-term environmental changes, gas dynamics, not to mention interactions with other processes such as star formation, blackholes etc. In contrast, time averaged quantities are accumulated more smoothly over time, making them less sensitive to short-term fluctuations. Here, the total variation as computed with Method 2 is plotted, shows a high variability (σ_{gmcold} is of the order 0.4-0.8) throughout the evolution of the simulation and is several factors higher for the relatively higher stellar feedback cases, particularly in earlier times. The increased variation for higher feedback causes a variation in the SFR, ultimately resulting in a variation in stellar mass in the galaxy population within each individual simulation. Thus high stellar feedback, in simulations without blackholes, increases the stochastic

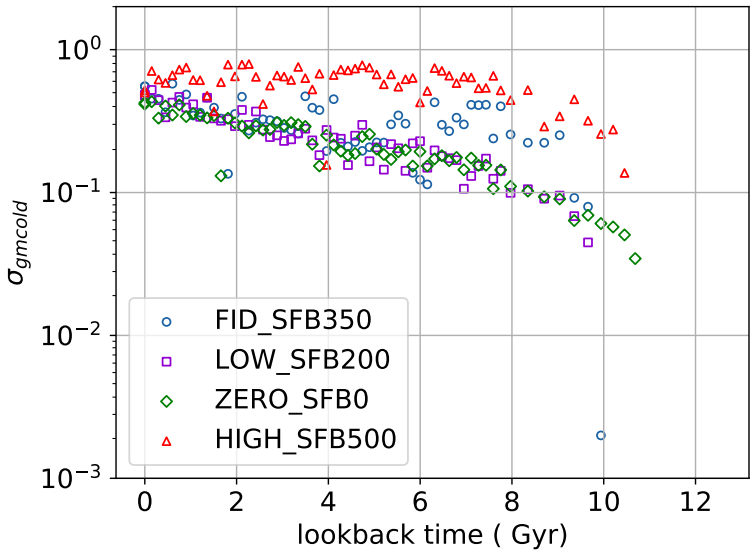

Figure 4.2.15: Total variation (Method 2) in galaxy cold gas mass for the four feedback variations: Fiducial, LOW, ZERO, HIGH stellar feedback.

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 the total variation approaches the lower limit of the shot noise, see figure 4.2.13 for ZERO feedback case. Also, these galaxies may show less variability in their star formation histories (albeit unrealistically high SF), as star formation proceeds smoothly and continuously until all the gas is exhausted. This reasoning is particularly discussed in [76] (see summarized in introduction: 3.2), 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; 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 identical simulation for both 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. See figure 4.2.16, where the variation in stellar mass is plotted in their dwarf galaxy simulations (128 identical runs) run without feedback. Each row considers a different gas fraction, starting with the lowest gas fraction at the top $f_g = 0.05$. Red dashed line indicates a 10% variation from the mean of stellar mass distribution of the clones. Left panel plots galaxy stellar mass and the right panel shows variation computed as ratio of stellar mass between pairs. The variations in this case is well below 10% when the gas fractions are high enough to maintain a high star formation efficiency, but as seen for the cases with lower gas fractions, the variation grows much higher than what is expected from shot noise alone.

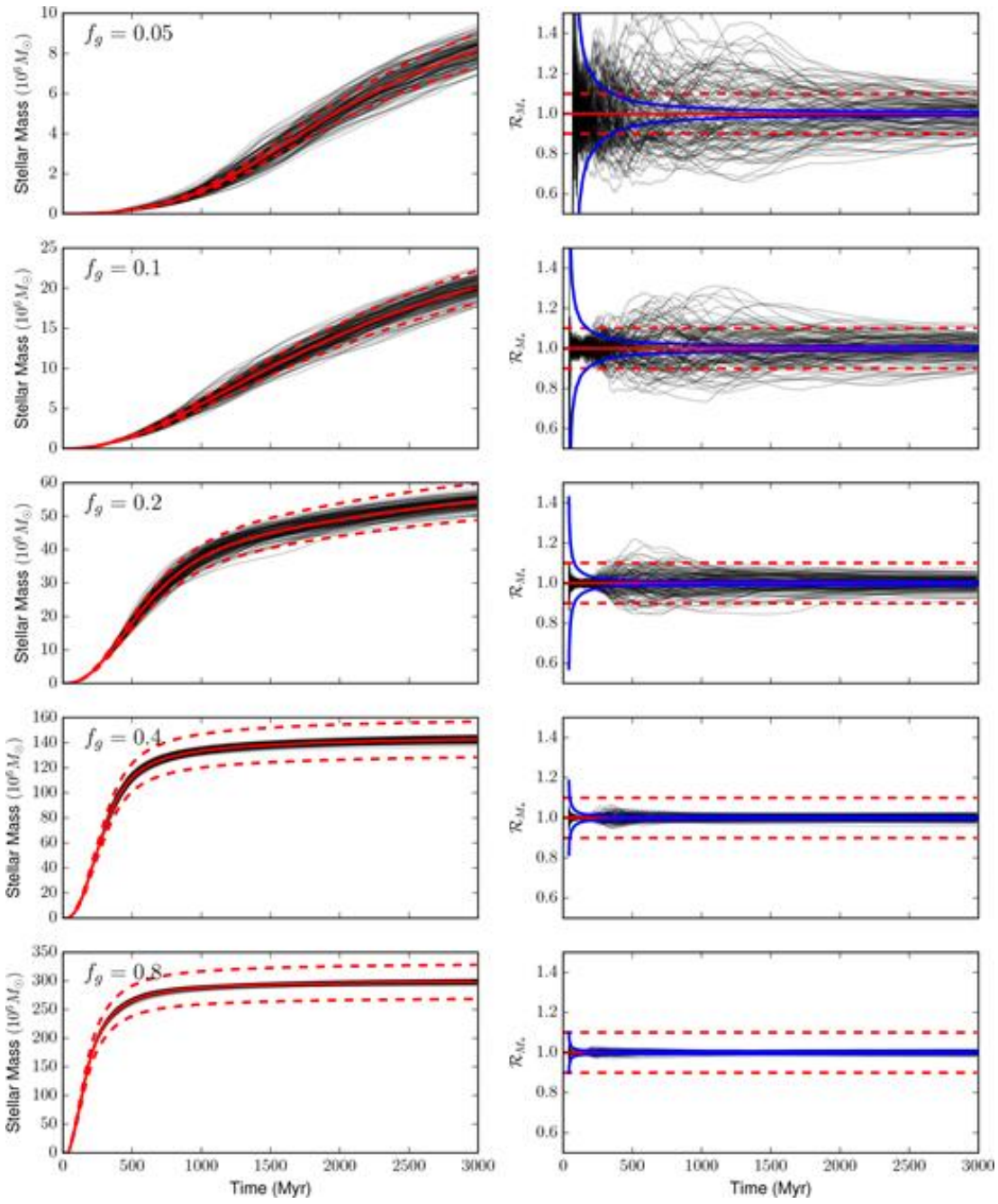

Figure 4.2.16: Variation in stellar mass as seen with varying gas fractions (f_g) in dwarf galaxy simulations (128 identical runs) run without feedback. Each row considers a different gas fraction. Left panel shows the stellar mass and the right panel shows the ratio of stellar mass between pairs. Red dashed line indicates a 10% variation from the mean. Credits:[76].

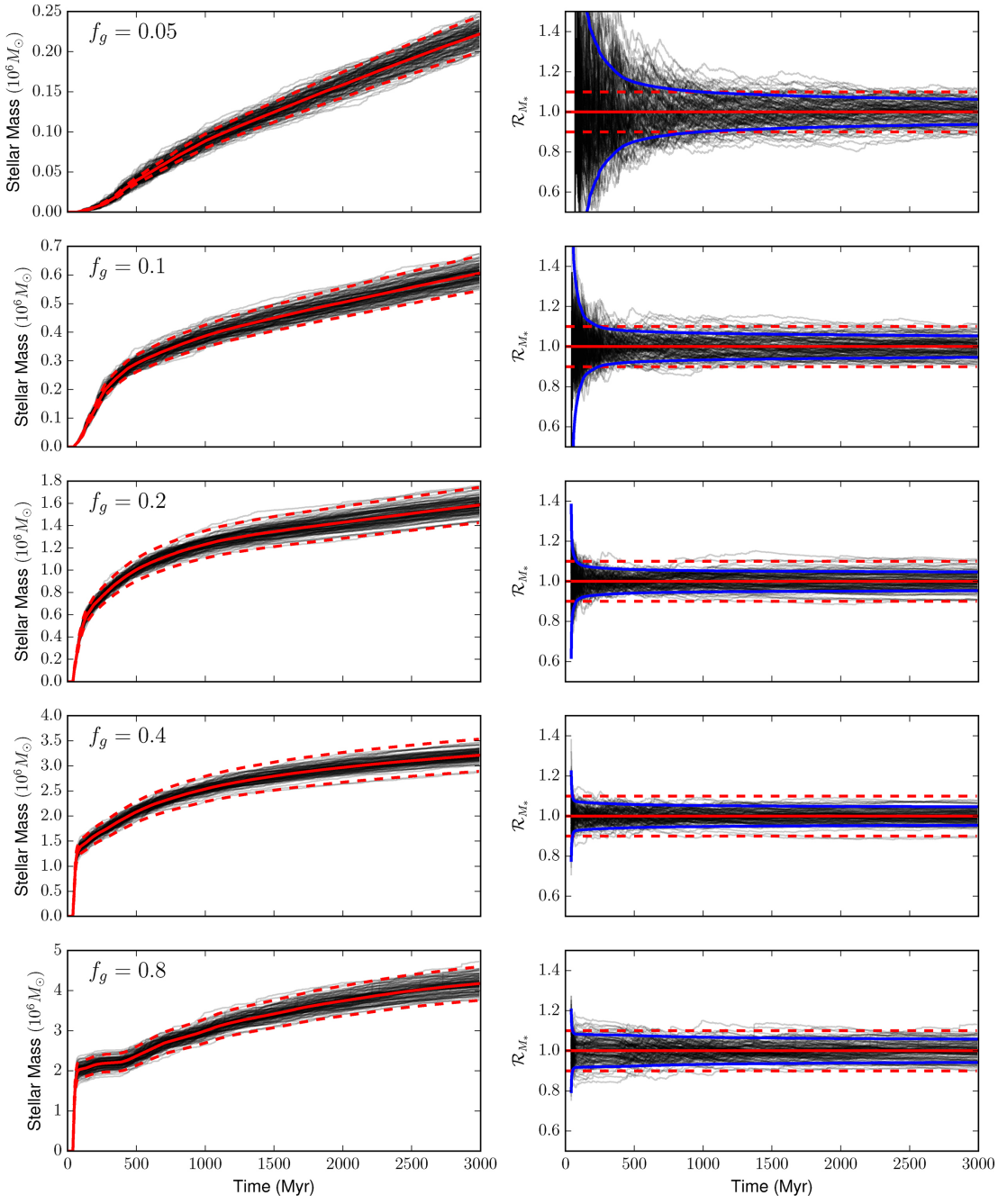

Figure 4.2.17: Same as 4.2.16, except feedback is included. Credits:[76].

But with stellar feedback, see figure 4.2.17, the fluctuations in stellar mass are maintained or reduced regardless of the gas fraction. These findings are reported for the case of isolated galaxy, but this effect on a galaxy evolved in a cosmological context is not shown in [76]. They do report the lower variations in cosmological zoom-in galaxy for unregulated feedback but not the behavior of the system with lower gas fraction with respect to the presence of feedback or lack thereof. It could very well be, that if the the test with varied gas fraction is made, that this particular result in a cosmological context could not be visible. This is considering the fact that in a cosmological setting, galaxy evolution is much more complex with many mergers, which greatly affect variations.
Nevertheless, additionally to the method of varying the gas fraction in the initial conditions, the relation of feedback and variation can be explored in presence of blackholes which could disrupt the one-to-one relation between high SFE-gas depletion-and lowered scatter in galaxy stellar mass.
The randomness introduced by stellar winds dominate in the absence of additional regulating mechanisms like black hole feedback. Black hole feedback, when included, can suppress star formation and moderate turbulence, thereby reducing unchecked variation. Testing for the effect at higher resolutions could lead to improvements as finer hydrodynamic interactions are resolved, reducing reliance on subgrid approximations and mitigating noise. However, feedback dynamics may still retain some degree of stochasticity, as these processes inherently involve chaotic behavior. Importantly, how subgrid models scale with resolution remains a critical factor: while higher resolution may smooth feedback processes, the effects on stochastic noise require further testing. To solidify these conclusions, simulations at higher resolution as well as inclusion of blackholes will be explored in future work.

Variation in different galaxy mass bins

The results for σ_{M_*} shown so far considers a large distribution of galaxies above $M_* > 10^8 M_\odot$. In this part, I consider how the stellar mass can vary within galaxies belonging to different mass bins. In figure 4.2.18, 4.2.19 and 4.2.20, σ_{M_*} is plotted for galaxies belonging to mass bins $[10 < logM_*(M_\odot/h) < 10.5]$, $[10.5 < logM_*(M_\odot/h) < 11]$ and $[11 < logM_*(M_\odot/h) < 11.5]$ respectively. Each mass bin always considers a sample size > 100; Galaxies outside these mass bins are excluded due to poor sample sizes. Note that this is why LOW and ZERO have some missing data at >4Gyr time in 4.2.18, considering unregulated or weakly regulated star formation leads to galaxies quickly evolving into ones with more significant stellar mass.
As seen in figure 4.2.18 and figure 4.2.19, the results of the full distribution is 4.2.12 is still reflected. Thus for the galaxies of stellar mass within $[10 < logM_*(M_\odot/h) < 11]$, 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 $[11 < logM_*(M_\odot/h) < 11.5]$, see figure 4.2.20, 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

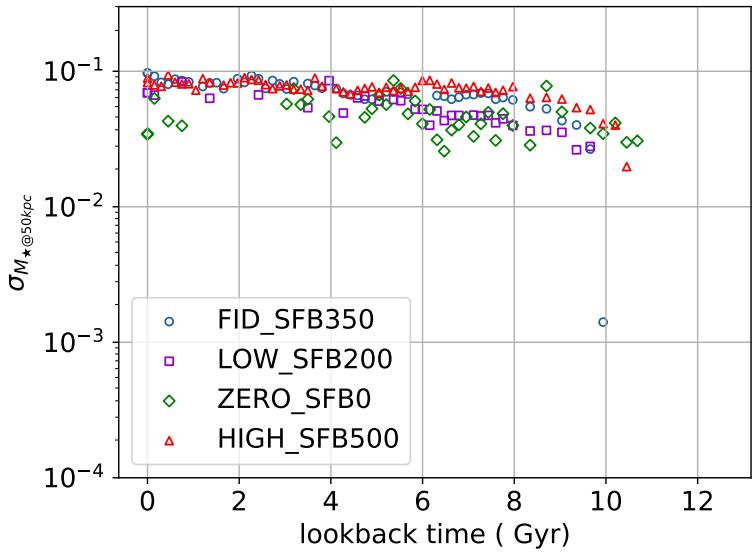

Figure 4.2.18: Total variation (Method 2) in galaxy stellar mass within a mass bin of $[10 < logM_*(M_\odot/h) < 10.5]$ for the four feedback variations: Fiducial, LOW, ZERO, HIGH stellar feedback.

experience a stronger impact from stellar feedback due to shallower gravitational potential wells; the energy and momentum of supernovae feedback or stellar winds can drive gas out of the galaxy more efficiently, thus limiting the available gas for star formation and thus having a 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. In these systems AGN feedback typically is the more dominant effect for regulating star formation. Thus, for the relatively larger mass bin $[11 < logM_*(M_\odot/h) < 11.5]$, the effect of HIGH feedback is not evident and thus the stochastic variation is more or less similar to a lower or absent stellar feedback.
Thus, for the specific set of simulations configuration I tested, the implementation of stellar feedback and configuration of the feedback strength: HIGH (also fiducial) stellar feedback particularly affects the galaxies with $M_* < 10^{11}\, M_\odot$ increasing the stochastic variation, whereas more massive galaxies are unaffected. This illustrates the complex interaction of physical processes, i.e the type and strength of feedback and its impact on stochastic processes and thus the respective variation. Broader conclusions will require extensive tests across various configurations of feedback models and parameters.

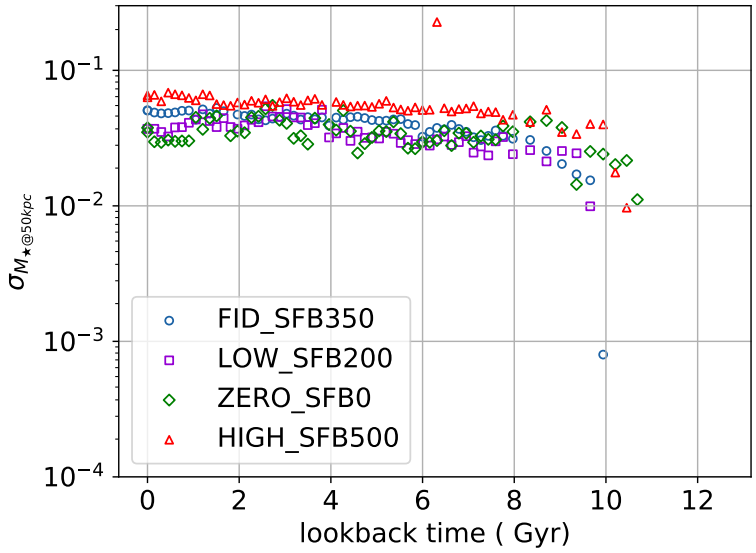

Figure 4.2.19: Same as 4.2.18 but for mass bin of $[10.5 < logM_*(M_\odot/h) < 11]$

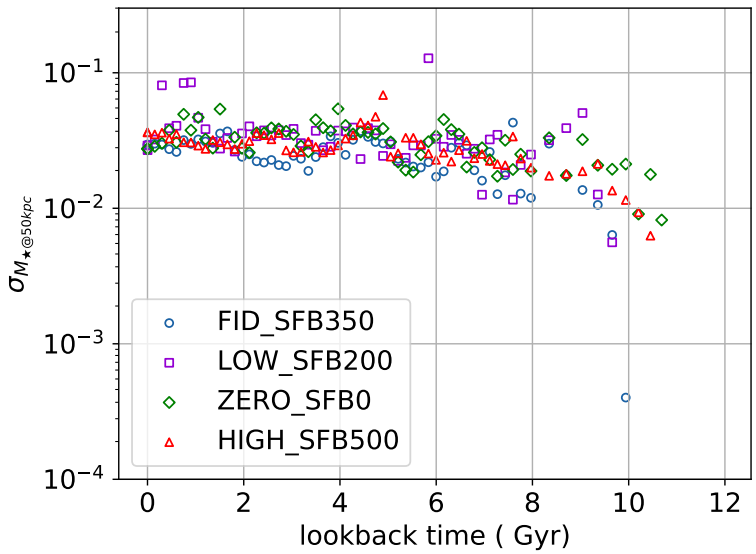

Figure 4.2.20: Same as 4.2.18 but for mass bin of $[11 < logM_*(M_\odot/h) < 11.5]$

Part 2

Chapter 5

Cosmological hydrodynamical simulations archive

5.1 Introduction

Cosmological numerical simulations have been a predominant constituent in the field of Cosmology. Studying structure formation and evolution at cosmic scales is a complicated problem. The multi-scale complexity, involving diverse non-linear physical processes is something that can be systematically addressed with the methods of numerical simulations.
Simulations have gained grounds in modeling various important physical phenomena such as effects of star formation, chemical enrichment, and feedback processes, and more, thanks to the combination of the gravity and hydrodynamics processes with so-called "sub-resolution" models.
Their usage is applicable to the majority of activities related to astrophysical data: from the reduction and analysis of astronomical instruments data up to their interpretation and comparison to theoretical predictions, including simulations and theoretical modeling. They are vital during the preparatory and operational phases of new scientific experiments, efficiently guiding and shaping instruments and observatories. They can also help to capture and analyze the torrent of complex observational data that the new generation of observatories will produce (e.g. the Square Kilometer Array (SKA), MeerKAT+ and LOFAR2.0 observatories, the Cherenkov Telescope Array (CTA), the European Extremely Large Telescope (E-ELT), the Vera Rubin Observatory, the low-frequency space-born Laser Interferometer Space Antenna (LISA), the EUCLID and WFIRST satellite missions), providing new insights into astronomical phenomena, the formation and evolution of the universe, and the fundamental laws of physics. The new generation of instruments will produce exponentially more data than their predecessors, and will need outstanding resources to be post-processed, analyzed and stored.
This will lead to new IT laboratories (codes/algorithms/tools/services) able to investigate the physical processes behind the observed phenomena with unprecedented quality, resolution and reliability, allowing their interpretation and paving the way for scientific discovery.

For these laboratories, the efficient and effective exploitation of HPC computing capabilities is fundamental: the combination of high-performance data processing with post processing processes and data visualization will help in the exploitation of the expected large data volumes. The comparison of observations with large scale simulations will be possible, based also on the high performance archiving and data access capabilities necessary to make the next generation astrophysical data stores easily accessible to the whole astronomical community.

Archiving and sharing numerical simulations results is then, of primary importance in the scientific exploitation of the numerical data and making scientific discovery. Cosmological simulations are currently mostly hosted at HPC centers, wherein access is restricted to close collaborators. Additionally there is a general lack of HPC expertise and technical know-how of interpreting simulation results and a sparsity of tools needed by astronomers to expedite validations and comparisons to observational data.

In this aspect, it's not just that one simply ensures an easier data access for the researchers. A crucial, yet often overlooked, step in data publishing, is to regularize the access so that said data is indeed usable and reusable in the long term. Such concepts of ensuring usability of published data is encapsulated with the so called FAIR guidelines principals (refer [154]) and Virtual Observatory (VO, see https://www.ivoa.net/) implementation. FAIR guidelines mandate a digital identifier for any published data, annotate it with rich metadata and use standardized framework and communication protocols. This ensures that the data is Findable, Accessible, Interoperable, and Reusable (FAIR) both by machines and their human counterparts. VO principles, which goes hand in hand with FAIR guidelines, work towards an objective wherein astronomical data sets should work coherently with other data sets and resources, such that all the distributed data centers and services throughout the world form one single, global virtual observatory. This is accomplished by the set of technical standards, all laid out by the VO which help establish a global standardized infrastructure. The idea here is not one of getting a specific set of software or data access means to work but rather to enable the data centres worldwide to altogether offer a variety of compatible tools and services. This means, one could be tapping into a wealth of powerful tools and resources to visualize and perform further analysis on any data, all available within the VO infrastructure.

This work is aimed towards creating an open ever-growing, curated, verified and validated simulated data archive, managing both raw simulation outputs and post-processed products, that can be also extended in future and complemented with new features such as including catalogs of mock images mimicking real observations to compare with. FAIR principles for data management and stewardship will be followed, strongly contributing to their promotion within the community and beyond. Also, standardized protocols will be used in the design of the data access channels as well as to provide some post-processing services for data analysis.

The Data Archive is the where simulation inputs, outputs, metadata and higher level products are stored and indexed.

There already exists some archives hosting cosmological simulations e.g. G. Lem-

son and the Virgo Consortium 2006 [86], data access portal for MillenniumTNG Project [100]. Among these, MultiDark [120] is an example of a generic database for cosmological simulations and is a VO oriented database hosting several cosmological simulations.

The fundamental objective of this work is designing and deploying a flexible and expandable platform for hosting cosmological simulations, that implement the VO specific standards for simulations apart from the more generic VO standards. And thus, to accommodate cosmological simulations from a variety of projects. Rather than creating a rich interface and analysis tools, the focus is on standard interfaces and protocols that will allow machine consuming of the data archive resources for post processing and visualization, but also for comparing simulated data with observations. As a matter of fact, interface and analysis tools are often very specific to the type of simulation data that they are run on, on the other side generic services that applies to all simulations will be continuously identified and integrated as part of the archive post processing services. But, this in essence is the objective of the VO simulation standards, to provide standard services that require lightweight interfaces at the client side, custom made, to access the data which is heterogeneous in nature.

This part of the thesis is organized as follows: Section 5.2 provides a brief introduction to FAIR and VO standards. Section 5.3 describes the design of a database for simulations and a web interface. Section 5.4 is regarding the implementation of the data standards.

5.2 Data standards in astronomy and astrophysics

A key development in the scientific management of data is the adoption of the *FAIR data principles*, which emphasize that data should be **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (FAIR). Simultaneously, the **International Virtual Observatory Alliance (IVOA)** has been instrumental in establishing a comprehensive set of standards that support these principles, ensuring that astronomical data is universally shareable and usable across diverse platforms and disciplines.

In regards to other existing data sharing frameworks (other than that of IVOA), we have HDF5 (Hierarchical Data Format 5; widely used for efficient storage and organization of large datasets, providing high performance in reading and writing hierarchical data), NetCDF (Network Common Data Form; supports self-describing and portable data formats, often utilized in climate and earth sciences), OpenDAP (Open-source Project for a Network Data Access Protocol; enables remote access and subsetting of large datasets) etc. While these frameworks excel in format efficiency, they lack the domain-specific metadata interoperability offered by IVOA standards, which align closely with FAIR principles to enhance scientific discovery and reproducibility. This section introduces the FAIR data principles and IVOA standards in the context of astronomical research with respect to data management, reproducibility and fostering better interoperability.

5.2.1 The FAIR Data Principles

The FAIR data principles, first outlined by [154], provide a guiding framework for ensuring that scientific data is well-managed and optimally reusable by both humans and machines. The FAIR principles consist of four core pillars:

1. Findability: Data should be easy to locate by both humans and automated systems. This is achieved through the use of persistent identifiers (e.g., digital object identifier, DOIs) and metadata that enable effective search and retrieval processes. A key aspect of this principle is the availability of rich, descriptive metadata that aids in data discovery.

2. Accessibility: Once found, data should be easily accessible, ideally through open-access protocols. Accessibility does not necessarily imply that data should be open without restriction, but that its terms of use are clear, and it is retrievable using standardized communication protocols.

3. Interoperability: To be useful, data must be compatible with other datasets and tools. This requires the use of standardized data formats and vocabularies that allow for cross-disciplinary integration. For example, astronomical data in the Flexible Image Transport System (FITS) format is interoperable with many software tools and archives, enhancing its utility across the field.

4. Reusability: To maximize the impact of data, it should be reusable for future research and not loose meaning over time. This involves detailed metadata, and sufficient documentation of data provenance and quality, allowing for replication of existing results and fostering further discoveries.

5.2.2 The International Virtual Observatory Alliance (IVOA) standards

As astronomy has become increasingly data-driven, the need for a global framework for data interoperability has grown. The **International Virtual Observatory Alliance (IVOA)** was founded in 2002 to address this need by developing and maintaining standards that support the exchange and integration of astronomical data worldwide. The IVOA standards align closely with the FAIR principles, particularly in ensuring the findability, accessibility, and interoperability of astronomical data.
The Virtual Observatory (VO) is the result of these efforts- an international framework that allows astronomers to access and analyze data from multiple archives and observatories through a unified interface. The IVOA oversees the development of protocols and tools that facilitate this global infrastructure, ensuring seamless integration and cross-compatibility of data across different platforms. Some of the most critical IVOA standards include:

- Registry [35]: facilitates the discovery of astronomical resources, including datasets, services, and other scientific assets. Registries provide a struc-

tured way for resource providers to describe their offerings using standardized metadata, making it easier for users to search and retrieve relevant data. Each resource is registered using a Resource Metadata format [108], allowing automated tools and researchers to query registries and locate services or datasets of interest across the Virtual Observatory. Registries are distributed but interoperable, enabling global discoverability of astronomical data and services.

- Data model : The IVOA Data Model defines the standards for how astronomical data and metadata are structured and described. It provides a consistent framework for representing various types of data, including observational data, catalogs, and simulations. By standardizing data structures, the Data Model ensures that data from different sources can be interpreted and processed uniformly, supporting interoperability across the Virtual Observatory. Key examples include the Observation Data Model (ObsDM [91]) and the Simulation Data Model (SimDM [87]), each providing specific ways to describe observational or simulation-based data.

- Vocabulary: For instance, Unified Content Descriptors (UCD [33]) provide a standardized vocabulary for describing the content of datasets. UCDs are standardized labels used to describe the semantics of data columns in astronomical datasets. UCDs enable consistent interpretation of data by associating each column or data product with a well-defined meaning. For instance, a column that stores celestial coordinates might be labeled with a UCD like "pos.eq.ra", making it clear to any user or software what kind of data is present. UCDs are essential for facilitating the automatic discovery and interpretation of data in the VO, ensuring that datasets from different archives use a common language for describing content.

- Data access: refers to a suite of protocols and standards developed by the IVOA to enable uniform access to astronomical data. DAL services allow users to query, retrieve, and interact with different types of data, such as images, spectra, or catalogs, using standard interfaces. Some of the key DAL protocols include:

  - Simple Image Access Protocol (SIAP [36]) for accessing images.
  - Simple Spectral Access Protocol (SSAP [148]) for retrieving spectral data.
  - Table Access Protocol (TAP [37]) for querying and retrieving tabular data.

- VOTable [104]: is a standard XML format for representing tabular data, clearly defining both data and metadata. It is widely used in astronomical research for sharing data between different services and tools. VOTable is flexible, extending beyond pure XML to being compatible with FITS format as well.

Each of these technologies plays a crucial role in the Virtual Observatory, supporting various facets of discovery, access, and interpretation of different datasets. The relation between the FAIR principles and IVOA standards lies at the core of astronomical data management. Both frameworks aim to improve the accessibility, integration, and reuse of data: While the FAIR principles provide a high-level blueprint for data stewardship across scientific domains, the IVOA standards offer the practical tools and protocols required for implementing these principles in the specific context of astronomy.
Thus FAIR is emulated within IVOA as,

1. **Findability** is enhanced through the use of persistent identifiers in IVOA-compliant datasets and rich metadata standards such as the Resource Metadata model.

2. **Accessibility** is facilitated by the Simple Access Protocols (e.g., SIAP for images, SSAP for spectra), which ensure that data can be retrieved using standardized methods from distributed archives.

3. **Interoperability** is ensured by common formats like VOTable, as well as the use of UCDs for metadata description, enabling data from different sources and wavelengths to be analyzed together.

4. **Reusability** is supported through the IVOA's emphasis on clear data provenance and the use of comprehensive metadata, ensuring that datasets are accompanied by the necessary contextual information for future reuse.

The adoption of the FAIR data principles, in conjunction with the implementation of IVOA standards, has transformed the way astronomical data is managed, shared, and reused. As we enter the era of "big data" astronomy with projects like the LSST and the James Webb Space Telescope (JWST), ensuring that data adheres to these principles will be essential for maximizing scientific output and fostering global collaboration. With the FAIR principles at its core, IVOA is at the forefront of proper data management in astronomy, ensuring that future generations of astronomers will have the tools and resources needed to use and re-use astronomical datasets.

5.3 Custom Database and Interface Implementation

A preliminary custom database and a web-based interface were developed as an initial implementation serving as a foundation for understanding the structural requirements, file organization, and associated challenges. Although not standardized, this prototype highlights essential insights that will inform future efforts to align the system with best practices in data management and interoperability. As an example, a set of 29 hydrodynamical simulations on galaxy clusters called the "Dianoga" (See introduced in methadology of Part 1: 4.1.1) have been considered for the first ingestion into the archive database. These hydrodynamical simulations, conducted using the GADGET code (see Section 2.4.1), model various physical processes, including star formation, feedback mechanisms, and

cooling. This example provided valuable insights into parsing simulation outputs and tailoring the archive to accommodate varying file structures across datasets. This section provides some generic implementation details of the application.

Metadata and Archive Setup: Accurate metadata is essential for effectively annotating data objects, providing a comprehensive description of the dataset. In the context of simulations, this includes not only the final data products but also details about the setup, input parameters, and configuration used during the simulation process. Hydrodynamical simulations involve numerous configuration and output files. Rather than storing exhaustive parameter lists, the database highlights key parameters like particle masses, numbers, and softening lengths, which help users efficiently identify relevant simulations.
The archive organizes data hierarchically using a relational database, see Figure 5.3.2 which shows the database design. At the top level, a "suite" table collates metadata common to a group of simulations, while a "simulation" table stores information specific to individual runs. Cosmological parameters common to a suite, such as Ω_m, Ω_b, box size etc., are stored in the suite table, while run-specific data is extracted from snapshot files and cataloged in the simulation table.

Snapshot Files : These raw data products capture particle states i.e. position, velocity, mass, density, energy etc., at specified times or redshifts. Depending on simulation settings, snapshots may be stored in a single binary file or split across multiple files. Customized scripts are required to parse snapshots of a given simulation considering there is no standard format for simulation outputs. For Gadget simulations, snapshot file format can be a standard binary block structure, an extended variant with optional blocks, or in hierarchical data format (HDF).

Group Catalogs : These files, generated using clustering algorithms like friends-of-friends (FOF) and Subfind (see Section 2.4.7), identify virialized halos and subhalos within snapshot files. Group catalogs are stored in a similar format to snapshots and provide information on properties such as mass, radius, temperature etc.
For parsing Dianoga snapshot and Subfind files I use g3read and g3matcha libraries (Refer: https://github.com/aragagnin/g3read). Snapshot and group catalog files are classified as Level 0 raw data in the archive, organized by redshift or time. Select properties of prominent halos and subhalos are extracted and stored as Level 1 data, making them more accessible for analysis.
To improve accessibility, Level 2 data includes visual summaries and pre-made plots, such as 2D halo maps and radial density profiles. These previews give users an intuitive understanding of the simulation content, complementing raw data exploration.

Archive User Interface The database is accessible via a web application built using Django, a Python-based framework. The interface allows users to browse simulation metadata and download files. A typical simulation landing page (e.g.,

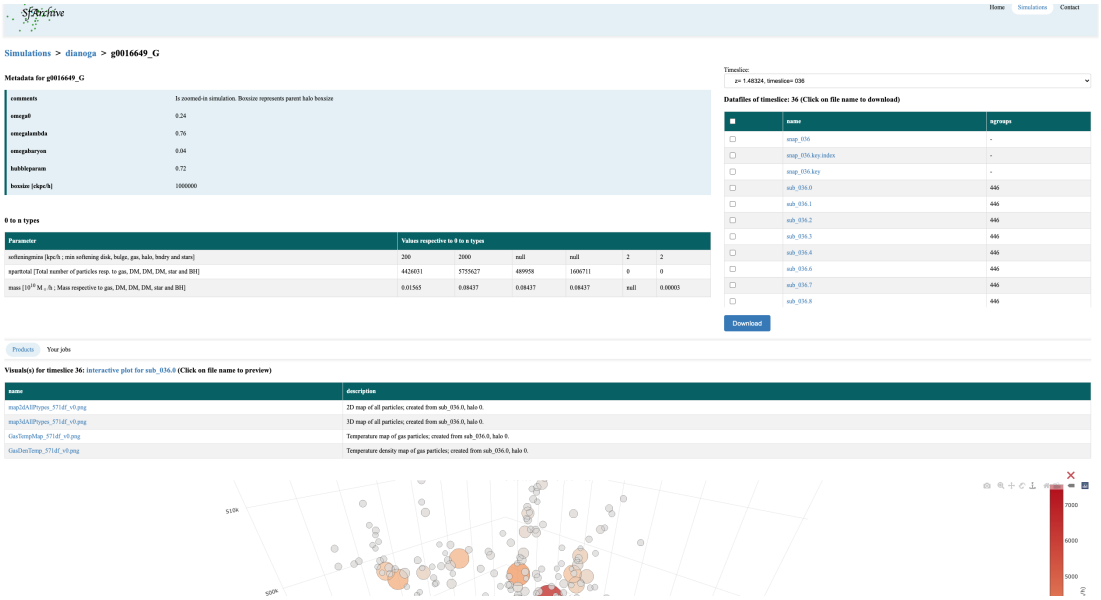

Figure 5.3.1: Archive front-end.

Figure 5.3.1) displays configuration parameters and lists raw data files by redshift or time.
Additional features include:

- An interactive plot of select catalog files shows the main halos with their characteristic information such as mass, radius etc.

- Pre-made visualizations of select halos.

These tools enable quick exploration of simulation outputs while offering server-side options, minimizing the need for users to process raw data locally.

Deployment and Implementation To streamline the deployment process and enhance reliability across diverse infrastructures, the archive software employs containerization— an approach that simplifies application deployment by bundling code and dependencies into isolated units known as containers. Unlike traditional deployment methods, containerization abstracts the operating system, enabling applications to share the host OS while maintaining lightweight and efficient operation.
For this project, Docker, a widely used open-source containerization platform, has been selected to manage and deploy the archive's components. Docker containers encapsulate the database service, web application server, task queue, and user interface as individual "images." These images ensure consistent behavior across development, testing, staging, and production environments, regardless of the underlying hardware or operating system. All of the multiple containers here are seamlessly managed with docker compose which is a one stop point to build and run each service as well as manage the life-cycle of the collective application. This approach guarantees portability and reliability for the software.

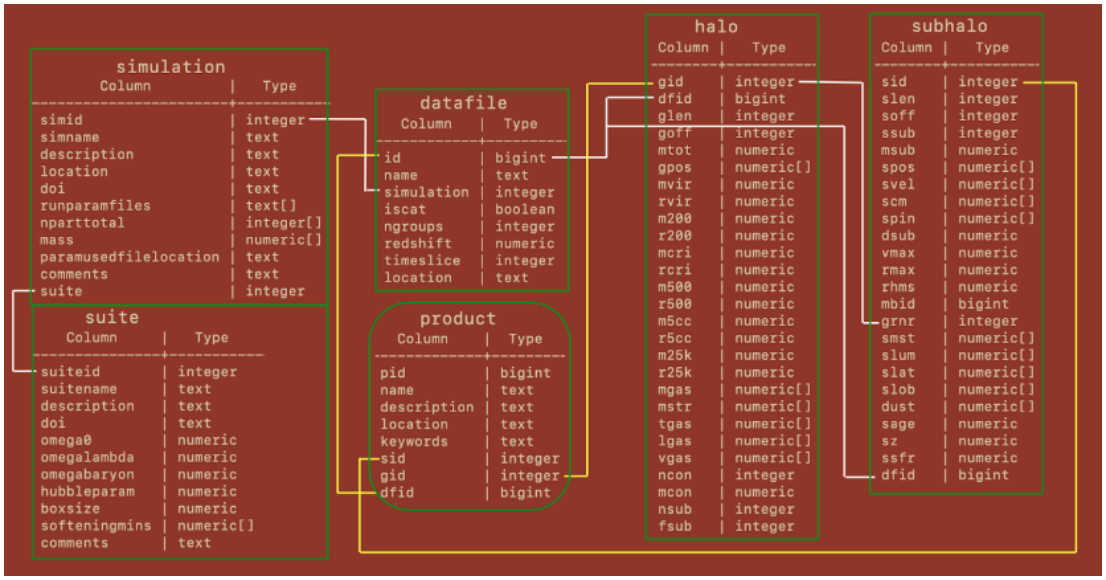

Figure 5.3.2: Database design of the archive.

5.4 Standard implementations

The archive will be deployed on an infrastructure that assigns DOIs to the simulation datasets, ensuring findability and accessibility aspects of the FAIR data principles. For interoperability and reusability, the adoption of VO standards will enable standardized metadata descriptions and seamless integration with existing tools and services. This section covers the details of the work related to adoption of VO standards.

5.4.1 First steps of standards implementation

The first steps into standardization involve implementing some foundational IVOA standards that are required in the VO environment and in wider scheme of things. Two important standards that were implemented as initial steps are the VOTable format and the Universal Worker Service (UWS) protocol.

VOTable , as already introduced in 5.2.2, is a IVOA standard designed for the storage and exchange of tabular data, such as simulation outputs or observational data. VOTable uses XML-based syntax to represent tables where each column is described with metadata, allowing easy interpretation of the data's structure. It is particularly useful in the VO framework because it for instance, supports embedding Uniform Content Descriptors (UCDs), which semantically describe the content of each column, and it handles complex data types like arrays or binary structures efficiently.

Figure 5.4.3 shows the output of this implementation, a list of simulations in the archive database, represented in VOTable format.

Figure 5.4.3: Simulations listed in VOTable format

Universal Worker Service

Few features that are added to provide additional services to archive users. These include access to outputs or plots as well as creating new ones. Depending on the input parameters, we can expect prolonged execution times for given requests. These services are implemented exclusively as asynchronous services. This is given, within a context where data which is handled, i.e. simulation data, is large and thus the response time is expected to be long.
This aspect of a given service is described by Universal Worker Service (UWS [56]), which is another key IVOA standard. This protocol is for service requests that takes more than the time taken for rendering a dynamic response to the user. The request instead is registered and replied to with the registered id. At the service backend, the request prompts the creation of a job, which is executed upon availability of required resources and other criteria. The communication henceforth between the user and service is followed as per the UWS protocol, i.e. checking the status of the sent request, edit or abort an initiated job. Upon completion of the job the user can download the results from the given location provided by the details of the completed job.
Figure 5.4.4 provides a pictorial representation of the various states of a UWS job:

- PENDING: The job is registered but awaits resource allocation.
- QUEUED: Resources are allocated, and the job awaits execution.
- EXECUTING: The job is actively running.
- HELD: The job is temporarily paused, but not aborted.

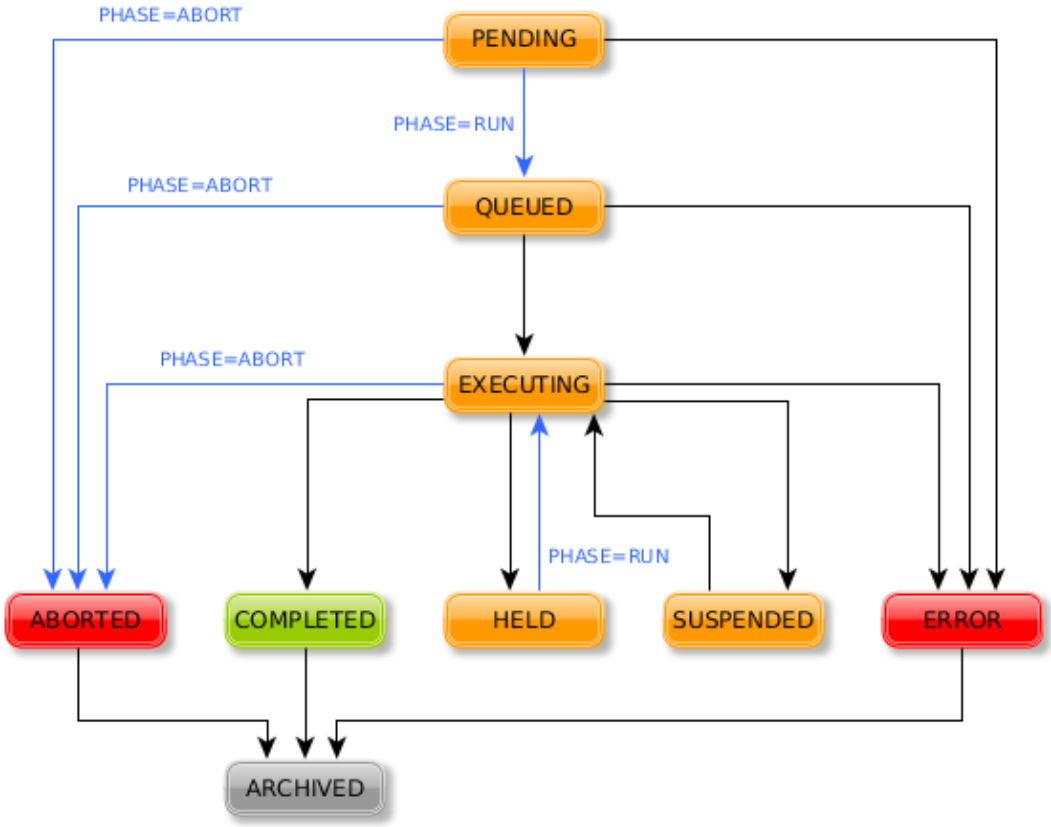

Figure 5.4.4: Various phases of a UWS job where color of the arrows signify transitions initiated by client(blue) and by server(black). Figure taken from [56].

- SUSPENDED: Execution is stopped, requiring intervention or reconfiguration.
- COMPLETED: The job successfully finishes, and outputs are available.
- ABORTED: The job is terminated before completion, often by user request.
- ERROR: The job fails due to issues during execution.
- ARCHIVED: Completed or aborted jobs are stored for retrieval.

By using UWS, the system can manage multiple requests efficiently without requiring the user to wait for immediate results, improving overall scalability and user experience. Figure 5.4.5 shows the related implementation where the details of a registered job is listed, along with its status.

5.4.2 IVOA simulation standards

In regards to data sharing frameworks closely related to simulation, we have HDF5 (Hierarchical Data Format 5; widely used for efficient storage and organization of large datasets, providing high performance in reading and writing hierarchical data), NetCDF (Network Common Data Form; supports self-describing and

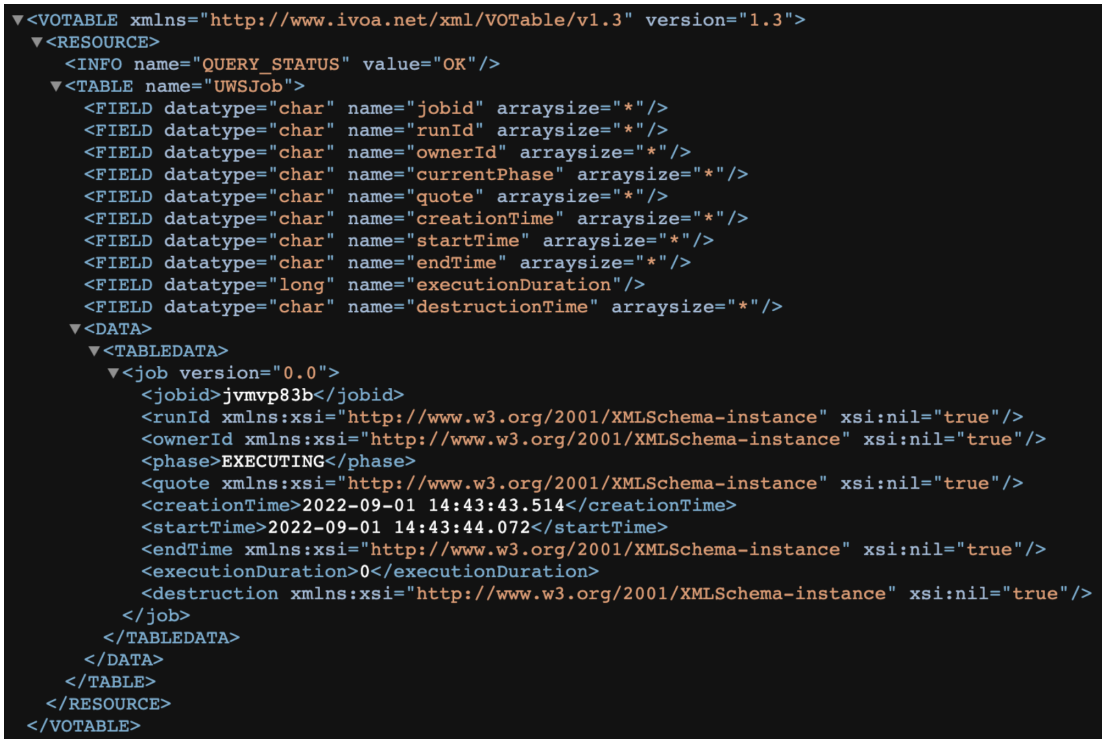

Figure 5.4.5: Details from UWS service on executing job

portable data formats, often utilized in climate and earth sciences), OpenDAP (Open-source Project for a Network Data Access Protocol) enables remote access and subsetting of large datasets etc. While these frameworks excel in format efficiency, they lack the domain-specific metadata interoperability offered by IVOA standards, which align closely with FAIR principles to enhance scientific discovery and reproducibility. Metadata interoperability is dependent on community requirements and thus is tuned better on the needs of the specific community.
The VO standards for simulation data are Simulation Data Model (SimDM) and Simulation Data Access Layer (SimDAL). These are the key frameworks developed in an attempt to standardize access of simulation data. SimDM, introduced by [87], provides a structured model for describing simulations, their configuration, parameters, and results, ensuring that simulation metadata can be consistently organized, annotated with the right metadata and thus, shared. It supports the interoperability of diverse simulation datasets, facilitating their use in broader astronomical research. Building on this, SimDAL, developed by [82], extends these capabilities by defining a protocol for accessing simulation data through a unified interface. In this section, specific details and related implementation are described for SimDM and SimDAL.

Simulation Data Model, SimDM

SimDM provides a comprehensive description of the simulation data, i.e proprietary information, scientific objective, codes, physical processes modeled, the setup and parameter settings, raw data output information, post processing etc,.

The format of the metadata in SimDAL components are in the form of xml serializations of SimDM definitions of said metadata.
SimDM is represented as an ontology with a Unified Modeling Language (UML) diagram that visualizes the relationships between the different classes. The UML representation aids in understanding how different entities within the model relate to each other, ensuring clarity in the structure of the model. Some key classes at the top of the SimDM hierarchy include Simulator which defines the code or software used for the simulation, Simulation which is the main entity that encapsulates the details of a simulation, ParameterSet which describes the input parameters, Data products which links to the outputs generated from simulations (e.g., raw snapshot files, object catalogs, images). These SimDM entities are represented in figure 5.4.6; at the top of the hierarchy are Protocol and Experiment which represent the simulations and the suite-setup respectively. The Protocol (SimDM:/resource/protocol/Protocol) describes the configuration used to run any given experiment. Protocol encapsulates a set of algorithms, description of the input parameters (SimDM:/resource/protocol/Protocol.parameter), description of the output (SimDM:/resource/protocol/Protocol.outputType) etc,. Protocol is written for both simulations as well as postprocessing as presented in the diagram. Then each of the runs or postprocessing executed will instantiate the protocol used when describing their respective output data.

Modelling simulation metadata In this section the data entities from cosmological simulations, taking Dianoga as an example, are modelled into SimDM structures. Each SimDM object class hereafter is mentioned in the brackets along with its full type description when referencing for the first time.
Figure 5.4.7 is a top-level view of the instance diagram for Dianoga suite of simulations. The raw data output of Dianoga are the snapshots and catalog files available from the beginning of the simulations to redshift=0, at incremental time-steps. These are described by the OutputDataset (SimDM/resource/-experiment/OutputDataset). OutputDataset can represent both snapshot as well as outputs from postprocessing, thus the type of each is given by OutputDataset.objectType where the data is described along with its properties. The description of the objectType is clarified with OutputDataObjectType (SimDM/resource/protocol/OutputDataObjectType; a category of
OutputDataset.objectType). As an example, the snapshots are shown in the instance diagram, with type set as: `http://ivoa.net/rdf/theory/DataObject Types/20-19-02-27/DataObjectTypes.html#Snapshot`. This is a term which is maintained in the SKOS (Simple Knowledge Organization System) vocabulary which generically describes all output types which is a dump of the state of the given experiment at a given time. The snapshot contains all the particle information and the properties included are dependent on the type of the particle with information captured for a given redshift or time. For brevity, a complete list of all the particle properties is not included in the instance diagram.
Figure 5.4.8 represents one of the Dianoga simulation run (example named as D1-1x) with the protocol described in 5.4.7. Each of the simulations run with the Dianoga protocol instantiates the configuration input, algorithms used and

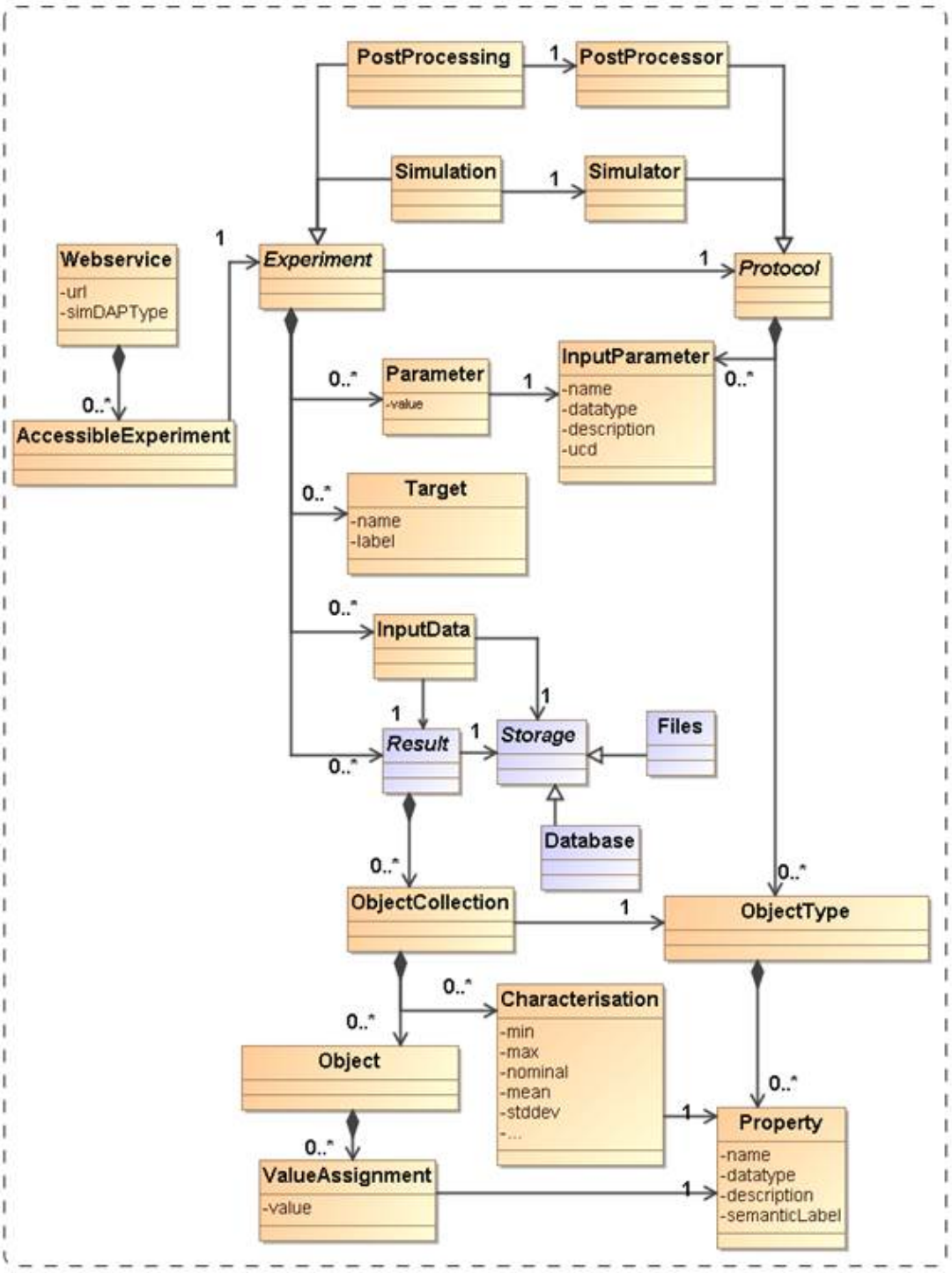

Figure 5.4.6: SimDM model classes, figure taken from SimDM specification, taken from: [87]

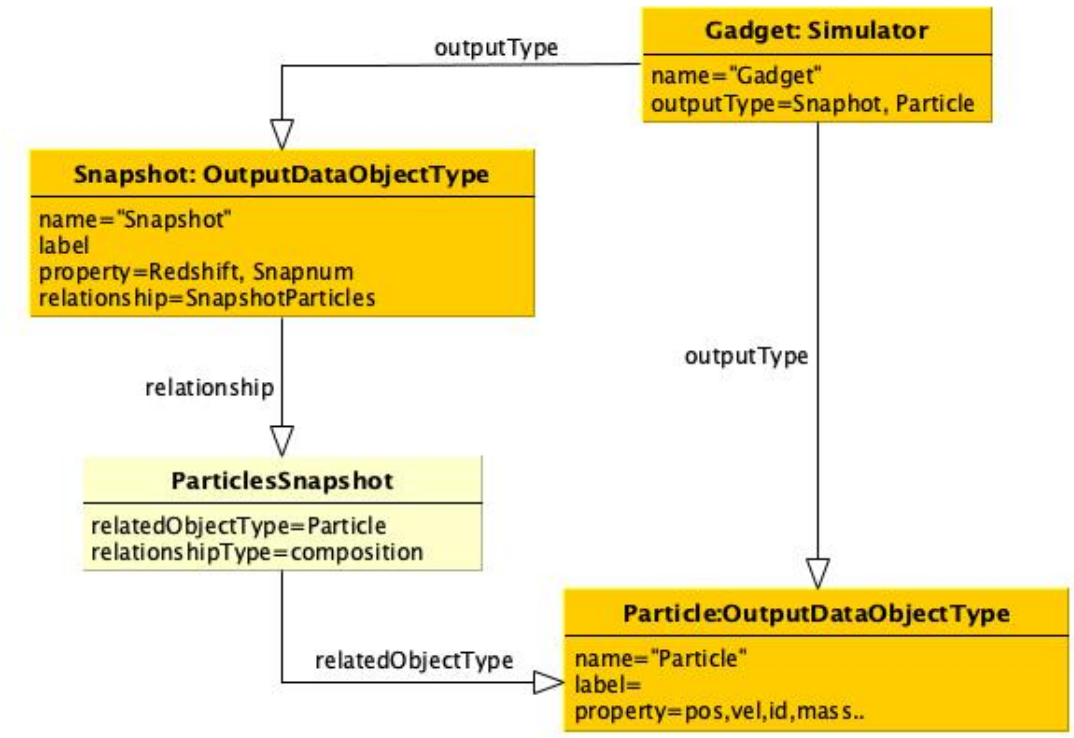

Figure 5.4.7: Instance diagram showing the output types of Dianoga simulation suite.

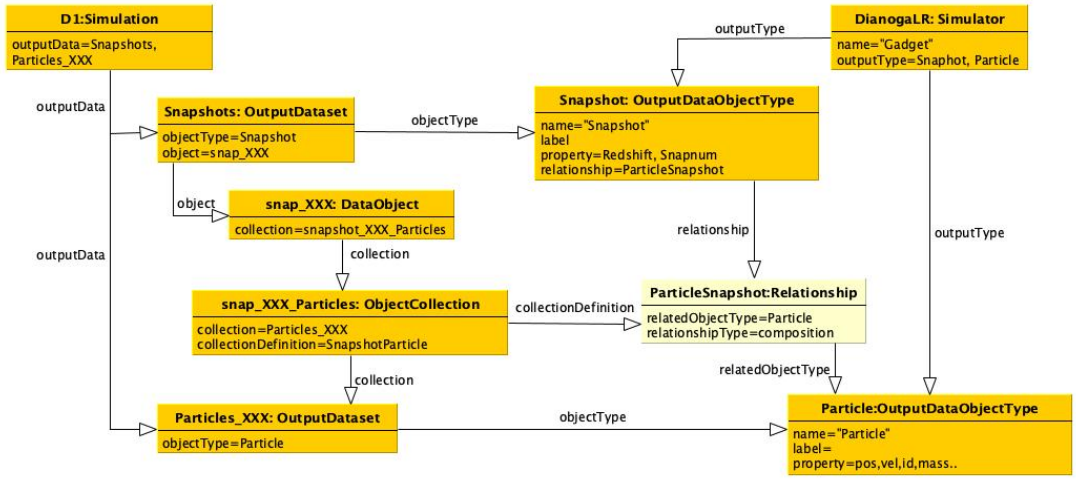

Figure 5.4.8: Instance diagram showing one of Dianoga runs referencing the output types of the suite shown in Figure 5.4.7.

the specific implementation of physical processes which are described in the protocol and included in the run. The outputTypes outlined in the referenced Protocol, which are the generated in each run need to be described. For D1-1x run, the raw outputs, i.e. snapshots produced are SimDM/resource/experiment/DataObject i.e. representing instances of the data which are the specific types of output (OutputDataObjectType) described in the protocol. On the other hand, the data can also be just presented in-terms of their statistical values (SimDM:/resource/experiment/StatisticalSummary) using SimDM classes. This is the way each particle information of D1-1x are described, considering that listing the values for all particles is simply not a feasible option for cosmological simulations. Thus each particle : gas, dark matter, boundary, star and blackhole particles are described with a statistical summary of their properties. All the particles are listed in collections of outputs (SimDM:/resource/experiment /ObjectCollection) and attached to the snapshot file metadata description (SimDM/resource/experiment/DataObject). The properties (SimDM/object/Property) of the snapshots are also instantiated, in case of Dianoga runs it is the information of the time at which the snapshot was generated during the simulation.
This modeling is illustrated in Figure 5.4.9 which is a snapshot of Simulator.xml file, prepared for the Dianoga suite of simulations. This file encapsulates key metadata about the simulation code, including the specific algorithms used, input parameters, and types of output that are generated as described by the respective Simdm classes. It provides a structured description of the simulator itself, detailing elements such as the software version, numerical methods, and the various parameter types, as well as the expected forms of output.
In Figure 5.4.10, we see the Simulation.xml file, representing one of the simulation runs from the Dianoga suite. This file directly references properties defined in the Simulator.xml, instantiating them for the specific run. For instance, it includes the actual values assigned to each parameter type defined in the simulator, the choice of algorithms used as well as the description of the data produced.

Simulation Data Access Layer, SimDAL

SimDAL protocols are established to facilitate the three phases of data retrieval: peruse, filter and access.
Figure 5.4.11 shows the 3 components of SimDAL which covers these phases:

- a registry which lists published simulations high level details.
- search component which is based on filtering for required values of parameters related to the metadata or statistics of the quantities of simulations. The user is to find IDs of the simulation datasets that they are interested in using search services, so that they can use the ID in the data retrieval phase.
- access: retrieving schemas, raw data or cutouts.

These protocols operate based on the standard formats described by the SimDM. To start with, I begin with the implementation of the data access component, which is detailed next.

```
<simulator xmlns:ns2="http://www.ivoa.net/xml/SimDM/v1.0/resource/experiment" xmlns:ns3="http://www.ivoa.net/xml/SimDM/v1.0/resource/protocol">
  <identity xmlId="gadgetId1"/>
  <name>GADGET</name>
  <code>
    https://ui.adsabs.harvard.edu/link_gateway/2005MNRAS.364.1105S/arxiv:astro-ph/0505010
  </code>
  <version>GADGET-3, Upgraded SPH scheme</version>
  <algorithm>
    <identity xmlId="palgo1"/>
    <name>Upgraded SPH</name>
    <description>Upgraded SPH scheme ref: Beck et al. 2015</description>
    <label>
      http://ivoa.net/rdf/theory/Algorithms#TreeSmoothParticleHydrodynamics
    </label>
  </algorithm>
  <algorithm>
    <identity xmlId="palgo2"/>
    <name>SF and SN feedback</name>
    <description>
      Star formation and feedback from a multiphase ISM, ref: Springle & Hernquist 2003, MNRAS, 339, 289
    </description>
    <label>http://ivoa.net/rdf/theory/Algorithms#Algorithm</label>
  </algorithm>
  <algorithm>
    <identity xmlId="palgo3"/>
    <name>Metal production</name>
    <description>
      Metal production from SN-II, SN-Ia and asymptotic-giantbranch (AGB) stars .ref: Tornatore et al. 2007, MNRAS, 382, 1050 and Planelles et al. 2014
    </description>
    <label>http://ivoa.net/rdf/theory/Algorithms#Algorithm</label>
  </algorithm>
  <algorithm>
    <identity xmlId="palgo4"/>
    <name>AGN feedback</name>
    <description>ref: Steinborn et al. 2015, MNRAS, 448, 1504</description>
    <label>http://ivoa.net/rdf/theory/Algorithms#Algorithm</label>
  </algorithm>
  <parameter>
    <identity xmlId="p1"/>
    <name>hubbleparam</name>
    <datatype>real</datatype>
    <description>Hubble Parameter</description>
    <label>
      http://ivoa.net/rdf/theory/PhysicalQuantities#HubbleParameter
    </label>
```

Figure 5.4.9: Part of Simulator.xml prepared for Dianoga suite of simulations

```
<ns2:simulation xmlns:ns2="http://www.ivoa.net/xml/SimDM/v1.0/resource/experiment" xmlns:ns3="http://www.ivoa.net/xml/SimDM/v1.0/resource/protocol">
  <name>D1</name>
  <description>D1 region, LR 1x run</description>
  <referenceURL>
    https://wwwmpa.mpa-garching.mpg.de/hydrosims/doku.php?id=dianoga_initial_conditions
  </referenceURL>
  <contact>
    <role>publisher</role>
    <party xmlId="21"/>
  </contact>
  <target xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:ns5="http://www.ivoa.net/xml/SimDM/v1.0/resource" xsi:type="ns5:TargetObjectType">
    <description>Clusters of galaxies</description>
    <multiplicity>1..*</multiplicity>
    <identityName>cl-g</identityName>
    <label>http://astrothesaurus.org/uat/584</label>
  </target>
  <parameter>
    <numericValue>
      <value>0.7200000286102295</value>
      <unit>km/s/Mpc</unit>
    </numericValue>
    <inputParameter xmlId="p1"/>
  </parameter>
  <parameter>
    <numericValue>
      <value>0.0399999910593033</value>
      <unit>kg/m3</unit>
    </numericValue>
    <inputParameter xmlId="p2"/>
  </parameter>
  <parameter>
    <numericValue>
      <value>0.7599999904632568</value>
      <unit>kg/m3</unit>
    </numericValue>
    <inputParameter xmlId="p3"/>
  </parameter>
  <parameter>
    <numericValue>
      <value>0.23999999463558197</value>
      <unit>kg/m3</unit>
    </numericValue>
    <inputParameter xmlId="p4"/>
  </parameter>
  <parameter>
```

Figure 5.4.10: Part of Simulation.xml prepared for one of Dianoga simulations run with the protocol of 5.4.9

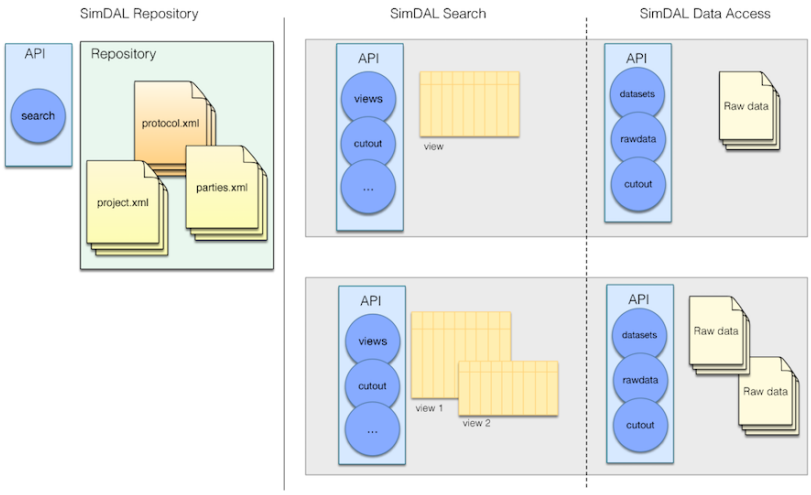

Figure 5.4.11: Simulation Data Access Layer components figure from: Version 1.0 [82]

Data access Users can make use of the existing simulations in many different ways depending on their specific needs. However the first fundamental step is to access the raw data; either download the data as is, or a part of the data. This essential part of data access is defined by SIMDAL's data access standard. The simulation outputs, described in the SimDM model as "OutputDatasets," are accessible through specific resources that enable users to query, retrieve, and cut out (select) sections of data based on defined axes of interest. To clarify with an example, if a Subfind catalog file is the dataset in question, then a query can be made to obtain only the velocity and mass of the most massive galaxies which is of interest to the user.
The Data Access component defines several types of resources to facilitate this functionality:

- datasets: This resource lists all available datasets, allowing clients to navigate the data catalog without a prior search step. It includes links to dataset metadata and schemas, helping users identify relevant datasets independently of other SimDAL services.
- cutouts and async/cutouts: Optional resources for retrieving specific data subsets. These are synchronous and asynchronous services for data cutouts along defined axes. This functionality is essential for users needing only certain parts of a large dataset, reducing data volume in transfer.
- rawdata: Allows direct downloading of raw data files, suitable for users needing complete datasets without any pre-processing or sub-selection.
- VOSI (Virtual Observatory Support Interface) resources: Include /availability and /capabilities endpoints to provide information on service status and capabilities.

Thus the datasets i.e the simulation outputs are provided along with an exhaustive description of the content i.e. a schema in the form of an xml seri-

alisation of all the properties made available for the given dataset. This allows the discovery of the associated services related to the dataset, i.e. rawdata, cutout which is required to finally access the data. The schema relating to the datasets provides information of the query parameters to be used to make the data access service requests. Along with these components, the XML serializations of the simulation protocol, the runs, the project and the party (SimDM:/resource/experiment/Simulation, SimDM:/resource/protocol/Simulator, SimDM:/resource/Project and SimDM:/resource/Party) need to be provided in the data access component. SimDM Party here represents individuals or organizations involved in a simulation project, essentially capturing the authorship, ownership, and responsibility for the simulation data. Meanwhile, the Project class defines the broader research initiative or context under which the simulations are conducted

I began implementing the SimDAL Data Access component by developing support for VOTable (to represent data and service communications) and UWS (Universal Worker Service). These foundational steps, already described earlier in this section, are preparing the ground for complete SimDAL Data Access functionality.

Once further progress is made in finalizing the SimDM serializations (described in the previous section) for simulation metadata, specifically for "simulator" and "simulation," I will proceed with completing the Data Access component. This will include ensuring that key XML serialization files—such as ‘project.xml‘, ‘protocol.xml‘, and ‘parties.xml‘—are made available via SimDAL.

5.4.3 Handling Metadata in Cosmological Simulations: Challenges and Considerations

In this section, I explore various challenges and considerations related to the implementation of SimDM and SimDAL standards for managing cosmological simulation data. While these standards provide a valuable framework, several aspects remain open for discussion and refinement, particularly when applied to large-scale simulations. Drawing on personal experience as a developer and previous discussions with VO authors, this section outlines key points that require further dialogue to ensure the standards evolve in a way that effectively supports the unique needs of the cosmological simulation community

1. In designing the simulations archive, I adopted an approach aligned with the SimDB concept, leveraging database structures for long-term metadata management. However, challenges arose when referring to the IVOA implementation note for SimDM and SimDAL, which outlines implementation using VO-URP and MagicDraw. These tools are outdated and not maintained. This led to wasted time and efforts attempting to implement an obsolete method. Such issues highlight a critical shortcoming in the SimDM and SimDAL ecosystem: the lack of clearly marked, up-to-date, and practical implementation guidance. Addressing these gaps is essential to prevent developers from encountering similar obstacles and to promote broader adoption of these standards.

2. Handling collections of large, distributed snapshots and post-processing file —common in cosmological simulations —remains a significant challenge under current SimDAL specifications. For example, raw data download of objects spanning multiple distributed files is not addressed. These challenges also intersect with SimDM, where each distributed snapshot file would seemingly require a separate metadata file. However, since these files are distributed based on domain decomposition and snapshots represent volumes following a space-filling curve, it is unclear how metadata-defined boundaries for individual files would enable effective cutouts. These gaps highlight the need for enhanced SimDAL and SimDM standards to better support the unique requirements of large-scale cosmological simulations.

3. SimDAL cutout services are structured with URLs like: `http://<simdal-data-uri>/DM51NoPAH_20_cloud/async/cutouts/?format=votable&RUNID=j1`.This approach requires the service URL to be explicitly tied to a specific dataset, which imposes an unnecessary constraint on developers. This rigidity creates an avoidable hindrance without providing any clear benefits, as such implementation details should ideally be left to the discretion of developers. Furthermore, this limitation restricts extending the service to valuable use cases, such as querying across multiple redshifts to retrieve progenitor data —a capability that would significantly enhance its utility.

4. The use of SKOS in SimDM presents challenges, as it is not actively maintained. Several terms are missing, and there is ambiguity about which terms to use in certain contexts.

5. The volume of metadata generated by large cosmological simulations with numerous snapshots is something worth considering. While creating SimDM metadata files for all snapshots could enable high-level querying, it is uncertain whether the payoff justifies the effort, particularly for complex cutout services. This question warrants further discussion with domain experts and VO authors to assess its feasibility and value. For example, cutout services centered on galaxies —requiring precise object positions derived from Subfind and FoF post-processing -highlight the potential utility of combining raw data for user-defined queries. Addressing such needs would significantly enhance usability for the cosmological simulations community, suggesting a broader discussion is needed to identify and develop services that align with domain-specific requirements.

It is important to note that the points discussed in this section are not intended as recommendations or critique. Instead, these are areas that warrant further dialogue with the VO community and the authors, as highlighted in previous interactions on points 2 and 4. As a developer, rather than a standards expert, I am not in a position to define these specifications. The development of standards is a complex, iterative process that often extends beyond the scope of a PhD timeline. Had the standards been fully mature and applicable, a more comprehensive implementation could have been pursued. However, due to the constraints of a

research deadline, it was not feasible to participate in working groups. The next step, as discussed with the VO authors, was to create SimDM templates, followed by collaborative discussions to identify and address gaps from both sides. This process will help refine and create a more practical and effective implementation for managing simulation data.

Chapter 6

Summary and final conclusions

In this thesis, I have explored two distinct topics, each addressing a critical aspect of simulation research and data management. The first part focused on quantifying chaotic variations in galaxy cluster simulations. The second part tackled the challenge of developing a data-sharing infrastructure for simulations, aligning with Virtual Observatory (VO) standards to enhance data accessibility and interoperability. In this final chapter for each topic, I conclude with summary of my work, the implications of the results for the field as well as outlining the directions for future work.

6.1 Chaotic effects in cosmological hydrodynamical simulations

Understanding and quantifying chaotic effects in galaxy cluster simulations are essential because stochastic variations can significantly impact our interpretation of astrophysical phenomena and the reliability of simulation results. Such variations, often resulting from small-scale physical processes, numerical approximations, and sensitivity to initial conditions, can influence individual galaxy properties and affect various observables. Assessing the inherent uncertainties due to these chaotic effects is crucial for determining the confidence we can place in simulated data and for distinguishing true physical trends from consequences of the system's chaotic behavior.

In this study, I investigated the role of numerical chaos in galaxy cluster simulations using the open-source OpenGadget3 code. I executed four identical simulations at the lowest resolution setting (1x) for one of the Dianoga galaxy clusters on the HotCat cluster. Individual galaxies were matched across these runs, and variations in their properties—such as stellar and dark matter mass—were analyzed using two statistical methods: pairwise differences (Method 2) and a mixed linear model (Method 1). This approach both quantifies the expected variability and also helps validate the results.

The primary objective was to establish a baseline measure of chaotic variation in simulations run with OpenGadget3. Analyzing the set of four identical simulations without black holes (the *Fiducial* model), the results reflect that the

ensemble averages of galaxy properties—such as the galaxy stellar mass function and star formation rate—are faithfully reproduced despite underlying chaotic dynamics (Figures 4.2.4 and 4.2.5). The variations in dark matter mass and stellar mass between matched Brightest Cluster Galaxies (BCGs) remained relatively small, with stellar mass variation ($\sigma_{M_*} = 0.04$–0.08) slightly higher than dark matter mass variation ($\sigma_{M_{\rm dm}} = 0.03$–$0.06$) (Figures 4.2.7 and 4.2.8). This increased variation in stellar mass is attributed to shot noise due to the lower number of stellar particles and the influence of highly non-linear processes like star formation and feedback.

My analysis showed that the run-to-run variation is minimal ($\sigma < 0.004$), 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 chaotic divergence, and the observed variations are primarily due to stochastic noise and shot noise rather than chaos amplifying differences between runs.

To explore the interplay between chaos and stellar feedback, I conducted feedback tests by varying the stellar wind velocity to create low, high, and zero feedback scenarios (Table 4.1.2). The results indicate that changing the stellar feedback has a negligible effect on the variation in dark matter mass (Figure 4.2.11) but significantly impacts the variation in stellar mass (Figure 4.2.12). Specifically, higher stellar feedback increases the stochastic variation in stellar mass within each simulation. This increased variation is attributed to heightened stochastic noise arising from the localized and highly non-linear nature of stellar feedback processes (Figures 4.2.14 and 4.2.15).

Simulations with lower feedback levels displayed reduced stochastic noise, potentially linked to smooth and high star formation rates and consequent gas exhaustion. This observation aligns with findings from Keller et al. (2019), who suggested that unregulated star formation can dampen variations due to heightened star formation efficiency, although this effect might not be as visible in a cosmological environment. These results can be particular interest especially incorporating blackholes in the simulation. Blackholes also compete for gas and could disrupt this dynamic of very high unregulated star formation and dampened variation.

Black holes play a pivotal role in shaping feedback processes, often amplifying chaotic effects in simulations. For instance, Genel et al. (2019) observed that feedback enhances chaos, but this result may partially arise from the absence of black holes in their no-feedback simulations, while full-physics runs included them. This underscores the difficulty of disentangling the contributions of different sub-grid physics from those of related to feedback —an undertaking that demands careful, focused analysis.

Incorporating black holes, particularly within the framework of OPENGADGET3, would have introduced additional layers of complexity, including challenges such as the wandering black hole problem and the substantial variability they introduce. These factors could obscure the core findings by conflating the sources of variation. As such, the decision to run the initial set of simulations without black holes was a deliberate choice, aimed at establishing a baseline of variability unrelated to their influence. The inclusion of black holes, to allow a

meaningful comparison, is a natural next step and is already underway. However, due to the constraints of the project timeline this aspect was not incorporated within the scope of this thesis.
I also examined the variation in stellar mass across different galaxy mass bins. For lower stellar-mass galaxies ($10 < \log M_(M_\odot/h) < 11$), higher stellar feedback leads to increased stochastic variation (Figures 4.2.18 and 4.2.19). In contrast, for higher stellar-mass galaxies ($11 < \log M_(M_\odot/h) < 11.5$), the effect of stellar feedback on stochastic variation is negligible (Figure 4.2.20). This is evidently linked to the deferential impact of stellar feedback for galaxies belonging to relatively lower mass bins and thereby induce variations. Galaxies in the highest mass bins, are relatively unaffected, showing no significant trends in stochastic variation with respect to stellar feedback. In these massive galaxies, AGN feedback may play a more prominent role, potentially exhibiting different variation patterns when included in the 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 generalizability of these 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.
In conclusion, this study provides a benchmark for the expected magnitude of variation due to chaotic effects in simulations run with OpenGadget3. The results highlight the importance of considering stochastic variations and shot noise when interpreting simulation outcomes, especially for properties sensitive to highly non-linear processes like star formation and feedback. Understanding these variations is crucial for accurately comparing simulation results with observational data and improving the reliability of predictions made by cosmological models.
Future work should extend this analysis to higher-resolution simulations and include the effects of black hole feedback to fully disentangle the complex interplay between chaos and various feedback mechanisms in galaxy formation. Additionally, exploring a broader range of subgrid models and feedback parameters will help clarify the conditions under which feedback mitigates or exacerbates chaotic behavior, thereby enhancing our understanding and interpretation of results of galaxy evolution in cosmological simulations.

6.2 Development of a cosmological hydrodynamical simulations archive

The importance of sharing simulation data is becoming increasingly clear as simulations play a pivotal role in astrophysics and cosmology. However, accessing these simulations remains a significant challenge. Regularizing access to this data not only facilitates scientific collaboration and reproducibility but also enables re-

searchers to perform independent analyses, verify results, and build on existing work. The Virtual Observatory (VO) standards and FAIR (**F**indable, **A**ccessible, **I**nteroperable, and **R**eusable) principles are critical in making simulation data universally accessible and interpretable. These standards ensure that data from diverse sources can be shared, cross-compared, and utilized without extensive preprocessing, fostering a collaborative and transparent scientific environment. The motivation behind this work —developing a data archive for simulations -stems from this need for efficient, standardized data sharing within the astrophysics community.

In this part of my PhD work, I address this critical need to create an open, curated, and standardized simulation data archive. To start with, I developed a custom database and user interface. This foundational work provided valuable insights into the data requirements, best practices, and common challenges associated with organizing simulation data. Using the "Dianoga" suite of hydrodynamical simulations as a case study, I organized the simulation data into a relational database, capturing essential metadata and structuring the data to facilitate user access and discovery. The web interface, built using the Django framework, provides users with an intuitive platform to explore, visualize, and retrieve simulation data (see Figure 5.3.1). This initial implementation facilitates the understanding of the complexities involved in managing simulation data and laid the groundwork for future enhancements.

To enhance the robustness and portability of the archive, I employed containerization technologies, specifically Docker, to streamline deployment and ensure consistent performance across different computing environments. This approach facilitates scalability and eases maintenance, making the archive more sustainable in the long term.

Building on this foundation, I began implementing several key VO data standards, such as, VOTable for data representation and the Universal Worker Service (UWS) protocol for managing asynchronous service requests. Additionally, I developed preliminary templates for the Simulation Data Model (SimDM), the VO's model for describing simulation metadata. These implementations ensure that the archive can effectively interface with other VO-compliant services and tools, enhancing interoperability.

However, during implementation, I encountered challenges due to ambiguities in certain VO simulation standards. While SimDM is designed to describe simulation metadata, it lacks clarity when applied to large cosmological datasets. Furthermore, it is not entirely clear how SimDM and SimDAL can manage large, distributed files typical of simulation outputs, such as multiple snapshot and post-processing files. Managing these collections effectively within a VO-compliant framework presents a technical and organizational challenge. Another area of concern is the SKOS vocabulary used within SimDM; it has not been consistently maintained, leading to gaps and ambiguities in metadata representation that complicate data encoding.

To address these issues, I will proceed as a collaborative effort with the authors of the VO standards. Together, we aim to identify and address gaps in the standards, particularly focusing on developing SimDM templates tailored to

representative simulations like the Dianoga runs. This joint approach enables to tackle challenges from both developer and standards perspectives, paving the way for a more robust, flexible, and comprehensive data-sharing framework. My efforts will not only support the specific needs of this project but also contribute to the broader goal of improving VO standards for simulation data, ultimately enhancing the accessibility and usability of simulation data archives.

Overall, this work represents a foundational step towards creating a comprehensive, standardized archive for cosmological hydrodynamical simulations. By adhering to FAIR principles and IVOA standards, I have established a framework that not only preserves valuable simulation data but also makes it accessible and useful to the broader scientific community.

Future work will involve completing the implementation of SimDAL, refining metadata descriptions, and integrating additional simulations into the archive. I also anticipate expanding the archive's capabilities by incorporating advanced data analysis and visualization tools, further enhancing its utility for researchers aiming to compare simulations with observational data or conduct new analyses. Additionally, ongoing collaboration with the VO standards community will be crucial to address challenges specific to cosmological simulations, such as handling large, distributed datasets and refining metadata standards to capture the richness of simulation outputs.

For a given simulation code, exploring variations in resolution settings, subgrid physics, and configuration parameters (such as numerical precision, MPI task allocation, and random number generators) is crucial for defining the error bars associated with different simulation setups. Understanding these error bars often requires running multiple simulations to assess the inherent variability and chaotic effects introduced by these factors. In this context, the proposed cosmological simulations archive could play a pivotal role by serving as a repository for repeated run simulation-suites for different codes, configurations, and parameter spaces. By hosting such simulations alongside metadata detailing their variability and chaotic properties, the archive could become a foundational resource for researchers. It would allow future studies to build on established knowledge, better interpret their own results, and understand the inherent variability in their simulations. To maximize its utility, simulation data standards may need to evolve to include detailed setup parameters and enable advanced querying across multiple simulations. Such advancements would not only support reproducibility but also provide a robust framework for studying and mitigating chaotic effects. While current standards may not fully support these capabilities, their development represents a promising direction for future research.

Acknowledgements

Firstly, I would like to thank Giuliano Taffoni, Stefano Borgani, Gianluigi Granato for providing me with this project opportunity, ensuring all resources and tools for my work and mainly for agreeing to guide me. I am immensely grateful to my supervisor, Giuliano Taffoni in particular, for his mentorship, support, and continuous encouragement.
I would also like to thank the Trieste cosmological hydrodynamical simulations team, specifically, Guiseppe Murante, Milena Valentini, David Goz as well as Cinthia Ragone Figueroa, Francesco Pauli and Antonio Raganin for their invaluable input during critical discussions, often on short notice, which has immensely helped my research. In particular, I owe a ton of gratitude to Antonio Raganin, who has guided me whenever I needed throughout my PhD —practically teaching me "how to fish," so to speak.
I would like to express my sincere gratitude to the reviewers, Mario Spera and Claudio Gheller for their valuable feedback and thorough consideration of the thesis, which greatly contributed to the improvement of this work.
Special thanks to Francois Bonnarel, Mirelle Louys, Marco Molinaro and Sara Bertocco for their kindness and support in helping me get here. Finally, words cannot express my love and appreciation for my family and friends, both the furry and less-furry ones; I am deeply grateful to them.

List of Figures

Bibliography

[1] S. J. Aarseth. "Dynamical evolution of clusters of galaxies, I". In: *Monthly Notices of the Royal Astronomical Society* 126 (Jan. 1, 1963). Publisher: OUP ADS Bibcode: 1963MNRAS.126..223A, p. 223. ISSN: 0035-8711. DOI: 10.1093/mnras/126.3.223. URL: https://ui.adsabs.harvard.edu/abs/1963MNRAS.126..223A (visited on 08/01/2024).

[2] Steven W. Allen, August E. Evrard, and Adam B. Mantz. "Cosmological Parameters from Observations of Galaxy Clusters". In: *Annual Review of Astronomy and Astrophysics* 49.1 (Sept. 22, 2011), pp. 409–470. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev-astro-081710-102514. arXiv: 1103.4829[astro-ph]. URL: http://arxiv.org/abs/1103.4829 (visited on 10/28/2024).

[3] Raul E. Angulo and Simon D. M. White. "One simulation to fit them all - changing the background parameters of a cosmological N-body simulation". In: *Monthly Notices of the Royal Astronomical Society* (Apr. 2010). ISSN: 00358711, 13652966. DOI: 10.1111/j.1365-2966.2010.16459.x. arXiv: 0912.4277[astro-ph]. URL: http://arxiv.org/abs/0912.4277 (visited on 08/24/2024).

[4] P. Anninos and M. L. Norman. *Nested grid methods for cosmological hydrodynamic and N-body systems.* Pages: 24 Publication Title: Numerical Simulations in Astrophysics ADS Bibcode: 1994nsa..book...24A. Jan. 1, 1994. URL: https://ui.adsabs.harvard.edu/abs/1994nsa..book...24A (visited on 07/29/2024).

[5] J. H. Applegate et al. "The outer solar system for 200 million years." In: *The Astronomical Journal* 92 (July 1, 1986). Publisher: IOP ADS Bibcode: 1986AJ.....92..176A, pp. 176–194. ISSN: 0004-6256. DOI: 10.1086/114149. URL: https://ui.adsabs.harvard.edu/abs/1986AJ.....92..176A (visited on 08/13/2024).

[6] Ronald-Louis G Ballouz. "NUMERICAL SIMULATIONS OF GRANULAR PHYSICS IN THE SOLAR SYSTEM". In: ().

[7] J. M. Bardeen et al. "The Statistics of Peaks of Gaussian Random Fields". In: *The Astrophysical Journal* 304 (May 1, 1986). Publisher: IOP ADS Bibcode: 1986ApJ...304...15B, p. 15. ISSN: 0004-637X. DOI: 10.1086/164143. URL: https://ui.adsabs.harvard.edu/abs/1986ApJ...304...15B (visited on 10/11/2024).

[8] Josh Barnes and Piet Hut. "A hierarchical O(N log N) force-calculation algorithm". In: *Nature* 324 (Dec. 1, 1986). ADS Bibcode: 1986Natur.324..446B, pp. 446–449. ISSN: 0028-0836. DOI: 10.1038/324446a0. URL: https://ui.adsabs.harvard.edu/abs/1986Natur.324..446B (visited on 07/29/2024).

[9] L. Bassini et al. "Black hole mass of central galaxies and cluster mass correlation in cosmological hydro-dynamical simulations". In: *Astronomy & Astrophysics* 630 (Oct. 1, 2019). Publisher: EDP Sciences, A144. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201935383. URL: https://www.aanda.org/articles/aa/abs/2019/10/aa35383-19/aa35383-19.html (visited on 08/20/2024).

[10] L. Bassini et al. "The DIANOGA simulations of galaxy clusters: characterising star formation in protoclusters". In: *Astronomy & Astrophysics* 642 (Oct. 1, 2020). Publisher: EDP Sciences, A37. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202038396. URL: https://www.aanda.org/articles/aa/abs/2020/10/aa38396-20/aa38396-20.html (visited on 08/20/2024).

[11] M. R. Bate, G. Lodato, and J. E. Pringle. "Chaotic star formation and the alignment of stellar rotation with disc and planetary orbital axes". In: *Monthly Notices of the Royal Astronomical Society* 401.3 (Jan. 21, 2010), pp. 1505–1513. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2009.15773.x. URL: https://doi.org/10.1111/j.1365-2966.2009.15773.x (visited on 08/17/2024).

[12] Konstantin Batygin and Gregory Laughlin. *On the Dynamical Stability of the Solar System*. Apr. 11, 2008. DOI: 10.48550/arXiv.0804.1946. arXiv: 0804.1946. URL: http://arxiv.org/abs/0804.1946 (visited on 10/12/2024).

[13] Konstantin Batygin, Alessandro Morbidelli, and Matthew J. Holman. "CHAOTIC DISINTEGRATION OF THE INNER SOLAR SYSTEM". In: *The Astrophysical Journal* 799.2 (Jan. 2015). Publisher: The American Astronomical Society, p. 120. ISSN: 0004-637X. DOI: 10.1088/0004-637X/799/2/120. URL: https://dx.doi.org/10.1088/0004-637X/799/2/120 (visited on 08/13/2024).

[14] D. Benhaiem et al. "Particle number dependence in the non-linear evolution of N-body self-gravitating systems". In: *Monthly Notices of the Royal Astronomical Society* 473.2 (Jan. 11, 2018), pp. 2348–2354. ISSN: 0035-8711. DOI: 10.1093/mnras/stx2444. URL: https://doi.org/10.1093/mnras/stx2444 (visited on 08/14/2024).

[15] F. Bernardeau et al. *Large-Scale Structure of the Universe and Cosmological Perturbation Theory*. Dec. 27, 2001. DOI: 10.48550/arXiv.astro-ph/0112551. arXiv: astro-ph/0112551. URL: http://arxiv.org/abs/astro-ph/0112551 (visited on 10/11/2024).

[16] M. Bernardi et al. “The massive end of the luminosity and stellar mass functions: dependence on the fit to the light profile”. In: *Monthly Notices of the Royal Astronomical Society* 436.1 (Nov. 21, 2013), pp. 697–704. ISSN: 0035-8711. DOI: 10.1093/mnras/stt1607. URL: https://doi.org/10.1093/mnras/stt1607 (visited on 10/26/2024).

[17] S Bertocco et al. “INAF Trieste Astronomical Observatory Information Technology Framework”. In: (2020).

[18] V Biffi et al. “The origin of ICM enrichment in the outskirts of present-day galaxy clusters from cosmological hydrodynamical simulations”. In: *Monthly Notices of the Royal Astronomical Society* 476.2 (May 11, 2018), pp. 2689–2703. ISSN: 0035-8711. DOI: 10.1093/mnras/sty363. URL: https://doi.org/10.1093/mnras/sty363 (visited on 08/20/2024).

[19] V. Biffi et al. “The history of chemical enrichment in the intracluster medium from cosmological simulations”. In: *Monthly Notices of the Royal Astronomical Society* 468.1 (June 11, 2017), pp. 531–548. ISSN: 0035-8711. DOI: 10.1093/mnras/stx444. URL: https://doi.org/10.1093/mnras/stx444 (visited on 08/20/2024).

[20] A. Bonafede et al. “A non-ideal magnetohydrodynamic GADGET: simulating massive galaxy clusters”. In: *Monthly Notices of the Royal Astronomical Society* 418 (Dec. 1, 2011). ADS Bibcode: 2011MNRAS.418.2234B, pp. 2234–2250. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2011.19523.x. URL: https://ui.adsabs.harvard.edu/abs/2011MNRAS.418.2234B (visited on 08/10/2022).

[21] H. Bondi. “On Spherically Symmetrical Accretion”. In: *Monthly Notices of the Royal Astronomical Society* 112.2 (Apr. 1, 1952), pp. 195–204. ISSN: 0035-8711. DOI: 10.1093/mnras/112.2.195. URL: https://doi.org/10.1093/mnras/112.2.195 (visited on 10/11/2024).

[22] Josh Borrow et al. “The impact of stochastic modelling on the predictive power of galaxy formation simulations”. In: *Monthly Notices of the Royal Astronomical Society* 526.2 (Dec. 1, 2023), pp. 2441–2457. ISSN: 0035-8711. DOI: 10.1093/mnras/stad2928. URL: https://doi.org/10.1093/mnras/stad2928 (visited on 08/17/2024).

[23] Sownak Bose et al. “No cores in dark matter-dominated dwarf galaxies with bursty star formation histories”. In: *Monthly Notices of the Royal Astronomical Society* 486.4 (July 11, 2019), pp. 4790–4804. ISSN: 0035-8711. DOI: 10.1093/mnras/stz1168. URL: https://doi.org/10.1093/mnras/stz1168 (visited on 08/24/2024).

[24] F. R. Bouchet et al. “Perturbative Lagrangian approach to gravitational instability.” In: *Astronomy and Astrophysics* 296 (Apr. 1, 1995). ADS Bibcode: 1995A&A...296..575B, p. 575. ISSN: 0004-6361. DOI: 10.48550/arXiv.astro-ph/9406013. URL: https://ui.adsabs.harvard.edu/abs/1995A&A...296..575B (visited on 10/11/2024).

[25] G. L. Bryan and M. L. Norman. *A Hybrid AMR Application for Cosmology and Astrophysics*. Oct. 16, 1997. DOI: 10.48550/arXiv.astro-ph/9710187. arXiv: astro-ph/9710187. URL: http://arxiv.org/abs/astro-ph/9710187 (visited on 07/29/2024).

[26] R. Cen. "Hydrodynamic Simulations of Cosmological Systems". In: 22 (Sept. 1, 1990). Conference Name: Bulletin of the American Astronomical Society ADS Bibcode: 1990BAAS...22.1305C, p. 1305. URL: https://ui.adsabs.harvard.edu/abs/1990BAAS...22.1305C (visited on 07/29/2024).

[27] Planck Collaboration et al. "Planck 2018 results. VI. Cosmological parameters". In: *Astronomy & Astrophysics* 641 (Sept. 2020), A6. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209[astro-ph]. URL: http://arxiv.org/abs/1807.06209 (visited on 07/26/2024).

[28] Jonathan J Davies, Robert A Crain, and Andrew Pontzen. "Quenching and morphological evolution due to circumgalactic gas expulsion in a simulated galaxy with a controlled assembly history". In: *Monthly Notices of the Royal Astronomical Society* 501.1 (Feb. 11, 2021), pp. 236–253. ISSN: 0035-8711. DOI: 10.1093/mnras/staa3643. URL: https://doi.org/10.1093/mnras/staa3643 (visited on 04/09/2024).

[29] Jonathan J. Davies, Andrew Pontzen, and Robert A. Crain. *Are the fates of supermassive black holes and galaxies determined by individual mergers, or by the properties of their host haloes?* Nov. 6, 2023. arXiv: 2301.04145[astro-ph]. URL: http://arxiv.org/abs/2301.04145 (visited on 08/17/2024).

[30] Jonathan J. Davies, Andrew Pontzen, and Robert A. Crain. "Galaxy mergers can initiate quenching by unlocking an AGN-driven transformation of the baryon cycle". In: *Monthly Notices of the Royal Astronomical Society* 515.1 (July 22, 2022), pp. 1430–1443. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stac1742. arXiv: 2203.08157[astro-ph]. URL: http://arxiv.org/abs/2203.08157 (visited on 08/17/2024).

[31] M. Davis et al. "The evolution of large-scale structure in a universe dominated by cold dark matter". In: *The Astrophysical Journal* 292 (May 1, 1985). Publisher: IOP ADS Bibcode: 1985ApJ...292..371D, pp. 371–394. ISSN: 0004-637X. DOI: 10.1086/163168. URL: https://ui.adsabs.harvard.edu/abs/1985ApJ...292..371D (visited on 10/11/2024).

[32] E. J. Delgado-Donate, C. J. Clarke, and M. R. Bate. "The dependence of the substellar initial mass function on the initial conditions for star formation". In: *Monthly Notices of the Royal Astronomical Society* 347.3 (Jan. 21, 2004), pp. 759–770. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2004.07259.x. URL: https://doi.org/10.1111/j.1365-2966.2004.07259.x (visited on 08/17/2024).

[33] Sébastien Derriere et al. *IVOA Recommendation: An IVOA Standard for Unified Content Descriptors Version 1.1*. Oct. 3, 2011. arXiv: 1110.0525. URL: http://arxiv.org/abs/1110.0525 (visited on 10/17/2024).

[34] K. Dolag et al. "Substructures in hydrodynamical cluster simulations". In: *Monthly Notices of the Royal Astronomical Society* 399.2 (Oct. 21, 2009), pp. 497–514. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2009.15034.x. URL: https://doi.org/10.1111/j.1365-2966.2009.15034.x (visited on 10/11/2024).

[35] Theresa Dower et al. *Registry Interfaces Version 1.1*. DOI: 10.5479/ADS/bib/2018ivoa.spec.0723D. URL: http://ui.adsabs.harvard.edu/abs/2018ivoa.spec.0723D/abstract (visited on 10/17/2024).

[36] Patrick Dowler et al. *IVOA Simple Image Access Version 2.0*. DOI: 10.5479/ADS/bib/2015ivoa.spec.1223D. URL: http://ui.adsabs.harvard.edu/abs/2015ivoa.spec.1223D/abstract (visited on 10/17/2024).

[37] Patrick Dowler et al. *Table Access Protocol Version 1.1*. Aug. 26, 2019.

[38] John Dubinski. "A parallel tree code". In: *New Astronomy* 1.2 (Oct. 1, 1996), pp. 133–147. ISSN: 1384-1076. DOI: 10.1016/S1384-1076(96)00009-7. URL: https://www.sciencedirect.com/science/article/pii/S1384107696000097 (visited on 10/28/2024).

[39] Martin J. Duncan and Thomas Quinn. "The long-term dynamical evolution of the solar system." In: *Annual Review of Astronomy and Astrophysics* 31 (Jan. 1, 1993). ADS Bibcode: 1993ARA&A..31..265D, pp. 265–295. ISSN: 0066-4146. DOI: 10.1146/annurev.aa.31.090193.001405. URL: https://ui.adsabs.harvard.edu/abs/1993ARA&A..31..265D (visited on 08/13/2024).

[40] A. Einstein. "Die Grundlage der allgemeinen Relativitätstheorie". In: *Annalen der Physik* 354 (Jan. 1, 1916). ADS Bibcode: 1916AnP...354..769E, pp. 769–822. ISSN: 0003-3804. DOI: 10.1002/andp.19163540702. URL: https://ui.adsabs.harvard.edu/abs/1916AnP...354..769E (visited on 07/29/2024).

[41] A. E. Evrard et al. *Galaxy Clusters in Hubble Volume Simulations: Cosmological Constraints from Sky Survey Populations*. Mar. 12, 2002. DOI: 10.48550/arXiv.astro-ph/0110246. arXiv: astro-ph/0110246. URL: http://arxiv.org/abs/astro-ph/0110246 (visited on 10/22/2024).

[42] Christoph Federrath and Ralf S. Klessen. "THE STAR FORMATION RATE OF TURBULENT MAGNETIZED CLOUDS: COMPARING THEORY, SIMULATIONS, AND OBSERVATIONS". In: *The Astrophysical Journal* 761.2 (Dec. 2012). Publisher: The American Astronomical Society, p. 156. ISSN: 0004-637X. DOI: 10.1088/0004-637X/761/2/156. URL: https://dx.doi.org/10.1088/0004-637X/761/2/156 (visited on 08/17/2024).

[43] Mitchell Feigenbaum. “Quantitative universality for a class of nonlinear transformations”. In: *J. Stat. Phys.* 19 (July 1, 1978), pp. 25–52. DOI: 10.1007/BF01020332.

[44] A. Friedmann. “Über die Krümmung des Raumes”. In: *Zeitschrift fur Physik* 10 (Jan. 1, 1922). Publisher: Springer ADS Bibcode: 1922ZPhy...10..377F, pp. 377–386. DOI: 10.1007/BF01332580. URL: https://ui.adsabs.harvard.edu/abs/1922ZPhy...10..377F (visited on 07/29/2024).

[45] R. Fux. “Order and chaos in the local disc stellar kinematics induced by the Galactic bar”. In: *Astronomy & Astrophysics* 373.2 (July 1, 2001). Number: 2 Publisher: EDP Sciences, pp. 511–535. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361:20010561. URL: https://www.aanda.org/articles/aa/abs/2001/26/aa1098/aa1098.html (visited on 08/17/2024).

[46] Sam Geen et al. “On the indeterministic nature of star formation on the cloud scale”. In: *Monthly Notices of the Royal Astronomical Society* 481.2 (Dec. 1, 2018), pp. 2548–2569. ISSN: 0035-8711. DOI: 10.1093/mnras/sty2439. URL: https://doi.org/10.1093/mnras/sty2439 (visited on 08/17/2024).

[47] Andrew Gelman. “Data Analysis Using Regression And Multilevel/Hierarchical Models”. In: *Cambridge Universty Press*. Vol. 3. Journal Abbreviation: Cambridge Universty Press. Nov. 30, 2006. ISBN: 978-0-521-86706-1. DOI: 10.1017/CBO9780511790942.

[48] Shy Genel et al. “A Quantification of the Butterfly Effect in Cosmological Simulations and Implications for Galaxy Scaling Relations”. In: *The Astrophysical Journal* 871.1 (Jan. 2019). Publisher: The American Astronomical Society, p. 21. ISSN: 0004-637X. DOI: 10.3847/1538-4357/aaf4bb. URL: https://dx.doi.org/10.3847/1538-4357/aaf4bb (visited on 08/09/2024).

[49] Facundo A. Gómez et al. “Streams in the Aquarius stellar haloes”. In: *Monthly Notices of the Royal Astronomical Society* 436.4 (Dec. 21, 2013), pp. 3602–3613. ISSN: 0035-8711. DOI: 10.1093/mnras/stt1838. URL: https://doi.org/10.1093/mnras/stt1838 (visited on 08/17/2024).

[50] Jeremy Goodman, Douglas C. Heggie, and Piet Hut. “On the Exponential Instability of N-Body Systems”. In: *The Astrophysical Journal* 415 (Oct. 1, 1993). Publisher: IOP ADS Bibcode: 1993ApJ...415..715G, p. 715. ISSN: 0004-637X. DOI: 10.1086/173196. URL: https://ui.adsabs.harvard.edu/abs/1993ApJ...415..715G (visited on 08/12/2024).

[51] L Greengard and V Rokhlin. “A fast algorithm for particle simulations”. In: *Journal of Computational Physics* 73.2 (Dec. 1, 1987), pp. 325–348. ISSN: 0021-9991. DOI: 10.1016/0021-9991(87)90140-9. URL: https://www.sciencedirect.com/science/article/pii/0021999187901409 (visited on 08/13/2024).

[52] Frederick Groth et al. “The cosmological simulation code OPENGADGET3 - implementation of meshless finite mass”. In: *Monthly Notices of the Royal Astronomical Society* 526 (Nov. 1, 2023). Publisher: OUP ADS Bibcode: 2023MNRAS.526..616G, pp. 616–644. ISSN: 0035-8711. DOI: 10.1093/mnras/stad2717. URL: https://ui.adsabs.harvard.edu/abs/2023MNRAS.526..616G (visited on 08/21/2024).

[53] V. G. Gurzadian and G. K. Savvidi. “On the problem of relaxation of stellar systems.” In: *Akademiia Nauk SSSR Doklady* 277 (Jan. 1, 1984). ADS Bibcode: 1984DoSSR.277...69G, pp. 69–73. ISSN: 0002-3264. URL: https://ui.adsabs.harvard.edu/abs/1984DoSSR.277...69G (visited on 10/21/2024).

[54] V. G. Gurzadian and G. K. Savvidy. “Collective relaxation of stellar systems.” In: *Astronomy and Astrophysics* 160 (May 1, 1986). ADS Bibcode: 1986A&A...160..203G, pp. 203–210. ISSN: 0004-6361. URL: https://ui.adsabs.harvard.edu/abs/1986A&A...160..203G (visited on 08/13/2024).

[55] Alan H. Guth. *The Inflationary Universe: The Quest for a New Theory of Cosmic Origins*. Google-Books-ID: Jz7eR4wu9hEC. Vintage, 1998. 378 pp. ISBN: 978-0-09-995950-2.

[56] P. A. Harrison, G. Rixon, and P. A. Harrison. *Universal Worker Service Pattern Version 1.1*. DOI: 10.5479/ADS/bib/2016ivoa.spec.1024H. URL: http://ui.adsabs.harvard.edu/abs/2016ivoa.spec.1024H/abstract (visited on 10/17/2024).

[57] Douglas C. Heggie. “The N-Body Problem in Stellar Dynamics”. In: *Long-Term Dynamical Behaviour of Natural and Artificial N-Body Systems*. Ed. by Archie E. Roy. Dordrecht: Springer Netherlands, 1988, pp. 329–347. ISBN: 978-94-009-3053-7. DOI: 10.1007/978-94-009-3053-7_30. URL: https://doi.org/10.1007/978-94-009-3053-7_30 (visited on 08/09/2024).

[58] Marc Hemsendorf and David Merritt. “Instability of the Gravitational N-Body Problem in the Large-N Limit”. In: *The Astrophysical Journal* 580 (Nov. 1, 2002). Publisher: IOP ADS Bibcode: 2002ApJ...580..606H, pp. 606–609. ISSN: 0004-637X. DOI: 10.1086/343027. URL: https://ui.adsabs.harvard.edu/abs/2002ApJ...580..606H (visited on 08/13/2024).

[59] Michel Henon and Carl Heiles. “The applicability of the third integral of motion: Some numerical experiments”. In: *The Astronomical Journal* 69 (Feb. 1964), p. 73. ISSN: 00046256. DOI: 10.1086/109234. URL: http://adsabs.harvard.edu/cgi-bin/bib_query?1964AJ.....69...73H (visited on 08/02/2024).

[60] R. W. Hockney, S. P. Goel, and J. W. Eastwood. "Quiet High-Resolution Computer Models of a Plasma". In: *Journal of Computational Physics* 14 (Feb. 1, 1974). ADS Bibcode: 1974JCoPh..14..148H, pp. 148–158. ISSN: 0021-9991. DOI: 10.1016/0021-9991(74)90010-2. URL: https://ui.adsabs.harvard.edu/abs/1974JCoPh..14..148H (visited on 07/29/2024).

[61] Yehuda Hoffman and Erez Ribak. "Constrained Realizations of Gaussian Fields: A Simple Algorithm". In: *The Astrophysical Journal* 380 (Oct. 1, 1991). Publisher: IOP ADS Bibcode: 1991ApJ...380L...5H, p. L5. ISSN: 0004-637X. DOI: 10.1086/186160. URL: https://ui.adsabs.harvard.edu/abs/1991ApJ...380L...5H (visited on 08/17/2024).

[62] S E Jaffa et al. "Chaotic star formation: error bars for the star formation efficiency and column density PDF". In: *Monthly Notices of the Royal Astronomical Society* 511.2 (Apr. 1, 2022), pp. 2702–2707. ISSN: 0035-8711. DOI: 10.1093/mnras/stac131. URL: https://doi.org/10.1093/mnras/stac131 (visited on 08/17/2024).

[63] James Hopwood Jeans and George Howard Darwin. "I. The stability of a spherical nebula". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 199.312 (Jan. 1997). Publisher: Royal Society, pp. 1–53. DOI: 10.1098/rsta.1902.0012. URL: https://royalsocietypublishing.org/doi/10.1098/rsta.1902.0012 (visited on 10/11/2024).

[64] Pritish Jetley et al. "Massively parallel cosmological simulations with ChaNGa". In: *2008 IEEE International Symposium on Parallel and Distributed Processing*. Distributed Processing Symposium (IPDPS). ISSN: 1530-2075. Miami, FL, USA: IEEE, Apr. 2008, pp. 1–12. ISBN: 978-1-4244-1693-6. DOI: 10.1109/IPDPS.2008.4536319. URL: http://ieeexplore.ieee.org/document/4536319/ (visited on 10/11/2024).

[65] Pritish Jetley et al. "Scaling Hierarchical N-body Simulations on GPU Clusters". In: *Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*. SC '10. USA: IEEE Computer Society, Nov. 13, 2010, pp. 1–11. ISBN: 978-1-4244-7559-9. DOI: 10.1109/SC.2010.49. URL: https://doi.org/10.1109/SC.2010.49 (visited on 10/11/2024).

[66] Martin Jubelgas, Volker Springel, and Klaus Dolag. "Thermal conduction in cosmological SPH simulations". In: *Monthly Notices of the Royal Astronomical Society* 351 (June 1, 2004). Publisher: OUP ADS Bibcode: 2004MNRAS.351..423J, pp. 423–435. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2004.07801.x. URL: https://ui.adsabs.harvard.edu/abs/2004MNRAS.351..423J (visited on 12/19/2024).

[67] Henry E Kandrup and Haywood Smith. "ON THE SENSITIVITY OF THE V-BODY PROBLEM TO SMALL CHANGES IN INITIAL CONDITIONS". In: *ApJ. . .* 374.1 (1991).

[68] Henry E. Kandrup. "Divergence of Nearby Trajectories for the Gravitational N-Body Problem". In: *The Astrophysical Journal* 364 (Dec. 1, 1990). Publisher: IOP ADS Bibcode: 1990ApJ...364..420K, p. 420. ISSN: 0004-637X. DOI: 10.1086/169425. URL: https://ui.adsabs.harvard.edu/abs/1990ApJ...364..420K (visited on 08/13/2024).

[69] Henry E. Kandrup. "How fast can a galaxy "mix"?" In: *Physica A Statistical Mechanics and its Applications* 169 (Nov. 1, 1990). ADS Bibcode: 1990PhyA..169...73K, pp. 73–94. ISSN: 0378-4371. DOI: 10.1016/0378-4371(90)90217-G. URL: https://ui.adsabs.harvard.edu/abs/1990PhyA..169...73K (visited on 08/13/2024).

[70] Henry E. Kandrup. "The time scale for "mixing" in a stellar dynamical system". In: *Physics Letters A* 140 (Sept. 1, 1989). ADS Bibcode: 1989PhLA..140...97K, pp. 97–100. ISSN: 0375-9601. DOI: 10.1016/0375-9601(89)90497-0. URL: https://ui.adsabs.harvard.edu/abs/1989PhLA..140...97K (visited on 08/13/2024).

[71] Henry E. Kandrup, M. Elaine Mahon, and Haywood Smith Jr. "On the Sensitivity of the N-Body Problem toward Small Changes in Initial Conditions. IV." In: *The Astrophysical Journal* 428 (June 1, 1994). Publisher: IOP ADS Bibcode: 1994ApJ...428..458K, p. 458. ISSN: 0004-637X. DOI: 10.1086/174259. URL: https://ui.adsabs.harvard.edu/abs/1994ApJ...428..458K (visited on 08/12/2024).

[72] Henry E. Kandrup, Haywood Smith Jr., and David E. Willmes. "On the Sensitivity of the N-Body Problem to Small Changes in Initial Conditions. III." In: *The Astrophysical Journal* 399 (Nov. 1, 1992). Publisher: IOP ADS Bibcode: 1992ApJ...399..627K, p. 627. ISSN: 0004-637X. DOI: 10.1086/171954. URL: https://ui.adsabs.harvard.edu/abs/1992ApJ...399..627K (visited on 08/12/2024).

[73] A. I. Karakas. "VizieR Online Data Catalog: Updated stellar yields from AGB models (Karakas, 2010)". In: *VizieR Online Data Catalog* 740 (Oct. 1, 2010). ADS Bibcode: 2010yCat..74031413K, J/MNRAS/403/1413. URL: https://ui.adsabs.harvard.edu/abs/2010yCat..74031413K (visited on 10/11/2024).

[74] Manolis Katevenis et al. "Next generation of Exascale-class systems: ExaNeSt project and the status of its interconnect and storage development". In: *Microprocessors and Microsystems* 61 (Sept. 1, 2018), pp. 58–71. ISSN: 0141-9331. DOI: 10.1016/j.micpro.2018.05.009. URL: https://www.sciencedirect.com/science/article/pii/S0141933118300188 (visited on 07/29/2024).

[75] Alexander A. Kaurov. *Stochasticity in the 21cm power spectrum at the epoch of reionization and cosmic dawn*. Publication Title: arXiv e-prints ADS Bibcode: 2017arXiv170904353K. Sept. 1, 2017. DOI: 10.48550/arXiv.1709.04353. URL: https://ui.adsabs.harvard.edu/abs/2017arXiv170904353K (visited on 08/13/2024).

[76] B W Keller et al. “Chaos and variance in galaxy formation”. In: *Monthly Notices of the Royal Astronomical Society* 482.2 (Jan. 11, 2019), pp. 2244–2261. ISSN: 0035-8711. DOI: 10.1093/mnras/sty2859. URL: https://doi.org/10.1093/mnras/sty2859 (visited on 08/06/2024).

[77] Benjamin W. Keller and J. M. Diederik Kruijssen. “Uncertainties in supernova input rates drive qualitative differences in simulations of galaxy evolution”. In: *Monthly Notices of the Royal Astronomical Society* 512.1 (Mar. 14, 2022), pp. 199–215. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stac511. arXiv: 2004.03608[astro-ph]. URL: http://arxiv.org/abs/2004.03608 (visited on 08/17/2024).

[78] E. Komatsu et al. *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*. Nov. 9, 2010. DOI: 10.1088/0067-0049/192/2/18. arXiv: 1001.4538[astro-ph]. URL: http://arxiv.org/abs/1001.4538 (visited on 08/21/2024).

[79] Douglas Kothe, Stephen Lee, and Irene Qualters. “Exascale Computing in the United States”. In: *Computing in Science & Engineering* 21.1 (Jan. 2019). Conference Name: Computing in Science & Engineering, pp. 17–29. ISSN: 1558-366X. DOI: 10.1109/MCSE.2018.2875366. URL: https://ieeexplore.ieee.org/document/8528398 (visited on 07/29/2024).

[80] Andrey V. Kravtsov and Stefano Borgani. “Formation of Galaxy Clusters”. In: *Annual Review of Astronomy and Astrophysics* 50 (Sept. 1, 2012). ADS Bibcode: 2012ARA&A..50..353K, pp. 353–409. ISSN: 0066-4146. DOI: 10.1146/annurev-astro-081811-125502. URL: https://ui.adsabs.harvard.edu/abs/2012ARA&A..50..353K (visited on 10/28/2024).

[81] Cedric Lacey and Shaun Cole. “Merger rates in hierarchical models of galaxy formation”. In: *Monthly Notices of the Royal Astronomical Society* 262 (June 1, 1993). Publisher: OUP ADS Bibcode: 1993MNRAS.262..627L, pp. 627–649. ISSN: 0035-8711. DOI: 10.1093/mnras/262.3.627. URL: https://ui.adsabs.harvard.edu/abs/1993MNRAS.262..627L (visited on 10/11/2024).

[82] David Languignon et al. *Simulation Data Access Layer Version 1.0*. Mar. 20, 2017. DOI: 10.5479/ADS/bib/2017ivoa.spec.0320L. URL: http://ui.adsabs.harvard.edu/abs/2017ivoa.spec.0320L/abstract (visited on 10/17/2024).

[83] J. Laskar. “A numerical experiment on the chaotic behaviour of the Solar System”. In: *Nature* 338 (Mar. 1, 1989). ADS Bibcode: 1989Natur.338..237L, pp. 237–238. ISSN: 0028-0836. DOI: 10.1038/338237a0. URL: https://ui.adsabs.harvard.edu/abs/1989Natur.338..237L (visited on 08/13/2024).

[84] J. Laskar and M. Gastineau. “Existence of collisional trajectories of Mercury, Mars and Venus with the Earth”. In: *Nature* 459 (June 1, 2009). ADS Bibcode: 2009Natur.459..817L, pp. 817–819. ISSN: 0028-0836. DOI: 10.1038/nature08096. URL: https://ui.adsabs.harvard.edu/abs/2009Natur.459..817L (visited on 08/13/2024).

[85] Jacques Laskar, Thomas Quinn, and Scott Tremaine. “Confirmation of resonant structure in the solar system”. In: *Icarus* 95 (Jan. 1, 1992). ADS Bibcode: 1992Icar...95..148L, pp. 148–152. ISSN: 0019-1035. DOI: 10.1016/0019-1035(92)90196-E. URL: https://ui.adsabs.harvard.edu/abs/1992Icar...95..148L (visited on 08/13/2024).

[86] G. Lemson and the Virgo Consortium. “Halo and Galaxy Formation Histories from the Millennium Simulation: Public release of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the LambdaCDM cosmogony”. In: *arXiv e-prints* (Aug. 1, 2006). ADS Bibcode: 2006astro.ph..8019L, astro–ph/0608019. URL: https://ui.adsabs.harvard.edu/abs/2006astro.ph..8019L (visited on 11/08/2022).

[87] Gerard Lemson et al. *Simulation Data Model Version 1.0*. May 3, 2012. DOI: 10.5479/ADS/bib/2012ivoa.spec.0503L. URL: http://ui.adsabs.harvard.edu/abs/2012ivoa.spec.0503L/abstract (visited on 10/17/2024).

[88] Andrei Linde. “Prospects of Inflation”. In: *Physica Scripta* 2005 (T117 Jan. 1, 2005). Publisher: IOP Publishing, p. 40. ISSN: 1402-4896. DOI: 10.1238/Physica.Topical.117a00040. URL: https://iopscience.iop.org/article/10.1238/Physica.Topical.117a00040/meta (visited on 10/11/2024).

[89] Kong You Liow and Clare L Dobbs. “The role of collision speed, cloud density, and turbulence in the formation of young massive clusters via cloud–cloud collisions”. In: *Monthly Notices of the Royal Astronomical Society* 499.1 (Oct. 22, 2020), pp. 1099–1115. ISSN: 0035-8711. DOI: 10.1093/mnras/staa2857. URL: https://doi.org/10.1093/mnras/staa2857 (visited on 08/17/2024).

[90] Edward N. Lorenz. “Deterministic Nonperiodic Flow”. In: (Mar. 1, 1963). Section: Journal of the Atmospheric Sciences. ISSN: 1520-0469. URL: https://journals.ametsoc.org/view/journals/atsc/20/2/1520-0469_1963_020_0130_dnf_2_0_co_2.xml (visited on 07/30/2024).

[91] Mireille Louys et al. *Observation Data Model Core Components, its Implementation in the Table Access Protocol Version 1.1*. DOI: 10.5479/ADS/bib/2017ivoa.spec.0509L. URL: http://ui.adsabs.harvard.edu/abs/2017ivoa.spec.0509L/abstract (visited on 10/17/2024).

[92] Mark R Lovell et al. “The fraction of dark matter within galaxies from the IllustrisTNG simulations”. In: *Monthly Notices of the Royal Astronomical Society* 481.2 (Dec. 1, 2018), pp. 1950–1975. ISSN: 0035-8711. DOI: 10.1093/mnras/sty2339. URL: https://doi.org/10.1093/mnras/sty2339 (visited on 08/23/2024).

[93] Piero Madau and Mark Dickinson. “Cosmic Star-Formation History”. In: *Annual Review of Astronomy and Astrophysics* 52 (Volume 52, 2014 Aug. 18, 2014). Publisher: Annual Reviews, pp. 415–486. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev-astro-081811-125615. URL: https:

//www.annualreviews.org/content/journals/10.1146/annurev-astro-081811-125615 (visited on 10/26/2024).

[94] N. P. Maffione et al. "On the relevance of chaos for halo stars in the solar neighbourhood". In: *Monthly Notices of the Royal Astronomical Society* 453.3 (Nov. 1, 2015), pp. 2830–2847. ISSN: 0035-8711. DOI: 10.1093/mnras/stv1778. URL: https://doi.org/10.1093/mnras/stv1778 (visited on 08/17/2024).

[95] Benoit B. Mandelbrot. *The fractal geometry of nature*. Revised edition. OCLC: 7876824. San Francisco: W.H. Freeman, 1982. 460 pp. ISBN: 978-0-7167-1186-5. URL: http://catdir.loc.gov/catdir/bios/hol056/81015085.html (visited on 07/30/2024).

[96] Francesca Matteucci and Paolo Padovani. "Chemical Evolution of Galaxies and Quasar Metallicities". In: *The Astrophysical Journal* 419 (Dec. 1, 1993). Publisher: IOP ADS Bibcode: 1993ApJ...419..485M, p. 485. ISSN: 0004-637X. DOI: 10.1086/173502. URL: https://ui.adsabs.harvard.edu/abs/1993ApJ...419..485M (visited on 10/11/2024).

[97] Robert May. "Simple Mathematical Models With Very Complicated Dynamics". In: *Nature* 26 (July 1, 1976), p. 457. ISSN: 978-1-4419-2330-1. DOI: 10.1038/261459a0.

[98] Harshitha Menon et al. "Adaptive techniques for clustered N-body cosmological simulations". In: *Computational Astrophysics and Cosmology* 2.1 (Mar. 28, 2015), p. 1. ISSN: 2197-7909. DOI: 10.1186/s40668-015-0007-9. URL: https://doi.org/10.1186/s40668-015-0007-9 (visited on 10/11/2024).

[99] Andrea Milani and Anna M. Nobili. "An example of stable chaos in the Solar System". In: *Nature* 357.6379 (June 1992). Publisher: Nature Publishing Group, pp. 569–571. ISSN: 1476-4687. DOI: 10.1038/357569a0. URL: https://www.nature.com/articles/357569a0 (visited on 08/13/2024).

[100] *MillenniumTNG Project: high-precision predictions for matter clustering and halo statistics | Monthly Notices of the Royal Astronomical Society | Oxford Academic*. URL: https://academic.oup.com/mnras/article/524/2/2556/7226466?login=true (visited on 10/12/2024).

[101] R. H. Miller. "Irreversibility in Small Stellar Dynamical Systems." In: *The Astrophysical Journal* 140 (July 1, 1964). Publisher: IOP ADS Bibcode: 1964ApJ...140..250M, p. 250. ISSN: 0004-637X. DOI: 10.1086/147911. URL: https://ui.adsabs.harvard.edu/abs/1964ApJ...140..250M (visited on 08/01/2024).

[102] J. J. Monaghan. "Smoothed Particle Hydrodynamics". In: *Annual Review of Astronomy and Astrophysics* 30 (Volume 30, 1992 Sept. 1, 1992). Publisher: Annual Reviews, pp. 543–574. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev.aa.30.090192.002551. URL: https://www.annualreviews.org/content/journals/10.1146/annurev.aa.30.090192.002551 (visited on 07/30/2024).

[103] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. "A Universal density profile from hierarchical clustering". In: *Astrophys. J.* 490 (1997). _eprint: astro-ph/9611107, pp. 493–508. DOI: 10.1086/304888.

[104] Francois Ochsenbein et al. *VOTable Format Definition Version 1.3*. DOI: 10.5479/ADS/bib/2013ivoa.spec.09200. URL: http://ui.adsabs.harvard.edu/abs/2013ivoa.spec.09200/abstract (visited on 10/17/2024).

[105] P. J. E. Peebles. "The Black-Body Radiation Content of the Universe and the Formation of Galaxies." In: *The Astrophysical Journal* 142 (Nov. 1, 1965). Publisher: IOP ADS Bibcode: 1965ApJ...142.1317P, p. 1317. ISSN: 0004-637X. DOI: 10.1086/148417. URL: https://ui.adsabs.harvard.edu/abs/1965ApJ...142.1317P (visited on 07/28/2024).

[106] Annalisa Pillepich et al. "Simulating galaxy formation with the IllustrisTNG model". In: *Monthly Notices of the Royal Astronomical Society* 473.3 (Jan. 21, 2018), pp. 4077–4106. ISSN: 0035-8711. DOI: 10.1093/mnras/stx2656. URL: https://doi.org/10.1093/mnras/stx2656 (visited on 08/17/2024).

[107] S. Planelles et al. "Pressure of the hot gas in simulations of galaxy clusters". In: *Monthly Notices of the Royal Astronomical Society* 467.4 (June 1, 2017), pp. 3827–3847. ISSN: 0035-8711. DOI: 10.1093/mnras/stx318. URL: https://doi.org/10.1093/mnras/stx318 (visited on 08/20/2024).

[108] Raymond Plante et al. *VOResource: an XML Encoding Schema for Resource Metadata Version 1.1*. DOI: 10.5479/ADS/bib/2018ivoa.spec.0625P. URL: http://ui.adsabs.harvard.edu/abs/2018ivoa.spec.0625P/abstract (visited on 10/17/2024).

[109] Henri Poincaré. "Sur le problème des trois corps et les équations de la dynamique". In: *Acta mathematica* 13 (1890), p. 270.

[110] Douglas Potter, Joachim Stadel, and Romain Teyssier. "PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys". In: *Computational Astrophysics and Cosmology* 4.1 (May 18, 2017), p. 2. ISSN: 2197-7909. DOI: 10.1186/s40668-017-0021-1. URL: https://doi.org/10.1186/s40668-017-0021-1 (visited on 07/29/2024).

[111] William H. Press and Paul Schechter. "Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation". In: *The Astrophysical Journal* 187 (Feb. 1, 1974). Publisher: IOP ADS Bibcode: 1974ApJ...187..425P, pp. 425–438. ISSN: 0004-637X. DOI: 10.1086/152650. URL: https://ui.adsabs.harvard.edu/abs/1974ApJ...187..425P (visited on 10/11/2024).

[112] William H. Press and William T. Vetterling. *Numerical Recipes in C++: The Art of Scientific Computing*. Google-Books-ID: Hs9GPgAACAAJ. Cambridge University Press India, 2009. 1002 pp. ISBN: 978-81-7596-096-1.

[113] Daniel J. Price. “Smoothed particle hydrodynamics and magnetohydrodynamics”. In: *Journal of Computational Physics* 231 (Feb. 1, 2012). ADS Bibcode: 2012JCoPh.231..759P, pp. 759–794. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2010.12.011. URL: https://ui.adsabs.harvard.edu/abs/2012JCoPh.231..759P (visited on 07/30/2024).

[114] Adrian M. Price-Whelan et al. “Chaotic dispersal of tidal debris”. In: *Monthly Notices of the Royal Astronomical Society* 455.1 (Jan. 1, 2016), pp. 1079–1098. ISSN: 1365-2966, 0035-8711. DOI: 10.1093/mnras/stv2383. URL: http://academic.oup.com/mnras/article/455/1/1079/986793/Chaotic-dispersal-of-tidal-debris (visited on 08/17/2024).

[115] Adrian M. Price-Whelan et al. “SPENDING TOO MUCH TIME AT THE GALACTIC BAR: CHAOTIC FANNING OF THE OPHIUCHUS STREAM”. In: *The Astrophysical Journal* 824.2 (June 2016). Publisher: The American Astronomical Society, p. 104. ISSN: 0004-637X. DOI: 10.3847/0004-637X/824/2/104. URL: https://dx.doi.org/10.3847/0004-637X/824/2/104 (visited on 08/17/2024).

[116] C Ragone-Figueroa et al. “BCG mass evolution in cosmological hydrosimulations”. In: *Monthly Notices of the Royal Astronomical Society* 479.1 (Sept. 1, 2018), pp. 1125–1136. ISSN: 0035-8711. DOI: 10.1093/mnras/sty1639. URL: https://doi.org/10.1093/mnras/sty1639 (visited on 08/20/2024).

[117] Cinthia Ragone-Figueroa et al. “Brightest cluster galaxies in cosmological simulations: achievements and limitations of active galactic nuclei feedback models”. In: *Monthly Notices of the Royal Astronomical Society* 436.2 (Dec. 1, 2013), pp. 1750–1764. ISSN: 0035-8711. DOI: 10.1093/mnras/stt1693. URL: https://doi.org/10.1093/mnras/stt1693 (visited on 01/17/2024).

[118] E. Rasia et al. “COOL CORE CLUSTERS FROM COSMOLOGICAL SIMULATIONS”. In: *The Astrophysical Journal* 813.1 (Oct. 2015). Publisher: American Astronomical Society, p. L17. ISSN: 2041-8205. DOI: 10.1088/2041-8205/813/1/L17. URL: https://doi.org/10.1088/2041-8205/813/1/l17 (visited on 08/09/2022).

[119] Martin P. Rey and Andrew Pontzen. “Quadratic genetic modifications: a streamlined route to cosmological simulations with controlled merger history”. In: *Monthly Notices of the Royal Astronomical Society* 474.1 (Feb. 11, 2018), pp. 45–54. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stx2744. URL: http://academic.oup.com/mnras/article/474/1/45/4563625 (visited on 08/17/2024).

[120] K. Riebe et al. “The MultiDark Database: Release of the Bolshoi and MultiDark cosmological simulations”. In: *Astronomische Nachrichten* 334.7 (2013). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asna.201211900, pp. 691–708. ISSN: 1521-3994. DOI: 10.1002/asna.201211900. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asna.201211900 (visited on 10/26/2022).

[121] D. Romano et al. “Quantifying the uncertainties of chemical evolution studies - II. Stellar yields”. In: *Astronomy & Astrophysics* 522 (Nov. 1, 2010). Publisher: EDP Sciences, A32. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201014483. URL: https://www.aanda.org/articles/aa/abs/2010/14/aa14483-10/aa14483-10.html (visited on 10/11/2024).

[122] Nina Roth, Andrew Pontzen, and Hiranya V. Peiris. “Genetically modified haloes: towards controlled experiments in CDM galaxy formation”. In: *Monthly Notices of the Royal Astronomical Society* 455.1 (Jan. 1, 2016), pp. 974–986. ISSN: 1365-2966, 0035-8711. DOI: 10.1093/mnras/stv2375. URL: http://academic.oup.com/mnras/article/455/1/974/986737/Genetically-modified-haloes-towards-controlled (visited on 08/17/2024).

[123] Matthieu Schaller et al. “Baryon effects on the internal structure of CDM haloes in the EAGLE simulations”. In: *Monthly Notices of the Royal Astronomical Society* 451 (Aug. 1, 2015). Publisher: OUP ADS Bibcode: 2015MNRAS.451.1247S, pp. 1247–1267. ISSN: 0035-8711. DOI: 10.1093/mnras/stv1067. URL: https://ui.adsabs.harvard.edu/abs/2015MNRAS.451.1247S (visited on 08/24/2024).

[124] Matthieu Schaller et al. “SWIFT: A modern highly-parallel gravity and smoothed particle hydrodynamics solver for astrophysical and cosmological applications”. In: *Monthly Notices of the Royal Astronomical Society* 530.2 (Apr. 13, 2024), pp. 2378–2419. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stae922. arXiv: 2305.13380[astro-ph]. URL: http://arxiv.org/abs/2305.13380 (visited on 08/17/2024).

[125] Lucie Scharré, Daniele Sorini, and Romeel Davé. “The effects of stellar and AGN feedback on the cosmic star formation history in the simba simulations”. In: *Monthly Notices of the Royal Astronomical Society* 534.1 (Oct. 11, 2024), pp. 361–383. ISSN: 0035-8711. DOI: 10.1093/mnras/stae2098. URL: https://doi.org/10.1093/mnras/stae2098 (visited on 10/26/2024).

[126] Joop Schaye et al. “The EAGLE project: simulating the evolution and assembly of galaxies and their environments”. In: *Monthly Notices of the Royal Astronomical Society* 446.1 (Jan. 1, 2015), pp. 521–554. ISSN: 0035-8711. DOI: 10.1093/mnras/stu2058. URL: https://doi.org/10.1093/mnras/stu2058 (visited on 08/17/2024).

[127] Maarten Schmidt. “The Rate of Star Formation.” In: *The Astrophysical Journal* 129 (Mar. 1, 1959). Publisher: IOP ADS Bibcode: 1959ApJ...129..243S, p. 243. ISSN: 0004-637X. DOI: 10.1086/146614. URL: https://ui.adsabs.harvard.edu/abs/1959ApJ...129..243S (visited on 07/30/2024).

[128] J. A. Sellwood and Victor P. Debattista. “Bar-Halo Friction in Galaxies. II. Metastability”. In: *The Astrophysical Journal* 639.2 (Mar. 2006), p. 868. ISSN: 0004-637X. DOI: 10.1086/499482. URL: https://dx.doi.org/10.1086/499482 (visited on 08/17/2024).

[129] J. A. Sellwood and Victor P. Debattista. "Stochasticity in N -body simulations of disc galaxies". In: *Monthly Notices of the Royal Astronomical Society* 398.3 (Sept. 21, 2009), pp. 1279–1297. ISSN: 00358711, 13652966. DOI: 10.1111/j.1365-2966.2009.15219.x. URL: https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2009.15219.x (visited on 08/17/2024).

[130] Daniele Sorini and John A Peacock. "Extended Hernquist–Springel formalism for cosmic star formation". In: *Monthly Notices of the Royal Astronomical Society* 508.4 (Dec. 21, 2021), pp. 5802–5824. ISSN: 0035-8711. DOI: 10.1093/mnras/stab2845. URL: https://doi.org/10.1093/mnras/stab2845 (visited on 10/26/2024).

[131] Volker Springel. "E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh". In: *Monthly Notices of the Royal Astronomical Society* 401.2 (Jan. 11, 2010), pp. 791–851. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2009.15715.x. URL: https://doi.org/10.1111/j.1365-2966.2009.15715.x (visited on 10/11/2024).

[132] Volker Springel. "Smoothed Particle Hydrodynamics in Astrophysics". In: *Annual Review of Astronomy and Astrophysics* 48 (Volume 48, 2010 Sept. 22, 2010). Publisher: Annual Reviews, pp. 391–430. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev-astro-081309-130914. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-astro-081309-130914 (visited on 07/30/2024).

[133] Volker Springel. "The cosmological simulation code GADGET-2". In: *Monthly Notices of the Royal Astronomical Society* 364 (Dec. 1, 2005). ADS Bibcode: 2005MNRAS.364.1105S, pp. 1105–1134. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09655.x. URL: https://ui.adsabs.harvard.edu/abs/2005MNRAS.364.1105S (visited on 08/09/2022).

[134] Volker Springel, Tiziana Di Matteo, and Lars Hernquist. "Modelling feedback from stars and black holes in galaxy mergers". In: *Monthly Notices of the Royal Astronomical Society* 361 (Aug. 1, 2005). ADS Bibcode: 2005MNRAS.361..776S, pp. 776–794. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09238.x. URL: https://ui.adsabs.harvard.edu/abs/2005MNRAS.361..776S (visited on 03/12/2024).

[135] Volker Springel and Lars Hernquist. "Cosmological SPH simulations: A hybrid multi-phase model for star formation". In: *Monthly Notices of the Royal Astronomical Society* 339.2 (Feb. 21, 2003), pp. 289–311. ISSN: 0035-8711, 1365-2966. DOI: 10.1046/j.1365-8711.2003.06206.x. arXiv: astro-ph/0206393. URL: http://arxiv.org/abs/astro-ph/0206393 (visited on 03/07/2024).

[136] Volker Springel, Naoki Yoshida, and Simon D. M. White. "GADGET: a code for collisionless and gasdynamical cosmological simulations". In: *New Astronomy* 6.2 (Apr. 1, 2001), pp. 79–117. ISSN: 1384-1076. DOI: 10.1016/S1384-1076(01)00042-2. URL: https://www.sciencedirect.com/science/article/pii/S1384107601000422 (visited on 10/28/2024).

[137] Volker Springel et al. “Populating a cluster of galaxies - I. Results at z=0”. In: *Monthly Notices of the Royal Astronomical Society* 328 (Dec. 1, 2001). Publisher: OUP ADS Bibcode: 2001MNRAS.328..726S, pp. 726–750. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2001.04912.x. URL: https://ui.adsabs.harvard.edu/abs/2001MNRAS.328..726S (visited on 07/30/2024).

[138] Volker Springel et al. “Simulations of the formation, evolution and clustering of galaxies and quasars”. In: *Nature* 435 (June 1, 2005). ADS Bibcode: 2005Natur.435..629S, pp. 629–636. ISSN: 0028-0836. DOI: 10.1038/nature03597. URL: https://ui.adsabs.harvard.edu/abs/2005Natur.435..629S (visited on 10/21/2024).

[139] Joachim Gerhard Stadel. “Cosmological N-body simulations and their analysis”. Pages: 3657 Publication Title: Ph.D. Thesis ADS Bibcode: 2001PhDT.....21S. PhD thesis. Jan. 1, 2001. URL: https://ui.adsabs.harvard.edu/abs/2001PhDT.....21S (visited on 07/29/2024).

[140] Erland Myles Standish Jr. “Numerical Studies of the Gravitational Problem of N Bodies.” Publication Title: Ph.D. Thesis ADS Bibcode: 1968PhDT.....6S. PhD thesis. Yale University, Jan. 1, 1968. URL: https://ui.adsabs.harvard.edu/abs/1968PhDT.....6S (visited on 08/09/2024).

[141] Steven H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. 2nd ed. Boca Raton: CRC Press, May 23, 2019. 532 pp. ISBN: 978-0-429-49256-3. DOI: 10.1201/9780429492563.

[142] Kung-Yi Su et al. “Discrete effects in stellar feedback: Individual Supernovae, Hypernovae, and IMF Sampling in Dwarf Galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 480.2 (Oct. 21, 2018), pp. 1666–1675. ISSN: 0035-8711. DOI: 10.1093/mnras/sty1928. URL: https://doi.org/10.1093/mnras/sty1928 (visited on 08/05/2024).

[143] Kung-Yi Su et al. “Feedback first: the surprisingly weak effects of magnetic fields, viscosity, conduction and metal diffusion on sub-L* galaxy formation”. In: *Monthly Notices of the Royal Astronomical Society* 471.1 (Oct. 11, 2017), pp. 144–166. ISSN: 0035-8711. DOI: 10.1093/mnras/stx1463. URL: https://doi.org/10.1093/mnras/stx1463 (visited on 08/05/2024).

[144] Giuliano Taffoni et al. *CHIPP: INAF pilot project for HTC, HPC and HPDA*. Feb. 4, 2020. DOI: 10.48550/arXiv.2002.01283. arXiv: 2002.01283[astro-ph]. URL: http://arxiv.org/abs/2002.01283 (visited on 03/05/2024).

[145] Romain Teyssier. *Cosmological Hydrodynamics with Adaptive Mesh Refinement: a new high resolution code called RAMSES*. Nov. 19, 2001. DOI: 10.48550/arXiv.astro-ph/0111367. arXiv: astro-ph/0111367. URL: http://arxiv.org/abs/astro-ph/0111367 (visited on 10/30/2024).

[146] J. Thiébaut et al. “On the onset of stochasticity in cold dark matter cosmological simulations”. In: *Monthly Notices of the Royal Astronomical Society* 387.1 (June 11, 2008), pp. 397–406. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2008.13250.x. URL: https://doi.org/10.1111/j.1365-2966.2008.13250.x (visited on 08/09/2024).

[147] F.-K. Thielemann et al. “Supernova Nucleosynthesis and Galactic Evolution”. In: *From Twilight to Highlight: The Physics of Supernovae*. Ed. by Wolfgang Hillebrandt and Bruno Leibundgut. Berlin, Heidelberg: Springer, 2003, pp. 331–343. ISBN: 978-3-540-36427-6. DOI: 10.1007/10828549_46.

[148] Doug Tody et al. *IVOA Recommendation: Simple Spectral Access Protocol Version 1.1*. Mar. 26, 2012. DOI: 10.48550/arXiv.1203.5725. arXiv: 1203.5725. URL: http://arxiv.org/abs/1203.5725 (visited on 10/17/2024).

[149] Giuseppe Tormen, François R. Bouchet, and Simon D. M. White. “The structure and dynamical evolution of dark matter haloes”. In: *Monthly Notices of the Royal Astronomical Society* 286.4 (Apr. 21, 1997), pp. 865–884. ISSN: 0035-8711. DOI: 10.1093/mnras/286.4.865. URL: https://doi.org/10.1093/mnras/286.4.865 (visited on 08/20/2024).

[150] L. Tornatore et al. “Chemical enrichment of galaxy clusters from hydrodynamical simulations: Chemical enrichment of galaxy clusters”. In: *Monthly Notices of the Royal Astronomical Society* 382.3 (Nov. 28, 2007), pp. 1050–1072. ISSN: 00358711. 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 (visited on 10/11/2024).

[151] L. Tornatore et al. “Cooling and heating the intracluster medium in hydrodynamical simulations”. In: *Monthly Notices of the Royal Astronomical Society* 342.4 (July 11, 2003), pp. 1025–1040. ISSN: 0035-8711, 1365-2966. DOI: 10.1046/j.1365-8711.2003.06631.x. URL: https://academic.oup.com/mnras/article/342/4/1025/956309 (visited on 10/11/2024).

[152] Rainer Weinberger et al. “Simulating galaxy formation with black hole driven thermal and kinetic feedback”. In: *Monthly Notices of the Royal Astronomical Society* 465.3 (Mar. 1, 2017), pp. 3291–3308. ISSN: 0035-8711. DOI: 10.1093/mnras/stw2944. URL: https://doi.org/10.1093/mnras/stw2944 (visited on 08/17/2024).

[153] S. D. M. White and M. J. Rees. “Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering”. In: *Monthly Notices of the Royal Astronomical Society* 183.3 (July 1, 1978), pp. 341–358. ISSN: 0035-8711. DOI: 10.1093/mnras/183.3.341. URL: https://doi.org/10.1093/mnras/183.3.341 (visited on 10/11/2024).

[154] Mark D. Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific Data* 3.1 (Mar. 15, 2016). Number: 1 Publisher: Nature Publishing Group, p. 160018. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18. URL: https://www.nature.com/articles/sdata201618 (visited on 02/10/2023).

[155] S. E. Woosley and Thomas A. Weaver. "The Evolution and Explosion of Massive Stars. II. Explosive Hydrodynamics and Nucleosynthesis". In: *The Astrophysical Journal Supplement Series* 101 (Nov. 1, 1995). Publisher: IOP ADS Bibcode: 1995ApJS..101..181W, p. 181. ISSN: 0067-0049. DOI: 10.1086/192237. URL: https://ui.adsabs.harvard.edu/abs/1995ApJS..101..181W (visited on 10/11/2024).

[156] Amr El-Zant, Mark Everitt, and Summer Kassem. "Errors, chaos and the collisionless limit". In: *Monthly Notices of the Royal Astronomical Society* 484.2 (Apr. 1, 2019), pp. 1456–1474. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stz029. arXiv: 1804.06920[astro-ph]. URL: http://arxiv.org/abs/1804.06920 (visited on 08/14/2024).

[157] Ya. B. Zel'dovich. "Gravitational instability: An approximate theory for large density perturbations." In: *Astronomy and Astrophysics* 5 (Mar. 1, 1970). ADS Bibcode: 1970A&A.....5...84Z, pp. 84–89. ISSN: 0004-6361. URL: https://ui.adsabs.harvard.edu/abs/1970A&A.....5...84Z (visited on 07/28/2024).

[158] S. F. Portegies Zwart et al. "Chaos in self-gravitating many-body systems - Lyapunov time dependence of N and the influence of general relativity". In: *Astronomy & Astrophysics* 659 (Mar. 1, 2022). Publisher: EDP Sciences, A86. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202141789. URL: https://www.aanda.org/articles/aa/abs/2022/03/aa41789-21/aa41789-21.html (visited on 08/09/2024).