Postprint on the Current Status of Milky Way Numerical Simulations

Authors: Liu Wei, Shao Shi, Gu Qing, In machine learning, we consider the following optimization problem:

$$\min_{x \in \mathbb{R}^d} f(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x) \tag{1}$$

where each $f_i(x)$ denotes the loss function of the i-th sample. We assume that each f_i is convex and L-smooth, i.e., it satisfies the condition in (**??**).

In recent years, the development of deep learning has propelled the widespread application of stochastic gradient descent (SGD) and its variants. As pointed out in [**?**], variance reduction techniques can significantly improve convergence speed. Specifically, the SVRG algorithm reduces variance by periodically computing the full gradient, and its update rule is:

$$x_{t+1} = x_t - \eta(\nabla f_{i_t}(x_t) - \nabla f_{i_t}(\tilde{x}) + \nabla f(\tilde{x})) \tag{2}$$

where i_t is a randomly selected sample index and $\tilde{x}$ is a snapshot point.

The main contributions of this paper include:

- We propose a novel variance reduction algorithm that improves the convergence rate while maintaining computational efficiency.
- We prove the convergence of the algorithm on non-convex problems, as shown in Theorem **??**.
- We validate the effectiveness of the algorithm on multiple datasets through experiments.

Date: 2025-04-02T00:00:00+00:00

Abstract

The standard cosmological model is one of the most successful models for explaining and predicting the formation and evolution of large-scale structures. Over the past few decades, numerical simulations based on the standard cosmological model have continuously deepened our understanding of cosmic structure and galaxy formation, serving as a bridge between theoretical research and observational data. The Milky Way, as the galaxy most familiar to humanity and one that has been observed in detail, provides a valuable sample for studying the history of galaxy formation and evolution. This paper reviews the research progress in numerical simulations of the Milky Way over the past decade. First, it introduces the cosmological background of large-scale structures and the numerical techniques of N-body simulations, analyzes the theoretical models of galaxy formation and hydrodynamic simulation methods, and discusses in detail the main baryonic physical mechanisms affecting galaxy evolution and their numerical subgrid models. Subsequently, it focuses on introducing zoom-in simulation techniques for simulating the Milky Way and some typical Milky Way simulation projects, as well as the current research status of using Milky Way simulations to alleviate small-scale problems. Finally, it summarizes the prospects for Milky Way simulations and presents future outlooks.

Full Text

Preamble

PROGRESS IN ASTRONOMY Vol. 43, No. 1 Mar., 2025 doi: 10.3969/j.issn.1000-8349.2025.01.02

Progress on Numerical Simulations of the Milky Way

LIU Wei[1,2], SHAO Shi[1,2], GU Qing[1], GAO Liang[1,2]

(1. Key Laboratory for Computational Astrophysics, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; 2. School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract

The standard cosmological model is one of the most successful frameworks for explaining and predicting the formation and evolution of large-scale structures. Over the past few decades, numerical simulations based on the standard cosmological model have continuously deepened our understanding of cosmic structures and galaxy formation, serving as a bridge between theoretical research and observational data. The Milky Way, as the most familiar and extensively observed galaxy, provides a valuable sample for studying galaxy formation and evolution history. This review summarizes the research progress on numerical simulations of the Milky Way over the past decade. We first introduce the cosmological background of large-scale structures and the numerical techniques of N-body simulations, analyze theoretical models of galaxy formation and hydrodynamical simulation methods, and discuss in detail the main baryonic physical mechanisms affecting galaxy evolution and their numerical subgrid models. We then focus on zoom-in simulation techniques for the Milky Way and some typical Milky Way simulation projects, as well as the current status of using Milky Way simulations to alleviate small-scale problems. Finally, we summarize and propose prospects for the future of Milky Way simulations.

Keywords: Milky Way; numerical simulation; zoom-in simulation; galaxy formation and evolution; cold dark matter; hydrodynamics

1 Introduction

With the upgrading of observational facilities and innovations in observational techniques, increasingly subtle cosmic structures and phenomena have continuously deepened human understanding of the universe. Large amounts of observational data challenge existing theoretical models, demanding more accurate and comprehensive explanations of observational results. Over the past few decades, with the continuous improvement of computational power, numerical simulations have developed rapidly as a primary tool for cosmological theoretical research, providing strong support for our in-depth study of cosmic evolution and galaxy formation, and connecting modern observations with theoretical research.

However, although numerical simulations have made significant progress in large-scale studies, there remain challenging problems to be solved at small scales below 1 Mpc, such as galaxies and satellite galaxies [1, 2]. This requires not only capturing small-scale structural features and evolutionary processes observationally, but also using higher-resolution numerical simulations and more refined galaxy formation models to describe the subtle internal structures and dynamical states of galaxies and satellite galaxies, in order to more comprehensively understand the formation and evolution history of these important structures in the universe.

The Milky Way, as one of the most familiar and extensively observed galaxies, provides an important sample for studying galaxy formation and evolution processes, and holds unique importance in cosmological research. The structure, dynamics, and composition of the Milky Way have relatively complete observational data. Comparing numerical simulation results of the Milky Way with observational data is one of the important means to verify galaxy formation theoretical models, which can deepen our understanding of the formation and evolution processes of the Milky Way and other galaxies, and effectively constrain the nature of dark matter. For example, observations of the Milky Way's disk rotation curve provide important evidence for the existence of dark matter: there is a huge difference between the rotation curve calculated from observable baryonic matter and the actually observed rotation curve, indicating that in addition to baryonic matter, there exists unobservable matter in galaxies, namely dark matter.

Currently, the main research problems in Milky Way simulations include: 1) The "missing satellite problem": the number of observed satellite galaxies is significantly less than predicted by the standard model; 2) The "plane of satellite problem": the anisotropy in the spatial distribution of satellite galaxies—observed Milky Way satellite galaxies are approximately distributed in the same plane with highly consistent orbits, while the standard model predicts that the probability of this phenomenon occurring is extremely low; 3) The "core-cusp problem": the standard model predicts that the central density of dwarf galaxies is a steep cusp, while observations show a flat core; 4) The "too big to fail problem": the satellite galaxy masses predicted by the standard model are much larger than those observed.

This paper will review the latest progress in numerical simulations of the Milky Way. Section 2 mainly introduces the cosmological background of numerical simulations, namely the formation and evolution of large-scale cosmic structures (Section 2.1) and the N-body simulation methods for handling large-scale structures (Section 2.2). Section 3 systematically expounds the basic principles of galaxy formation (Section 3.1), hydrodynamical simulation techniques used to simulate galaxy formation (Section 3.2), and the most important physical processes affecting galaxy formation and evolution and popular subgrid models (Section 3.3). Section 4 focuses on the process of zoom-in simulations for specific objects of interest (Section 4.1) and common hydrodynamical simulation projects for the Milky Way (Section 4.2), then discusses the latest progress in using Milky Way simulations to study small-scale challenges (Section 4.3). Section 5 provides a summary and outlook for the prospects of Milky Way simulations.

2.1 Large-Scale Cosmic Structures

The theoretical framework of modern cosmology is built upon the cosmological principle and general relativity. The cosmological principle states that the universe is homogeneous and isotropic on large scales; general relativity holds that the spacetime structure of the universe is determined by the matter and energy within it, i.e., "matter tells spacetime how to curve, spacetime tells matter how to move." The most widely accepted standard model in modern cosmology is the ΛCDM (Cosmological Constant-Cold Dark Matter) model, which posits that the universe was born from a singularity state of infinite temperature and density. Primordial quantum fluctuations led to perturbations in the matter density distribution, which were rapidly expanded to macroscopic scales through inflation, becoming the seeds of structure formation. Under gravitational instability, high-density regions gradually accreted matter from low-density regions, becoming increasingly dense, while low-density regions became voids, eventually forming a web-like large-scale structure.

The Friedmann equation describes how the universe evolves over time and is the fundamental equation of the standard cosmological model, expressed as:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda c^2}{3}$$

where a is the scale factor, representing the size of cosmic space relative to the current cosmic scale, ρ and P are the energy density and pressure of matter in the universe respectively, G is the gravitational constant, Λ is the cosmological constant, and K is the spatial curvature parameter. Modern cosmological observations indicate that the universe is most likely flat [3–8], i.e., $K = 0$. Today's universe contains three components: baryonic matter forming observable galaxies and intergalactic medium, accounting for about 5% of the total; dark matter dominates the evolution of large-scale cosmic structures, forming the skeleton of the entire cosmic structure—the cosmic web, accounting for 27% of the total energy; the remaining approximately 68% is dark energy, which is the source of the universe's accelerated expansion and is simply treated as the cosmological constant in ΛCDM. As the universe continues to evolve, different energy components dominate different epochs (see Figure 1 [Figure 1: see original paper]). When redshift $z < z_{\text{eq}} \simeq 3500$, matter gradually dominates the cosmic evolution process, forming today's large-scale structures. Until $z \simeq 0.5$, about 5 billion years ago, dark energy began to drive the accelerated expansion of the universe.

Note: The shaded region represents dark energy with equation of state coefficient $w = -1 \pm 0.2$, calculated using natural units $c = 1$.

Figure 1 Evolution of energy density of radiation, matter, and dark energy with redshift.

In the early matter-dominated era, matter can be treated as a non-relativistic ideal fluid. Considering cosmic expansion effects, we define comoving coordinates $\mathbf{x} = \mathbf{r}/a(t)$, whose time evolution is described jointly by the continuity equation, Euler equation, and Poisson equation (see equations (2)–(4)):

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

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{\rho}\nabla P - \nabla \Phi,$$

$$\nabla^2 \Phi = 4\pi G a^2 (\rho - \bar{\rho})$$

where ∇ denotes the differential operator with respect to comoving coordinates $\mathbf{x}$, ρ is the mass density, $\bar{\rho}$ is the mean mass density of the universe, $\mathbf{v}$ is the comoving velocity, P is the pressure, and Φ is the gravitational potential in comoving coordinates. Compared with the mean density, the density field fluctuations during this period are small and can be studied through linear approximation. Under first-order linear approximation, i.e., $\rho = \bar{\rho}(1 + \delta)$, where δ is the density perturbation, we obtain its evolution law:

$$\delta(\mathbf{x}, t) = D_{\pm}(t)\delta(\mathbf{x})$$

where $D_{\pm}(t)$ are the growth and decay factors of density perturbations. Since the decay mode of density perturbations disappears rapidly with cosmic expansion, if we only consider the growth mode, equation (5) shows that during the linear evolution stage, the evolution of density perturbations with time is self-similar, i.e., the shape remains unchanged in comoving coordinates while the amplitude continues to grow.

Once the density fluctuation $\delta \approx 1$, linear evolution no longer applies and the evolution process enters the non-linear stage. At this point, matter in high-density regions begins to continuously aggregate into groups. The evolution in this stage is extremely complex. In the early period, several widely used simplified models can be obtained approximately, such as the spherical collapse model and the Zeldovich approximation. When the dark matter system reaches a non-linear quasi-equilibrium state, it forms the first small dark matter halos (simply called dark halos). These dark halos continuously accrete surrounding matter and merge with other dark halos to form larger dark halos, with masses reaching $10^{10} M_{\odot}$ or more. During the merger process, some slightly larger dark halos are not tidally disrupted after merging, but continue to exist in the larger dark halo as relatively gravitationally bound clumps, called subhalos. The merger history of dark halos can be well described by merger trees. Figure 2 [Figure 2: see original paper] shows how massive dark halos acquire mass and subhalos through mergers, with the left side representing the main branch of the merger tree and the right side representing subhalos containing subhalos. Dark matter forms the skeleton of today's large-scale structures through this "from small to large" growth pattern, which is the hierarchical clustering model of structure formation. For more details about dark halos and subhalos, please refer to the literature [11], which will not be discussed further here.

2.2 N-body Simulations

When matter enters the non-linear evolution stage, structure formation and evolution become extremely complex and difficult to describe precisely using analytical methods. Therefore, numerical simulations play a crucial role in understanding large-scale cosmic structures and their evolution. Angulo and Hahn [12] systematically introduced dark matter simulations of large-scale cosmic structures, including numerical methods, initial conditions, other dark matter candidates, and statistical methods linking simulations with observations. Considering that dark matter is basically only subject to gravitational interactions, N-body simulations discretize the dark matter density field into collisionless particles, assign masses and initial velocities to dark matter particles, and describe the evolution of dark matter by tracking the trajectories of particles. Considering the effect of background cosmic expansion, the equation of motion for a single dark matter particle can be expressed as:

$$\frac{d^2\mathbf{x}}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\mathbf{x}}{dt} = -\frac{1}{a^3}\nabla\Phi$$

Here, the evolution of the scale factor a can be obtained through equation (1), $\mathbf{g} = -\nabla\Phi$ is the gravitational force per unit mass experienced by the particle, and Φ can be obtained through equation (4), which is the key to solving the particle's equation of motion. Feng and Zhu [13] and Tang and Lin [14] provided detailed descriptions of solving N-body problems, which will only be briefly introduced here.

Figure 2 Merger tree, dark halo and its subhalos.

(1) Direct Summation Method

The direct summation method uses Newton's law of universal gravitation to calculate the gravitational force on each particle from other particles, i.e., $\mathbf{g}_i = \sum_j Gm_j\mathbf{r}_{ij}/r_{ij}^3$, where m_j is the mass of particle j and $\mathbf{r}_{ij}$ is the position vector from particle i to particle j. This method leads to significant two-body scattering effects when particles encounter each other, generally requiring the introduction of a softening length ϵ to smooth particle motion. For example, the Plummer form: $\mathbf{g}_i = Gm_j\mathbf{r}_{ij}/(r_{ij}^2 + \epsilon^2)^{3/2}$, makes it closer to the real situation. The direct summation method has high computational accuracy, but is limited by its huge computational complexity $O(N^2)$, making it suitable only for small-scale calculations.

(2) Particle Mesh Method [15]

The particle mesh method has faster computational efficiency with complexity $O(N \log N)$, where N is the number of grids. This method first grids space, and through specific assignment schemes such as NGP (nearest grid point), CIC (cloud in cell), and TSC (triangular shaped cloud), distributes particle masses to different grid points (see Figure 3 [Figure 3: see original paper]). It then solves the Poisson equation in Fourier space on the grid points, $\mathbf{k}^2\tilde{\Phi}(\mathbf{k}) = 4\pi G\tilde{\rho}(\mathbf{k})$, obtains the gravitational potential in real space through inverse Fourier transform, and uses finite difference approximation to obtain the gravitational force on the grid points. Finally, it uses the same assignment scheme to distribute the gravitational force from the grid points to particles. The particle mesh method is insufficient for dealing with situations where large numbers of particles gather. An improved method is adaptive mesh refinement, which automatically adjusts the grid scale according to the density field instead of using a fixed grid scale. This method has been applied in various simulation programs [16, 17].

Note: In the figure above, red dots are original particles, and the three lower panels show the distribution of particles on grid points under different assignment schemes, with color depth indicating the mass fraction assigned to each grid point.

Figure 3 Schematic diagram of mass assignment schemes in two dimensions.

(3) Hierarchical Tree Method [18]

The hierarchical tree method uses an octree to divide space layer by layer into small blocks until each block contains at most one particle, forming a hierarchical tree structure (see Figure 4 [Figure 4: see original paper]). If a node at a certain level satisfies the set criterion $r/d < \theta_c$ (where r is the size of the node, d is the distance from the node's center of mass to the particle experiencing the force, and θ_c is the accuracy parameter called the opening angle), then the particles within that node are combined into a particle cluster to calculate the gravitational force on the particle experiencing the force. The gravitational force on the particle from the entire space is calculated by traversing the tree structure. The advantage of this method is that computational error can be controlled by adjusting the number of expansion levels and the opening angle, with complexity $O(N \log N)$.

(4) Fast Multipole Method [19, 20]

The fast multipole method is an extension of the hierarchical tree method, further reducing the complexity to $O(N)$. Through multipole expansion of the gravitational potential, particle interactions are aggregated layer by layer onto nodes, then the effects of nodes are transferred to other nodes, and finally passed from nodes to the particles experiencing the force. Compared with the hierarchical tree method, this method requires higher-order expansion to achieve the same accuracy. The PKDGRAV3 program [21] uses this method to calculate gravity for 2×10^{12} particles.

Since different methods have accuracy differences at different distances, in practical applications the gravitational force is often decomposed into short-range and long-range components, i.e., $\Phi = \Phi_{\text{near}} + \Phi_{\text{far}}$, using different methods to calculate the gravitational force to balance accuracy and computational cost. For example, Springel in 2005 combined the hierarchical tree method with the particle mesh method to launch the simulation program GADGET2 [22], using direct summation in closer regions; GADGET4 [23] combines the fast multipole method with the particle mesh method.

Note: a) The red part represents the hierarchical tree method, with large red dots indicating the center of mass of the node, the size of the node being r, the distance from the center of mass to the particle experiencing the force being d, and the opening angle of the node for this particle being $\theta = r/d$; the blue part represents the fast multipole method, with blue dots indicating the center of mass of the node, where particles within the node transfer forces through the center of mass to other nodes; the green part corresponds to the tree structure in panel b). In panel b), white dots represent parent nodes, black dots are leaf nodes, and each leaf node represents a particle.

Figure 4 Schematic diagram of hierarchical tree method and fast multipole method in two dimensions.

3.1 Galaxy Formation

The mainstream framework of galaxy formation is the two-stage theory proposed by White and Rees [24]. In the first stage, dark matter components first cluster under gravity to form dark halos; in the second stage, dark halos continuously accrete surrounding gas, which cools rapidly in the gravitational potential well produced by the dark halo. When the pressure decreases to the point where it can no longer support gravity, the gas begins to fall into the center of the dark halo, becoming a rotationally supported gas cloud, which further forms a proto-galaxy under gravity. The mass of proto-galaxies is usually very small, about several million solar masses, and it is difficult to form observed massive galaxies through pure accretion of surrounding gas and dust. These small-mass proto-galaxies gradually gather under gravity and eventually collide and merge into a larger-mass galaxy.

When dark halos merge, the massive galaxy becomes the central galaxy of the new dark halo, while other smaller galaxies become satellite galaxies orbiting around the halo center. Due to dynamical friction, satellite galaxies may gradually lose energy and angular momentum, then fall into the center of the dark halo and merge with the central galaxy.

The two-stage theory can well explain the observed types of galaxies: when spiral galaxies of similar size merge, their spiral structures are easily destroyed by collisions, eventually forming an elliptical galaxy; while if a large spiral galaxy merges with a small dwarf galaxy, the result may simply be a larger spiral galaxy, almost the same as the original shape. During galaxy evolution, numerous physical effects affect galaxy morphology and scaling relations, such as stellar evolution and feedback, metal enrichment, radiative transfer, supermassive black holes, and quasar feedback. Fall and Efstathiou [25] introduced angular momentum on this basis to explain the formation process and observational characteristics of disk galaxies. White and Frenk [26] further summarized the galaxy formation process under the hierarchical clustering mechanism, finding that radiative cooling, star formation processes, and feedback have significant effects on galaxy formation and star formation rates, and by analyzing the galaxy luminosity function, roughly reproduced the relationship between characteristic luminosities of galaxies and galaxy clusters and galaxy properties.

3.2 Hydrodynamical Simulations

Since N-body simulations are only applicable to solving gravitational interactions of non-collisional particles and cannot resolve the complex baryonic physical processes during galaxy formation and evolution, it is generally necessary to introduce baryonic matter on the basis of N-body simulations and use hydrodynamical simulations to reconstruct the process of galaxy formation and evolution. Baryonic matter is mostly gas, composed of H and He elements, and is usually treated as a non-viscous ideal gas in hydrodynamical simulations, described jointly by the continuity equation (equation (2)), Euler equation (equation (3)), and the first law of thermodynamics (equation (7)):

$$\frac{\partial(\rho u)}{\partial t} + \mathbf{v} \cdot \nabla(\rho u) = -(\rho u + P)\nabla \cdot \mathbf{v} + \frac{P}{\rho}\frac{d\rho}{dt}$$

where u is the internal energy per unit mass. The ideal gas equation of state is: $P = (\gamma - 1)\rho u$, where γ is the adiabatic index, taken as 5/3 for monatomic gas.

Methods for solving the above equations in hydrodynamical simulations are divided into two categories: one is the grid-based Eulerian method, and the other is the Lagrangian method based on smoothed particle hydrodynamics [27–30] (SPH). Figure 5 [Figure 5: see original paper] shows the differences between Eulerian and Lagrangian methods at two adjacent time steps.

Note: a) represents the Eulerian method; b) represents the Lagrangian method.

Figure 5 Schematic diagram of Eulerian and Lagrangian methods at two adjacent time steps.

The Eulerian method uses the finite volume method to divide the entire space into fine grids, with fluid properties represented by physical quantities on the grids, and hydrodynamical equations calculated through finite difference methods. Hydrodynamical equations can generally be expressed as:

$$\frac{\partial \mathbf{q}}{\partial t} + \nabla \cdot \mathbf{F} = 0$$

where $\mathbf{q}$ represents mass density ρ, momentum density $\rho\mathbf{v}$, or total energy density ρu, and $\mathbf{F}$ is the corresponding flux density. Applying finite differences to equation (8) yields the physical quantity at the next time step:

$$\mathbf{q}(\mathbf{n}, t + \Delta t) = \mathbf{q}(\mathbf{n}, t) + \frac{\Delta t}{\Delta x} \sum_{k=x,y,z} [F_k(\mathbf{n} + 1/2, t) - F_k(\mathbf{n} - 1/2, t)]$$

Here, $\mathbf{n} = (n_1, n_2, n_3)$ is the grid coordinate, $\mathbf{n} \pm 1/2$ represents grid boundaries, and Δt and Δx are the time step and grid scale respectively. From equation (9), the key to obtaining the physical quantity at the next time step $\mathbf{q}(\mathbf{n}, t + \Delta t)$ is knowing the flux density at grid boundaries $\mathbf{F}(\mathbf{n} \pm 1/2, t)$. A simple method is to use the median of adjacent grids, i.e., $\mathbf{F}(\mathbf{n} + 1/2) = [\mathbf{F}(\mathbf{n} + 1) + \mathbf{F}(\mathbf{n})]/2$. However, since gas velocity during accretion can easily exceed the sound speed and produce shocks, causing discontinuities in physical quantities such as density and temperature at boundaries, using the median leads to unstable results. One approach is to introduce artificial viscosity to handle shocks [31], but this method has low accuracy and may introduce additional numerical errors; another more widely used method is to use the piecewise parabolic method [32] (PPM) to reconstruct the fluid field through interpolation in adjacent grids. However, since baryonic physics involves a very wide range of spatial scales, a single grid cannot meet the resolution requirements. Therefore, people proposed the adaptive mesh refinement method [33, 34] and applied it to cosmological simulations. This method can adjust grid resolution according to regional density, ensuring computational accuracy while saving computational resources, becoming the mainstream method for solving hydrodynamical equations using the Eulerian method. Widely used program packages include RAMSES [17] and ART [35]. Trac and Pen [36] provided a detailed introduction to Eulerian methods for computational fluid dynamics in astrophysics.

The Lagrangian method is a mesh-free method that describes fluids using particles, with the most commonly used being the smoothed particle hydrodynamics method. This method uses Monte Carlo random sampling to trace fluid particle mass elements, following the motion of tracer particles to obtain changes in physical quantities, and can therefore be combined with N-body methods. Even without artificial viscosity, conservation laws can be satisfied. In addition, since resolution is automatically adjusted with particle mass, it has good resolution in high-density regions, although introducing artificial viscosity at shocks suppresses turbulence formation and leads to decreased accuracy and resolution. The SPH method uses a smoothing kernel for nearby particles to obtain dynamical physical quantities. For a physical quantity $A(\mathbf{r})$, we have:

$$\langle A(\mathbf{r})\rangle = \int d^3r' A(\mathbf{r}')W(\mathbf{r}-\mathbf{r}', h)$$

where $W(\mathbf{r}-\mathbf{r}', h)$ is the smoothing kernel function and h is the smoothing length. Assuming we know the physical quantity A_j of tracer particle j, after discretization of equation (10) we obtain:

$$\langle A(\mathbf{r})\rangle = \sum_j \frac{m_j}{\rho_j} A_j W(\mathbf{r}-\mathbf{r}_j, h)$$

where m_j and ρ_j are the mass and density of particle j respectively. Letting $A(\mathbf{r}) = \rho(\mathbf{r})$, we have $\langle \rho(\mathbf{r})\rangle = \sum_j m_j W(\mathbf{r}-\mathbf{r}_j, h)$, through which the density field can be determined from the coordinates and masses of tracer particles. Generally, the smoothing length h is a function of particle and time, i.e., $h = h_i(t)$. The basis for determining the smoothing function is the number density of particles in the neighborhood of the tracer particle, for example, selecting a radius containing a fixed number of particles, although this is computationally expensive and often floats within a certain range. The adjustability of the smoothing length means that spatial resolution can be adjusted according to circumstances, which is precisely the advantage of the SPH method. However, traditional methods still have some problems, such as over-smoothing of shocks and discontinuities, and slow numerical convergence. Bauer and Springel [37] found that the SPH method is prone to large errors in gradient calculations, leading to incorrect results for turbulence in subsonic regions. Agertz et al. [38] compared the SPH method with the Eulerian method and found that SPH can hardly resolve and handle dynamical instabilities of gas. To address this problem, Beck et al. [39] adopted methods such as artificial thermal dissipation, time-dependent artificial viscosity, and high-order smoothing functions. Hopkins [40] proposed a new pressure-entropy formulation that can simultaneously conserve energy and entropy to resolve instabilities. This new P-SPH method has been widely applied in many simulation programs [41–43]. The SPH method is a very popular method in cosmological hydrodynamical simulations and is widely used in simulation programs including GADGET [22, 23], GASOLINE [44, 45], and CHANGA [46]. Detailed principles of the SPH method and its applications in astrophysics can be found in references [29, 30].

Since Eulerian and Lagrangian methods each have their own advantages and disadvantages when dealing with fluid problems, a hybrid method combining the two has been proposed. The moving mesh method represented by the AREPO program [47, 48] not only has the advantage of Eulerian methods in accurately handling shocks and fluid instability problems, but can also adjust spatial resolution and satisfy physical conservation laws like Lagrangian methods. This method uses Voronoi tessellation to fill space (see Figure 6 [Figure 6: see original paper]), where particles can move freely and continuously change the shape of the mesh as the particles move, enabling better tracking of fluid motion, thereby reducing numerical diffusion and advection errors and improving the handling of contact discontinuities. Duffell and MacFadyen [50] and Vandenbroucke and Rijcke [51] successively applied such methods to cosmological simulation programs. Hopkins [52] first introduced meshless finite-mass and meshless finite-volume methods into numerical simulations of galaxy formation and applied them in the GIZMO code. This method sets a specific smoothing kernel function $W(\mathbf{x} - \mathbf{x}_i, h(\mathbf{x}))$ for each particle to divide space, forming a Voronoi tessellation with smooth boundaries (see Figure 7 [Figure 7: see original paper]), and solves fluid equations by integrating over the volume of each particle or cell. Compared with traditional SPH, this method can better capture mixing flow instabilities, greatly reduce numerical viscosity effects, and is more accurate when calculating the evolution of subsonic flows and shocks. Compared with fixed-grid Eulerian methods, this method has the advantages of automatically and continuously adjusting resolution, reducing advection errors, and satisfying conservation laws.

Note: a) shows the Voronoi tessellation of 64 discrete points; b) shows the corresponding Delaunay tessellation; c) shows both tessellations simultaneously, with solid lines representing Voronoi tessellation and dashed lines representing Delaunay tessellation.

Figure 6 Schematic diagram of Voronoi and Delaunay tessellations under periodic boundary conditions in two dimensions.

Note: Black circles represent particles, with different colors representing particle regions. a) shows the meshless finite-volume and finite-mass methods, where corresponding particle regions are obtained by setting smoothing kernel functions on each particle; b) shows the moving mesh method, where the boundaries of divided particle regions are clear; c) shows the SPH method, where space division is centered on particles.

Figure 7 Schematic diagram of space division under periodic boundary conditions for three different methods.

3.3 Baryonic Physics Mechanisms

One of the key goals of cosmological numerical simulations is to reproduce the properties of observed galaxies. Galaxy properties are the result of the combined action of numerous astrophysical processes, including gas cooling, interstellar medium effects, star formation, supernova feedback, active galactic nucleus feedback, chemical evolution, etc. Due to limited resolution in numerical simulations, these processes cannot be simulated directly and are usually modeled using subgrid models. Common subgrid models include the EAGLE model [53, 54], IllustrisTNG model [55–57], FIRE [58–60] model, and SIMBA model [61]. These models usually require large amounts of observational data to calibrate parameters to match the characteristics of real galaxies. The most basic observational constraints are adjusting star formation rates through the galaxy stellar mass function and adjusting active galactic nucleus feedback efficiency through the black hole-stellar mass relation. Other constraints include galaxy size, hot gas mass fraction, etc. [62–64]. Different models usually need to be compared through other uncalibrated and easily observable galaxy properties to ensure model reliability, such as galaxy color distribution, scaling relations, and various properties of the intergalactic medium [65]. Wright et al. [66] compared the baryon cycles of EAGLE, IllustrisTNG, and SIMBA simulations, and by analyzing the intensity of gas flows inside and around central galaxies, pointed out that although the simulations can obtain roughly consistent stellar mass content and star formation rates at $z \approx 0$, this consistency may be produced by different physical reasons.

3.3.1 Gas Cooling

Gas cooling is key to galaxy formation. As gas falls into dark halos, potential energy is converted into kinetic energy, and velocity and pressure continuously increase, causing hot gas to no longer continue collapsing. Subsequently, these gases gradually dissipate internal energy and cool mainly through two-body radiative processes such as collisional excitation and ionization, bremsstrahlung, and inverse Compton scattering. For a completely spherically symmetric gas halo, the cooling timescale at radius r is:

$$t_{\text{cool}} = \frac{3\mu m_p k_B T}{2\rho_g(r)\Lambda(T,Z)}$$

where μm_p is the mean molecular mass, k_B is the Boltzmann constant, T is the gas temperature, $\rho_g(r)$ is the radial gas density profile, and $\Lambda(T,Z)$ is the cooling function, including various cooling mechanisms, generally related to temperature and metallicity. In cosmological simulations, it is usually assumed that gas is optically thin overall and in ionization equilibrium [53, 67–69], which is a good approximation for the period after reionization. However, Puchwein et al. [70], Oppenheimer et al. [71, 72], and Bieri et al. [73] also studied non-ionization equilibrium states of multiphase intergalactic medium. Some models (such as TNG and FIRE) also consider self-shielding effects of dense gas, where background radiation does not easily penetrate into the interior of the gas, making the gas cool faster. Different models use different methods to calculate cooling functions to regulate cooling efficiency. For example, FIRE-2 [59] considers 13 cooling mechanisms, including molecular cooling in low-temperature regions and the effect of cosmic rays on gas cooling, obtaining their respective cooling functions with a computable temperature range from 10 K to 10^{10} K. The FIRE-3 model [60] models the cold-phase molecular gas in more detail. As metallic elements in the gas increase, metal line cooling excited by heavy element particles in the galaxy radiation field dominates in gas with temperature ranges of $10^5 \sim 10^7$ K. EAGLE adopts the radiative cooling model of Wiersma et al. [74], using the photoionization software package CLOUDY to calculate cooling functions for 11 chemical elements separately. The Illustris model [56, 75] considers that cooling of individual elements has large uncertainties and directly calculates metal line cooling based on overall solar gas composition, with primordial gas cooling calculated directly through ionization equations. The subsequent TNG model still uses this method.

3.3.2 Star Formation

Star formation is the most important process in galaxy formation. When peripheral gas gradually cools and falls into the center of dark halos, forming huge dense molecular clouds, these cold molecular clouds fragment into smaller dense cores under gravitational instability, which further collapse to form stars. Through observations of the nearby universe, people find that only about 1% of cold molecular gas forms stars [76–80], and even within the dynamical time of galaxies, only about 2% of gas clouds become stars [81]. In numerical simulations, due to insufficient resolution to model different phases of the interstellar medium and star formation processes in detail, a portion of gas is usually converted into collisionless star particles through random sampling according to the initial stellar mass function, forming stellar populations with the same metallicity that evolve together. Although most models (such as EAGLE, TNG) set a gas density threshold as the condition for the start of star formation, different subgrid models have different conditions for gas to form stars. For example, the FIRE-2 model sets four criteria for gas to form stars, including self-gravity (collapsing under its own gravity), self-shielding, Jeans instability, and high density; FIRE-3 reduces this to three criteria, merging self-shielding and high density into cohesive flow, i.e., gas undergoing star formation should be gathering inward; the FLAMINGO model [68] uses gas density, overdensity, and pressure as conditions for star formation. The star formation rate is generally calculated through the Kennicutt-Schmidt law $\dot{\rho}_\star = \epsilon \rho_{\rm gas}/t_{\rm ff}$, where $\rho_{\rm gas}$ is the gas density, $t_{\rm ff}$ is the gas free-fall time, and ϵ is the gas conversion efficiency, usually set to $0.01 \sim 1$ according to star formation conditions and observational results in simulations [59, 61, 82, 83]. Since gas conversion efficiency affects long-term star formation, different models in practical applications usually adopt other methods to satisfy the Kennicutt-Schmidt law. For example, the EAGLE model makes the star formation rate depend on pressure rather than density, while TNG adopts the model proposed by Springel and Hernquist [82], using an effective equation of state to analyze the relationship between temperature and density of the interstellar medium.

3.3.3 Stellar Feedback

The interaction between stars and surrounding gas is the main means of regulating star formation. Stellar feedback mainly includes two stages: the first stage is continuous mass loss caused by photoionization effects from young massive stars and stellar winds from AGB stars during stellar evolution [84]; the second stage is at the end of stellar life, when supernova explosions feed back large amounts of energy, momentum, and heavy elements to the surrounding interstellar space, thereby increasing the metallicity of the interstellar medium and regulating gas star formation. Massive stars ($m > 13M_{\odot}$ generally return most of their mass to the interstellar medium through core-collapse supernova explosions, while low-mass stars do so through AGB stellar winds. The mass loss rate is usually a function of stellar age and metallicity, with stellar winds carrying some mass, momentum, and metals emitted to gas cells around the star particle. The EAGLE model determines the share of stellar mass loss allocated to gas based on the distance from the gas to the star particle, with momentum and energy ensuring conservation by adjusting the velocity and entropy of gas particles. The FIRE-2 model uses the same particle interaction surface as in hydrodynamical simulations to define weights for calculating the mass and momentum obtained by different gas particles. The TNG model defines the stellar recycling rate $f_{\rm rec}(m, Z)$ as the proportion of mass returned from star particles to the interstellar medium, calculating the total returned mass and different metal elements.

In the supernova feedback stage, subgrid models usually transfer large amounts of energy to surrounding gas in the form of thermal or kinetic energy. Models using thermal feedback, such as EAGLE, directly and randomly assign energy to surrounding gas particles. When gas particles receive feedback energy, the gas temperature increases accordingly. This method can avoid excessive cooling leading to overly uniform energy distribution by freely adjusting the probability $f_{\rm th}$ of gas particles being heated. Another method using kinetic feedback was proposed by Springel and Hernquist [82] and applied in the TNG model. This method diffuses energy through galactic winds, during which wind particles are no longer subject to dynamical interactions. When the density of wind particles is too low or after sufficient time has passed, wind particles merge with the gas particles they are in, completing the feedback process. The mass loading factor $\eta \equiv \dot{M}_{\rm wind}/\dot{M}_{\star}$ and wind velocity $v_{\rm wind}$ together determine the mass and degree of influence of surrounding gas affected by the wind. The wind velocity is determined by the local dark matter velocity dispersion, while setting a minimum velocity; the wind energy is determined by the metallicity of the gas cell where stars are forming, with gas cells of higher metallicity having lower wind energy. Wind particles in the TNG model also carry some thermal energy that dissipates during travel, effectively avoiding unphysical star formation and thereby affecting the stellar mass component of the wind. This method is also adopted by the SIMBA simulation [61] and the ASTRID simulation [43]. The FIRE-2 model differs from the above centralized processing of large amounts of supernova feedback by probabilistically determining whether supernova events occur within a certain time step, and then distributing the mass, metal mass, energy, and momentum of supernova feedback isotropically to the surroundings using a precisely conserved algorithm. At the same time, the FIRE-2 model considers the ionization and heating effects of stellar luminosity on gas and radiation pressure.

3.3.4 Active Galactic Nucleus Feedback

Supermassive black holes have been observed at the centers of massive galaxies and disk galaxies [85, 86]. Black holes convert large amounts of energy and matter into high-energy radiation, high-speed jets, and outflows through accretion, which propagate to surrounding regions of several megaparsecs. This entire accretion system is the active galactic nucleus (AGN) [87]. AGN feedback is the process by which AGN ejects large amounts of high-energy matter and energy to surrounding gas, heating the gas and driving it out of the galaxy, thereby affecting subsequent accretion processes of the black hole and regulating star formation and galaxy evolution [88]. Since the origin of supermassive black holes is not yet understood, numerical simulations usually set black hole seeds of $(10^4 \sim 10^5)M_\odot$ in dark matter halos with masses of about $(10^{10} \sim 10^{11})M_\odot$. The seeds continuously accrete mass from surrounding gas and evolve into massive black holes. The black hole accretion rate is usually the Bondi-Hoyle accretion rate limited by the Eddington limit:

$$\dot{M}_{\rm BH} = \frac{4\pi\alpha G^2 M_{\rm BH}^2 \rho}{(c_s^2 + v_{\rm rel}^2)^{3/2}}$$

where ρ and c_s are the gas density and sound speed respectively, and $v_{\rm rel}$ is the gas velocity relative to the black hole. Another way for black hole mass growth is merging. When two black holes are sufficiently close, usually within the smoothing length of the black hole, they merge into one more massive black hole; in numerical simulations, black hole merging is instantaneous, generally without considering relativistic effects.

AGN feedback mainly has two modes: quasar mode and radio mode [87, 89], as shown in Figure 8 [Figure 8: see original paper]. The quasar mode mainly heats and ionizes surrounding gas through high-energy radiation from the black hole accretion disk, with radiation energy usually proportional to the accretion rate: $\dot{E} = \epsilon_f \epsilon_r \dot{M}_{\rm BH} c^2$, where ϵ_f is the efficiency of radiation energy deposition into surrounding gas, and ϵ_r is the radiation efficiency of the accretion disk. The radio mode involves interaction between high-energy relativistic jet particles and gas, with kinetic energy several orders of magnitude higher than the radiation energy of the AGN, even capable of driving gas out of the galaxy, playing an important role in regulating star formation in massive galaxies. In subgrid models, the EAGLE model applies the quasar mode, randomly distributing radiation energy to surrounding gas particles. The FIRE-3 model mainly uses the radio mode, where for non-relativistic feedback, the outflow carries mass, energy, and momentum and is emitted from the accretion disk to surrounding gas at a specific ejection velocity, while relativistic jets are treated as cosmic rays with energy equal to one-thousandth of the black hole accretion mass. The TNG model incorporates both modes simultaneously. In low-accretion-rate states, AGN feedback uses the radio mode, where black hole-driven winds randomly feed energy and momentum to gas in certain directions; in high-accretion-rate quasar mode, surrounding gas is heated. In addition, as a strong ionizing radiation source, the radiation field around AGN differs from the uniform cosmological background radiation. The radiation mode of AGN feedback heats and ionizes distant gas through radiative transfer, which is considered in both the Illustris [56, 75] and FIRE-3 models.

Note: a) Quasar mode, usually appearing in bright AGN; b) Radio mode, usually appearing in low-excitation radio AGN.

Figure 8 Two main modes of AGN outflow feedback.

4.1 Zoom-in Simulations

A major challenge facing cosmological numerical simulations is the difficulty of fully covering both the dynamical evolution of galaxy disks on 1 kpc scales and the statistical distribution of galaxies in the universe on hundreds of Mpc scales. Generally, cosmological simulations have large volumes but low resolution, with baryonic particle masses above $10^6 M_\odot$ and spatial resolutions at the hundreds of pc level, which is clearly insufficient to describe baryonic physical processes at galaxy scales in detail. Zoom-in simulations can simultaneously solve the problems of low resolution and high computational cost. Zoom-in simulations select regions of interest (such as dark halos) from cosmological simulations for secondary simulations, adding more particles to increase resolution and resolve subhalos, galaxy shapes, and internal structures, while other regions maintain lower resolution. Low-resolution regions serve as the background for high-resolution regions, providing only long-range gravitational effects, thereby better helping us understand the formation and evolution processes of dark halos and galaxies. Figure 9 [Figure 9: see original paper] compares the mass resolution and spatial resolution of the latest cosmological hydrodynamical simulations and zoom-in simulations.

Note: Hollow symbols represent zoom-in simulations, solid symbols represent cosmological hydrodynamical simulations. For Lagrangian simulations, spatial resolution is defined as the minimum softening length of particles; for Eulerian simulations, it is defined as the minimum grid scale. For zoom-in simulations, the effective box size is defined as 5 times the virial radius of the largest dark halo; for cosmological hydrodynamical simulations, the effective box size is defined as the simulation box size. In panel a), dashed lines indicate approximate particle numbers in simulations; in panel b), dashed lines indicate the ratio of effective box size to spatial resolution.

Figure 9 Relationship between resolution and effective box size [11;43;53;57;59;61;68;90–115].

The process of zoom-in simulation consists of four steps. First, select samples meeting research conditions from large-scale cosmological parent simulations, determine information of all particles in the samples, and trace particles forward to initial conditions. Then, divide the zoom-in simulation region centered on the sample, introduce small-scale power spectra, and generate initial conditions for the zoom-in simulation using the same perturbation theory. Next, add corresponding baryonic physical processes. Finally, rerun the simulation program to obtain the zoom-in simulation results of the sample.

Sample selection depends on different research objects and scientific goals. Selected samples should be representative and satisfy known observational phenomena, while also considering the computational cost of sample zoom-in simulations. Analysis by Onorbe et al. [116] points out that low-resolution dark halos should preferably exceed 10^6 particles to ensure that Lagrangian regions do not undergo large changes due to increased resolution. The definition of Lagrangian region involves the spatial positions of high-resolution particles in the simulation, with particles of different mass resolutions filled at different distances around it. During zoom-in simulation, low-resolution particles near the Lagrangian region may enter the high-resolution region, affecting the motion of high-resolution particles and causing contamination of the high-resolution region. To avoid contamination from low-resolution particles, the high-resolution region should be sufficiently large to ensure that all particles passing through this region are high-resolution particles. Onorbe et al. [116] warn that if the mass fraction of low-resolution particles in a dark halo is greater than about 2%, it will cause deviations in the density profile, shape, angular momentum of the dark halo, especially gas properties; and this contamination will deepen with increasing resolution of the zoom-in simulation. After determining the positions of particles that need increased resolution at the initial redshift, a simple geometric body containing all particles can be used to represent the Lagrangian space. Onorbe et al. [116] analyzed the computational efficiency of four different shapes of Lagrangian space. Griffin et al. [117] developed a set of tools on this basis to systematically evaluate the computational complexity of different shapes of Lagrangian space, thereby selecting better spatial shapes. Springel et al. [118] used irregular Lagrangian region shapes composed of small cubes in the pure dark matter zoom-in simulation project Aquarius.

Generating initial conditions generally involves the following steps: First, determine the simulation spatial size and resolution, select appropriate cosmological models, determine model parameters (Ω_m, Ω_Λ, Ω_b, H_0, σ_8, n_s, etc.) and initial redshift z_{init}, add small-scale matter power spectra on the basis of parent simulation initial conditions, and regenerate primordial density perturbations. Then, determine the region where the research sample is located in the parent simulation, mark the particles therein, and trace back to the initial conditions to define the Lagrangian space at the time of particle initial conditions. Finally, distribute particles in the simulation space according to resolution and density perturbations. For dark matter particles, they can generally be directly traced to initial conditions based on particle identification numbers; for gas particles and star particles, a large number of massless tracer particles are used to mark gas cells. These tracer particles do not affect the intrinsic properties of gas but only follow gas advection, recording their positions and velocities through sampling interpolation methods [119–121], known as velocity field tracer particles. This method is widely used in cosmological hydrodynamical simulations [120–124] and hydrodynamical simulations of other astrophysical processes [119, 125–127].

4.2 Milky Way Simulation Projects

The Milky Way is not only the galaxy we live in, but also the galaxy we know best and most familiar with. The Milky Way holds important significance in astrophysical research for understanding the structure, evolution, and physical processes of the universe, providing an excellent experimental site for studying the mysteries and deep laws of the universe. For example, Zhou et al. [128] used N-body simulations to study the dynamical evolution of the Milky Way-Andromeda galaxy merger. This section introduces mainstream numerical simulation projects of the Milky Way and their conclusions and focuses. Table 1 shows the main parameters of mainstream Milky Way simulations.

APOSTLE (A Project Of Simulating The Local Environment) project [41] conducts zoom-in simulations of 12 regions with Local Group dynamical properties, attempting to answer cosmic puzzles in the Local Group, namely the "missing satellite" problem, "too big to fail" problem, and "plane of satellite" problem. The APOSTLE simulation selects Local Group samples with virial masses of two dark halos in $(5 \times 10^{11} \sim 2.5 \times 10^{12}) M_\odot$, and requires the distance between the two halos to be (800 ± 200) kpc, radial velocity $(0 \sim 250)$ km s^{-1}, tangential velocity below 100 km s^{-1}, and no larger dark halos within 2.5 Mpc. The APOSTLE simulation adopts the EAGLE galaxy formation model, accurately reproducing the observed Local Group stellar mass function and satellite galaxy rotation velocity function. At the same time, the statistical properties of satellite galaxies also match observations well, and they find a Milky Way-like system with a satellite galaxy plane, indicating that the observed anisotropic characteristics of the Milky Way satellite system still conform to the ΛCDM model, possibly related to the system's accretion history. In addition, APOSTLE predicts that the relationship between stellar mass and rotation velocity is affected by the galaxy environment.

Auriga project [96] selects 30 isolated Milky Way-sized dark halos from the pure dark matter EAGLE simulation for magnetohydrodynamical zoom-in simulations to study the properties and formation mechanisms of disk galaxies. The simulated galaxies have rotationally supported disk structures and flat rotation curves, consistent with the Tully-Fisher relation, and match observations in mass-metallicity relation and current star formation rate, although early stellar masses are relatively larger and disks are thicker than the Milky Way. In addition, Auriga also studies the relationship between star formation rate and stellar mass, radial distribution of star formation, and the effect of AGN on disk size. It is worth mentioning that the Auriga project analyzes the effect of simulation resolution on galaxy properties, with results showing that galaxy properties remain basically consistent under different mass resolutions. The subsequent Aurigaia project [131] matches 6 Milky Way samples from Auriga with Gaia DR2 satellite galaxy data, then uses two different methods to generate simulated image data, further analyzing the distribution characteristics of the young outer stellar disk and the rotation of the stellar halo. In 2021, Grand et al. [132] increased the baryonic resolution of one sample to $800M_{\odot}$, analyzed radial stellar surface density profiles and star formation histories, and studied the mass completeness and radial distribution of satellite galaxies. The results show that the Auriga model can successfully resolve extremely faint galaxies and reproduce Milky Way observational data on stellar velocity dispersion, half-light radius, and visible-band luminosity.

Latte simulation [97] completely adopts the FIRE-2 subgrid model to simulate a Milky Way-sized galaxy and its satellite galaxies, aiming to study the statistical properties of dwarf galaxies. The Latte project was the first to self-consistently resolve the spatial scale of half-light radii of Milky Way satellite galaxies with stellar mass $M_{\star} \gtrsim 10^5 M_{\odot}$, with a softening radius as low as 1 pc. Its simulation results are similar to observational properties of the Milky Way and Andromeda in the Local Group in three aspects: distribution of stellar mass and velocity dispersion and their relationship, relationship between stellar mass and metallicity, and star formation history and its dependence on stellar mass. For dwarf galaxies with stellar mass $M_{\star} \gtrsim 10^6 M_{\odot}$, the simulated galaxies' star formation history and metal enrichment history show no obvious differences from observations.

ARTEMIS (Assembly of high-ResoluTion Eagle-simulations of MIlky way-type galaxieS) simulation [95] adopts the EAGLE subgrid model, providing 42 Milky Way-mass galaxy samples, using multi-sample statistics of the Milky Way to study stellar halo structure. By adjusting stellar feedback model parameters to match the observed stellar mass-halo mass relation, other stellar halo properties such as density, surface brightness, metallicity, and color radial profiles match well with observed properties of nearby galaxies. The ARTEMIS simulation distinguishes between two components formed within galaxies: in-situ stars and accreted stars. In-situ stars dominate the inner part of the galaxy, with highly flattened distribution aligned with the disk direction, and their spatial distribution is the cause of metallicity, color, and age gradients in the galaxy. This provides new insights for studying complex interactions and mergers in galaxy formation processes.

HESTIA (High-resolutions Environmental Simulations of The Immediate Area) project [94] combines the large-scale structure and proper motion velocities of the nearby universe observed by Cosmicflows-2 [133] to constrain and reset numerical simulation initial conditions. It first conducts low-resolution pure dark matter simulations, then introduces the Auriga [96] galaxy formation model to perform high-resolution zoom-in simulations on 3 samples that satisfy Local Group characteristics. HESTIA's Local Group samples not only satisfy certain conditions in halo mass, distance, isolation, mass ratio, and radial velocity, but also remain consistent with the Virgo cluster on large scales, with no additional massive galaxy clusters within 20 Mpc. HESTIA's Local Group mass accretion history differs from that in simulations without constrained initial conditions, reproducing many observational characteristics of the nearby universe such as Local Group formation, dark halo mass, stellar disk mass, galaxy morphology, satellite galaxy distribution characteristics, and the Magellanic Clouds. This provides test samples for how large-scale environments affect the Local Group.

NIHAO-LG simulation [129] generates constrained initial conditions from the CLUES (Constrained Local UniversE Simulations) project [134, 135], adopts the subgrid model of the NIHAO (Numerical Investigation of a Hundred Astrophysical Objects) project [101], and tests whether the Local Group can provide unbiased samples for galaxy formation and evolution models by comparing the similarities and differences in properties between Local Group dwarf galaxies and field galaxies. NIHAO-LG dwarf galaxies have similar stellar statistical properties to NIHAO field galaxies, such as velocity dispersion, mean stellar age, star formation rate, etc., although most dwarf galaxies in NIHAO-LG formed all their stars at earlier times and then remained "quenched." The total gas content of dwarf galaxies in NIHAO-LG is also consistent with field galaxies, indicating that environment does not affect total gas content. However, most gas is cold and distributed within $0.2R_{200}$, which is related to its higher metallicity; higher metallicity makes it cool faster, thus forming stars, but stellar feedback drives away outer gas. In addition, interactions between dwarf galaxies in NIHAO-LG simulations cause high-metallicity gas to be mainly distributed at radii greater than $0.2R_{200}$, and the distribution of high-metallicity dwarf galaxies is random, with no correlation between their metallicity and the central region of the Local Group.

4.3 Using Milky Way Simulations to Study Small-Scale Challenges

Small-scale challenges involve the distribution and structure of the Milky Way and its satellite galaxies, and Milky Way simulations have natural advantages for studying small-scale challenges due to their higher resolution and more accurate Milky Way properties. Currently, Milky Way simulations based on the standard cosmological model have made considerable progress in alleviating small-scale challenges.

4.3.1 The "Missing Satellite" Problem

The APOSTLE simulation [41] finds that within 300 kpc of Andromeda and the Milky Way, the number of satellite galaxies with stellar mass $M_\star \gtrsim 10^5 M_\odot$ is 20^{+10}_{-6} and 18^{+8}_{-5} respectively, consistent with observed numbers. Simpson et al. [136] find in Auriga simulations that the luminosity distribution of Milky Way satellite galaxies also matches observations. Garrison-Kimmel et al. [137] use FIRE simulations to obtain a median number of 15.5 satellite galaxies in the Local Group, indicating no evidence for the "missing satellite" problem or "too big to fail" problem in the FIRE sample. Kim et al. [138] find by assuming an empirical relationship between stellar mass and dark halo mass that the number of satellite galaxies calibrated using SDSS data matches the number of luminous galaxies predicted by the cold dark matter model. The "missing satellite" problem seems to be solvable within the framework of the standard cosmological model, with baryonic physical mechanisms (such as supernova feedback and gas ionization processes) suppressing star formation in low-mass dark halos, leading to satellite galaxies "disappearing."

4.3.2 The "Plane of Satellite" Problem

Numerical simulations attempt to solve the "plane of satellite" problem by improving the phase-space correlation of satellite galaxies. Satellite galaxies originate from the cosmic environment, falling in from the same direction (such as along filaments) or originating from the same process (such as collective infall of a galaxy group), having similar orbits [139]. Samuel et al. [140] find from FIRE-2 simulations that at redshift $z = 0 \sim 0.2$, 1%–2% of snapshots have thin satellite planes, 5% have dynamically correlated planes, and the presence of the Large Magellanic Cloud increases this to 7%–16%, although there is no significant difference between isolated central galaxies and Local Group-like galaxies. Shao et al. [141] focus on a Milky Way-like system with a thin satellite plane in EAGLE simulations, finding that 8 out of 11 classical satellite galaxies are located in a highly clustered orbital plane, and the plane has a smaller axis ratio at satellite galaxy infall times. Shao et al. also point out that tidal torques caused by the non-spherical distribution of the main halo mass distribution guide satellite galaxy orbits onto the plane. Sawala et al. [142] use Gaia proper motion data to demonstrate that the anisotropy of Milky Way satellite galaxies is more common than reported, largely caused by unstable radial distributions, and their plane is temporary rather than rotationally supported.

4.3.3 The "Core-Cusp" Problem

Many numerical simulations show that baryonic feedback effects (such as supernova feedback) reduce the density at galaxy centers [83, 143–145]. If there is sufficient star formation in a galaxy, the energy produced by supernova explosions drives large amounts of gas away from the center, thereby changing the dark matter density profile and forming a "core" center. Benitez-Llambay et al. [146] point out that the density threshold for star formation affects the formation of the central core. If the threshold is low, more gas forms stars, leading to reduced gas that cannot dominate the gravitational potential at the center, thus having limited effect on the inner density profile. If the threshold is high, it can dominate the gravitational potential at the center before being driven away, but if there is too much gas, it will accelerate gravitational contraction and increase central density. Read et al. [83] simulation results show that if star formation lasts long enough, a "core" will form. Tollet et al. [144] use NIHAO simulations to analyze 3 dark halos with masses greater than $10^7 M_\odot$, finding that their inner density profiles hardly change with time and have flat core densities. FIRE simulations also find that in dark halos with $M_\star \approx 10^{6.3} M_\odot$, there are dark matter "cores" with constant density; in dark halos with $M_\star \approx 10^4 M_\odot$, "cusp" density profiles appear [145]. However, Bose et al. [147] find that in APOSTLE and Auriga simulations, no "cores" form regardless of the stellar mass of dwarf galaxies, and point out that star formation rate and stellar mass fraction have limited effect on the inner density profile of dark halos.

4.3.4 The "Too Big to Fail" Problem

Interactions between satellite galaxies and the Milky Way, such as tidal stripping, disk shocking, and ram pressure stripping, reduce the central mass of satellite galaxies. Tomozeiu et al. [149] simulations show that for dwarf galaxy halos with flat profiles, strong tidal stripping and mass redistribution promote their circular velocity profiles to be similar to those of Milky Way satellite galaxies. Latte simulations reproduce the number characteristics of dwarf galaxies with $M_\star \approx 10^5 M_\odot$, suggesting they are not troubled by the "too big to fail" problem. Dutton et al. [150] find that in NIHAO simulations, the circular velocities of galaxies with luminosity $L_V \gtrsim 2 \times 10^6 L_\odot$ show no systematic differences from satellite galaxies of the Milky Way and Andromeda, suggesting that other cosmological simulations may overestimate stellar mass. Similarly, the circular velocity distribution function of satellite galaxies in APOSTLE simulations is consistent with that of Milky Way satellite galaxies, while the corresponding pure dark matter simulation is higher. The reason may be that baryonic effects reduce subhalo mass, and subhalos with maximum circular velocity $v_{\max} < 30$ km s^{-1} cannot form stars or their galaxies are tidally stripped [41, 148]. It is worth mentioning that Papastergis and Shankar [151], based on observations of 21 cm line of H atoms in 90 field dwarf galaxies, argue that baryonic effects cannot solve the "too big to fail" problem. However, Verbeke et al. [152] point out that for low-mass galaxies, due to continuous turbulence in the interstellar medium caused by supernova explosions, the circular velocity profiles obtained using H atom dynamical data are lower than the real ones.

5 Summary and Outlook

Cosmological numerical simulation is one of the most powerful methods for studying galaxy formation and evolution. Over the past decade, cosmological numerical simulations have made tremendous progress in two aspects: (1) Large-volume simulations have successfully reproduced the large-scale structure of the universe while providing large samples of galaxies, helping to study the global statistical properties and scaling relations of galaxies; (2) High-resolution zoom-in simulations have revealed detailed galaxy structures, allowing more detailed study of the specific effects of astrophysical mechanisms on galaxy properties [69]. The Milky Way, as the only galaxy that can be observed in detail, has become the main research object of zoom-in simulations. Milky Way zoom-in simulations study important properties such as the Milky Way's star formation, internal structure, evolution history, and satellite galaxies, continuously deepening our understanding of the Milky Way and pointing the way for Milky Way observations, such as searching for fainter satellite galaxies.

Although different Milky Way projects may adopt different numerical calculation methods and subgrid models, they are basically consistent with observations in main statistical properties and can make predictions about Milky Way evolution. At the same time, Milky Way projects have also made considerable progress in alleviating small-scale challenges to the standard cosmological model. With the continuous advancement of numerous Milky Way observational projects at home and abroad, ground-based telescopes such as FAST, LAMOST, and space telescopes such as Gaia, JWST, as well as future telescopes like CSST and ELT, will provide more information about the Milky Way and nearby galaxies in radio, optical, and infrared bands. The internal and external structures of the Milky Way are gradually becoming clear. Based on these structural features to constrain galaxy formation model parameters, people have gained a deeper understanding of the Milky Way's formation and evolution history.

However, limited by current computational resources, current subgrid models still simplify physical processes considerably, mostly describing certain physical processes through parameterized methods. These adjustable parameters inevitably have certain degeneracies and may not truly capture the physical processes affecting galaxies or accurately reflect the real physical picture, which greatly limits the predictive power of subgrid models. One of the subsequent goals of Milky Way numerical simulations is to study how complex astrophysical processes affect simulation results to further explore the physical mechanisms of galaxy formation and evolution. To this end, on the one hand, more realistic Milky Way numerical simulations are needed, especially considering that the Milky Way may have unique characteristics, such as the peculiar spatial distribution of its satellite galaxies and the past merger of a massive galaxy (Gaia Sausage). Comparison between simulations and observations requires ensuring sample consistency, so it is necessary to simulate samples that match more Milky Way observational characteristics to obtain more realistic Milky Way simulation data.

On the other hand, with improving computational capabilities, increasing resolution has become a development trend in numerical simulations. Ultra-high-resolution simulations play an important role in revealing the internal structure of extremely low-mass galaxies and the chemical evolution of the interstellar medium. If resolution becomes sufficiently high, it will be possible to resolve the formation process of individual stars and even feedback mechanisms, fundamentally studying the physical mechanisms of galaxy formation and evolution without needing to use subgrid models for parametric modeling of physical mechanisms. Achieving this goal still requires overcoming many difficulties.

References

[1] Bullock J S, Boylan-Kolchin M. ARA&A, 2017, 55(1): 343 [2] Del Popolo A, Le Delliou M. Galaxies, 2017, 5(1): 17 [3] Planck C, Aghanim N, Akrami Y, et al. A&A, 2020, 641: A6 [4] Hinshaw G, Larson D, Komatsu E, et al. ApJS, 2013, 208(2): 19 [5] Nunes R C, Bernui A. Eur Phys J C, 2020, 80(11): 1025 [6] Liu Y, Cao S, Liu T, et al. ApJ, 2020, 901(2): 129 [7] Aiola S, Calabrese E, Maurin L, et al. JCAP, 2020, 12: 47 [8] Vardanyan M, Trotta R, Silk J. MNRAS, 2009, 397(1): 431 [9] Frieman J A, Turner M S, Huterer D. ARA&A, 2008, 46(1): 385 [10] Giocoli C, Tormen G, Sheth R K, et al. MNRAS, 2010, 404(1): 502 [11] Zavala J, Frenk C S. Galaxies, 2019, 7(4): 81 [12] Angulo R E, Hahn O. Living Rev Comput Astrophys, 2022, 8(1): 1 [13] Feng, L. & Zhu, W. Science China: Physics, Mechanics & Astronomy, 2013, 43(6): 687 [14] Tang, L. & Lin, W. Progress in Astronomy, 2018, 36(2): 136 [15] Hockney R W, Eastwood J W. Computer Simulation Using Particles. New York: McGraw-Hill, 1981: 1 [16] Bryan G L, Norman M L, O'Shea B W, et al. ApJS, 2014, 211(2): 19 [17] Teyssier R. A&A, 2002, 385(1): 337 [18] Barnes J, Hut P. Nature, 1986, 324: 446 [19] Dehnen W. Journal of Computational Physics, 2002, 179: 27 [20] Greengard L, Rokhlin V. Journal of Computational Physics, 1987, 73: 325 [21] Potter D, Stadel J, Teyssier R. Comput Astrophys, 2017, 4(1): 2 [22] Springel V. MNRAS, 2005, 364(4): 1105 [23] Springel V, Pakmor R, Zier O, et al. MNRAS, 2021, 506(2): 2871 [24] White S D M, Rees M J. MNRAS, 1978, 183: 341 [25] Fall S M, Efstathiou G. MNRAS, 1980, 193: 189 [26] White S D M, Frenk C S. ApJ, 1991, 379: 52 [27] Gingold R A, Monaghan J J. MNRAS, 1977, 181(3): 375 [28] Monaghan J J. ARA&A, 1992, 30: 543 [29] Springel V. ARA&A, 2010, 48(1): 391 [30] Bagheri M, Mohammadi M, Riazi M. Comp Part Mech, 2024, 11(3): 1163 [31] Ryu D, Vishniac E T, Chiang W-H. ApJ, 1990, 354: 389 [32] Colella P, Woodward P R. Journal of Computational Physics, 1984, 54: 174 [33] Berger M J, Oliger J. Journal of Computational Physics, 1984, 53(3): 484 [34] Berger M J, Colella P. Journal of Computational Physics, 1989, 82(1): 64 [35] Kravtsov A V, Klypin A A, Khokhlov A M. ApJS, 1997, 111: 73 [36] Trac H, Pen U. PASP, 2003, 115(805): 303 [37] Bauer A, Springel V. MNRAS, 2012, 423(3): 2558 [38] Agertz O, Moore B, Stadel J, et al. MNRAS, 2007, 380(3): 963 [39] Beck A M, Murante G, Arth A, et al. MNRAS, 2016, 455(2): 2110 [40] Hopkins P F. MNRAS, 2013, 428(4): 2840 [41] Sawala T, Frenk C S, Fattahi A, et al. MNRAS, 2016, 457(2): 1931 [42] Schaller M, Dalla Vecchia C, Schaye J, et al. MNRAS,

2015, 454(3): 2277 [43] Bird S, Ni Y, Di Matteo T, et al. MNRAS, 2022, 512(3): 3703 [44] Wadsley J W, Stadel J, Quinn T. New Astronomy, 2004, 9(2): 137 [45] Wadsley J W, Keller B W, Quinn T R. MNRAS, 2017, 471: 2357 [46] Menon H, Wesolowski L, Zheng G, et al. Comput Astrophys, 2015, 2(1): 1 [47] Weinberger R, Springel V, Pakmor R. ApJS, 2020, 248(2): 32 [48] Springel V. MNRAS, 2010, 401(2): 791 [49] Pen U-L. ApJS, 1998, 115: 19 [50] Duffell P C, MacFadyen A I. ApJS, 2011, 197(2): 15 [51] Vandenbroucke B, De Rijcke S. Astronomy and Computing, 2016, 16: 109 [52] Hopkins P F. MNRAS, 2015, 450(1): 53 [53] Schaye J, Crain R A, Bower R G, et al. MNRAS, 2015, 446: 521 [54] Crain R A, Schaye J, Bower R G, et al. MNRAS, 2015, 450: 1937 [55] Pillepich A, Springel V, Nelson D, et al. MNRAS, 2018, 473(3): 4077 [56] Vogelsberger M, Genel S, Sijacki D, et al. MNRAS, 2013, 436(4): 3031 [57] Springel V, Pakmor R, Pillepich A, et al. MNRAS, 2018, 475(1): 676 [58] Hopkins P F, Kereš D, Oñorbe J, et al. MNRAS, 2014, 445(1): 581 [59] Hopkins P F, Wetzel A, Kereš D, et al. MNRAS, 2018, 480(1): 800 [60] Hopkins P F, Wetzel A, Wheeler C, et al. MNRAS, 2023, 519: 3154 [61] Davé R, Anglés-Alcázar D, Narayanan D, et al. MNRAS, 2019, 486(2): 2827 [62] Crain R A, Schaye J, Bower R G, et al. MNRAS, 2015, 450(2): 1937 [63] McCarthy I G, Schaye J, Bird S, et al. MNRAS, 2017, 465(3): 2936 [64] Robertson B E. ARA&A, 2022, 60(1): 121 [65] Faucher-Giguère C-A. Nat Astron, 2018, 2(5): 368 [66] Wright R J, Somerville R S, Lagos C del P, et al. https://arxiv.org/pdf/2402.08408, 2024 [67] Crain R A, van de Voort F. ARA&A, 2023, 61(1): 473 [68] Schaye J, Kugel R, Schaller M, et al. MNRAS, 2023, 526(4): 4978 [69] Vogelsberger M, Marinacci F, Torrey P, et al. Nature Reviews Physics, 2020, 2: 42 [70] Puchwein E, Bolton J S, Haehnelt M G, et al. MNRAS, 2015, 450(4): 4081 [71] Oppenheimer B D, Crain R A, Schaye J, et al. MNRAS, 2016, 460(2): 2157 [72] Oppenheimer B D, Schaye J, Crain R A, et al. MNRAS, 2018, 481(1): 835 [73] Bieri R, Naab T, Geen S, et al. MNRAS, 2023, 523(4): 6336 [74] Wiersma R P C, Schaye J, Theuns T, et al. MNRAS, 2009, 399(2): 574 [75] Vogelsberger M, Genel S, Springel V, et al. MNRAS, 2014, 444(2): 1518 [76] Utomo D, Sun J, Leroy A K, et al. ApJ, 2018, 861(2): L18 [77] Evans N J, Heiderman A, Vutisalchavakul N. ApJ, 2014, 782(2): 114 [78] Lee E J, Miville-Deschênes M-A, Murray N W. ApJ, 2016, 833(2): 229 [79] Leroy A K, Walter F, Sandstrom K, et al. AJ, 2013, 146(2): 19 [80] Bigiel F, Leroy A, Walter F, et al. AJ, 2008, 136(6): 2846 [81] Kennicutt R C Jr. ApJ, 1998, 498: 541 [82] Springel V, Hernquist L. MNRAS, 2003, 339(2): 289 [83] Read J I, Agertz O, Collins M L M. MNRAS., 2016, 459(3): 2573 [84] Smith N. ARA&A, 2014, 52(1): 487 [85] Gehren T, Fried J, Wehinger P A, et al. ApJ, 1984, 278: 11 [86] Shields J C, Walcher C J, Böker T, et al. ApJ, 2008, 682(1): 104 [87] Fabian A C. ARA&A, 2012, 50(1): 455 [88] Heckman T M, Best P N. ARA&A, 2014, 52(1): 589 [89] Alexander D M, Hickox R C. New Astronomy Reviews, 2012, 56(4): 93 [90] Ramesh R, Nelson D. MNRAS, 2024, 528(2): 3320 [91] Applebaum E, Brooks A M, Christensen C R, et al. ApJ, 2021, 906(2): 96 [92] Nuñez-Castiñeyra A, Nezri E, Devriendt J, et al. MNRAS, 2021, 501(1): 62 [93] Dubois Y, Beckmann R, Bournaud F, et al. A&A, 2021, 651: A109 [94] Libeskind N I, Carlesi E, Grand R J J, et al. MNRAS, 2020, 498(2): 2968 [95] Font A S, McCarthy I G, Poole-Mckenzie

R, et al. MNRAS, 2020, 498(2): 1765 [96] Grand R J J, Gomez F A, Marinacci F, et al. MNRAS, 2017, 467: 179 [97] Wetzel A R, Hopkins P F, Kim J, et al. ApJ, 2016, 827(2): L23 [98] Roca-Fàbrega S, Valenzuela O, Colín P, et al. ApJ, 2016, 824(2): 94 [99] Fiacconi D, Madau P, Potter D, et al. ApJ, 2016, 824(2): 144 [100] Keller B W, Wadsley J, Couchman H M P. MNRAS, 2016, 463(2): 1431 [101] Wang L, Dutton A A, Stinson G S, et al. MNRAS, 2015, 454(1): 83 [102] Hernández-Aguayo C, Springel V, Pakmor R, et al. MNRAS, 2023, 524(2): 2556 [103] Feldmann R, Quataert E, Faucher-Giguère C-A, et al. MNRAS, 2023, 522(3): 3831 [104] Kannan R, Garaldi E, Smith A, et al. MNRAS, 2022, 511(3): 4005 [105] Pillepich A, Nelson D, Springel V, et al. MNRAS, 2019, 490(3): 3196 [106] Elahi P J, Welker C, Power C, et al. MNRAS, 2018, 475(4): 5338 [107] McCarthy I G, Schaye J, Bird S, et al. MNRAS, 2017, 465(3): 2936 [108] Tremmel M, Karcher M, Governato F, et al. MNRAS, 2017, 470(1): 1121 [109] Davé R, Thompson R, Hopkins P F. MNRAS, 2016, 462(3): 3265 [110] Dubois Y, Peirani S, Pichon C, et al. MNRAS, 2016, 463(4): 3948 [111] Feng Y, Di-Matteo T, Croft R A, et al. MNRAS, 2016, 455(3): 2778 [112] Bocquet S, Saro A, Dolag K, et al. MNRAS, 2016, 456(3): 2361 [113] Khandai N, Di Matteo T, Croft R, et al. MNRAS, 2015, 450(2): 1349 [114] Di Matteo T, Khandai N, DeGraf C, et al. ApJ, 2012, 745(2): L29 [115] Schaye J, Vecchia C D, Booth C M, et al. MNRAS, 2010, 402(3): 1536 [116] Onorbe J, Garrison-Kimmel S, Maller A H, et al. MNRAS, 2014, 437: 1894 [117] Griffin B F, Ji A P, Dooley G A, et al. ApJ, 2016, 818(1): 10 [118] Springel V, Wang J, Vogelsberger M, et al. MNRAS, 2008, 391: 1685 [119] Federrath C, Glover S C O, Klessen R S, et al. Phys Scr, 2008, 132: 4025 [120] Konstandin L, Federrath C, Klessen R S, et al. J Fluid Mech, 2012, 692: 183 [121] Vazza F, Gheller C, Brunetti G. A&A, 2010, 513: A32 [122] Dubois Y, Pichon C, Haehnelt M, et al. MNRAS, 2012, 423: 3616 [123] ZuHone J A, Markevitch M, Brunetti G, et al. ApJ, 2013, 762(2): 78 [124] Vazza F, Wittor D, Brunetti G, et al. A&A, 2021, 653: A23 [125] Li Y, Bryan G L. ApJ, 2014, 789: 153 [126] Bovard L, Rezzolla L. Classical Quant Grav, 2017, 34(21): 5005 [127] Tiede C, Zrake J, MacFadyen A, et al. ApJ, 2022, 932(1): 24 [128] Zhou, Y., Li, Z. & Chang, J. Progress in Astronomy, 2023, 41(4): 476 [129] Arora N, Macciò A V, Courteau S, et al. MNRAS, 2022, 512(4): 6134 [130] Agertz O, Renaud F, Feltzing S, et al. MNRAS, 2021, 503(4): 5826 [131] Grand R J J, Helly J, Fattahi A, et al. MNRAS, 2018, 481(2): 1726 [132] Grand R J J, Marinacci F, Pakmor R, et al. MNRAS, 2021, 507(4): 4953 [133] Tully R B, Courtois H M, Dolphin A E, et al. AJ, 2013, 146(4): 86 [134] Sorce J G, Gottlöber S, Yepes G, et al. MNRAS, 2016, 455(2): 2078 [135] Carlesi E, Sorce J G, Hoffman Y, et al. MNRAS, 2016, 458(1): 900 [136] Simpson C M, Grand R J J, Gómez F A, et al. MNRAS, 2018, 478(1): 548 [137] Garrison-Kimmel S, Hopkins P F, Wetzel A, et al. MNRAS, 2019, 487(1): 1380 [138] Kim S Y, Peter A H G, Hargis J R. Phys Rev Lett, 2018, 121(21): 1302 [139] Pawlowski M S. Mod Phys Lett A, 2018, 33(06): 1830004 [140] Samuel J, Wetzel A, Chapman S, et al. MNRAS, 2021, 504(1): 1379 [141] Shao S, Cautun M, Frenk C S. MNRAS, 2019, 488(1): 1166 [142] Sawala T, Cautun M, Frenk C, et al. Nat Astron, 2022, 7(4): 481 [143] Pontzen A, Governato F. MNRAS, 2012, 421(4): 3464 [144] Tollet E, Macciò A V, Dutton A A, et al. MNRAS, 2016, 456(4): 3542 [145]

Oñorbe J, Boylan-Kolchin M, Bullock J S, et al. MNRAS, 2015, 454(2): 2092 [146] Benítez-Llambay A, Frenk C S, Ludlow A D, et al. MNRAS, 2019, 488(2): 2387 [147] Bose S, Frenk C S, Jenkins A, et al. MNRAS, 2019, 486(4): 4790 [148] Fattahi A, Navarro J F, Sawala T, et al. MNRAS, 2016, 457(1): 844 [149] Tomozeiu M, Mayer L, Quinn T. ApJ, 2016, 827(1): L15 [150] Dutton A A, Macciò A V, Frings J, et al. MNRAS, 2016, 457(1): L74 [151] Papastergis E, Shankar F. A&A, 2016, 591: A58 [152] Verbeke R, Papastergis E, Ponomareva A A, et al. A&A, 2017, 607: A13

Note: Figure translations are in progress. See original paper for figures.