

MPA Lecture on Numerical Methods for Cosmology

Linda Blot

30 January 2020

Contents

1	Introduction	2
2	Simulating Dark Matter	2
2.1	Vlasov-Poisson system of equations	5
2.2	A 1D example	6
2.3	The N-body technique	9
2.4	The Lagrangian phase space technique	10
3	Gravity solvers	11
3.1	Particle-particle	11
3.2	Hierarchical Tree	12
3.3	Particle-Mesh	14
3.4	Hybrid approaches and refinement	16
4	Cosmological simulations	16
4.1	Initial Conditions	17
4.2	Identifying Halos	19
5	Statistical modelling of galaxy distribution	22
5.1	Abundance Matching	24
5.2	Halo Occupation Distribution	24

1 Introduction

In this lecture I will present the basics of numerical methods that are used for cosmological simulations. In particular, I will focus on gravity-only simulations targeted at large scale structure (LSS) studies. In this context one simulates large cubic portions of our Universe (the typical size today is of order a few Gpc per side) using random initial conditions that are compatible with CMB observations in terms of statistics of the initial density perturbations. Since we are simulating scales that always lie inside the horizon we are going to use Newtonian gravity, although some implementations of relativistic numerical methods already exist [see e.g. 3, 6]. In any case relativistic effects only affect the largest scales and can be neglected for most of cosmological applications.

While most of the LSS observables have been focused on the statistics of the density field in configuration space, modern galaxy surveys have enough galaxies for splitting them into fine redshift bins and thus the redshift space distortions due to peculiar velocities become an important part of the modelling. For this reason new techniques are being developed where more care is put into predicting the velocity field of the LSS. This is why I included here some new techniques that try to use the full phase space information (positions and velocities) both to solve for dark matter evolution and to define halos.

In Section 2 I will introduce different ways of representing the dark matter distribution in a simulation. Section 3 covers the different techniques developed to solve the Poisson equation. In Section 4 I will describe how to generate initial conditions and identify halos in cosmological simulations. Finally Section 5 outlines some simple methods one can use to populate gravity-only simulations with galaxies.

2 Simulating Dark Matter

There are two properties of dark matter that are going to be relevant for the following: dark matter is collisionless and cold. We expect dark matter to interact only gravitationally with itself, and even for self-interacting models the cross-section needs to be very small for the model to be compatible with

observations. This means that on cosmological scales we can consider dark matter as collisionless. We can imagine that some collisional effects could become important in the interior of halos, where the density of dark matter particles is very high.

To verify that this is not the case let us compute the two-body relaxation time for a dark matter halo composed of N dark matter particles. We want an order of magnitude estimate so we are going to make some rough approximations. Let us consider the trajectory of a test particle moving in the system (see Figure 1). Suppose the particle passes at a distance b

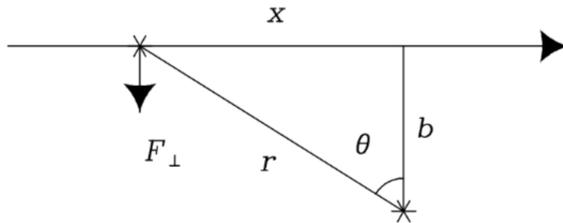


Figure 1: A test particle approaches the field particle at speed v and impact parameter b . We approximate the test particle's trajectory as a straight line. From [8].

from another particle of the system (field particle from now on). We want to compute the amount by which the encounter deflects the velocity of the test particle. We shall assume that $\delta v/v \ll 1$ and that the field particle is stationary during the encounter. This means that $\delta \mathbf{v}$ is perpendicular to \mathbf{v} since the parallel component averages out. If we approximate the trajectory of the test particle as a straight line we can compute the change in velocity by integrating the perpendicular force along the trajectory:

$$F_{\perp} = \frac{Gm^2}{r^2} \cdot \cos \theta = \frac{Gm^2}{x^2 + b^2} \cdot \cos \theta. \quad (1)$$

Using the law of cosines and that $x = vt$ we get:

$$F_{\perp} = \frac{Gm^2}{b^2} \left[1 + \left(\frac{vt}{b} \right)^2 \right]^{-3/2} \quad (2)$$

Integrating Newton's law in time:

$$m \delta v = \int_{-\infty}^{\infty} F_{\perp} dt \quad (3)$$

we arrive to:

$$\delta v = \frac{2Gm}{bv} \quad (4)$$

Eq. 4 is for one encounter. If the test particle crosses the halo once it will encounter many field particles and the deviations will add up. In the limit where the distribution of field particles is homogeneous the average deviation will be zero, but the scatter will not. So the impacts will cause a sort of diffusion that we call two-body relaxation. The surface density of field particles is:

$$\Sigma = \frac{N}{\pi R^2} \quad (5)$$

where R is the radius of the halo. The number of encounters with impact parameter between b and $b + db$ is then:

$$\delta n = \frac{N}{\pi R^2} 2\pi b db = \frac{2N}{R^2} b db. \quad (6)$$

Summing the square velocity variations over the encounters:

$$\sum \delta v^2 = \delta v^2 \delta n = \left(\frac{Gm}{vR} \right)^2 \frac{8N}{b} db \quad (7)$$

and integrating over all impact parameters we get the total square velocity variation in one crossing:

$$\Delta v^2 = \int_{b_{min}}^{b_{max}} \delta v^2 \delta n = 8N \left(\frac{Gm}{Rv} \right)^2 \log \left(\frac{b_{max}}{b_{min}} \right). \quad (8)$$

If Δv becomes of the same order as v then we can say collisions are important. The typical speed of a particle is one in a circular orbit at the edge of the halo:

$$v^2 \sim v_c^2 = \frac{GNm}{R}. \quad (9)$$

Using Eq. 9 to eliminate R from Eq. 8 we can rewrite Δv as:

$$\Delta v^2 = \frac{8 \log \Lambda v^2}{N} \quad (10)$$

where $\Lambda = b_{max}/b_{min}$. So for Δv to become of order v the test particle needs to cross the halo n_{relax} times:

$$n_{relax} = \frac{N}{8 \log \Lambda} \quad (11)$$

The relaxation time is then defined as:

$$t_{relax} = n_{relax} t_{cross} \quad (12)$$

The minimal scattering happens when the particle is at the edge of the halo so $b_{max} = R$. The maximal scattering happens when $\delta v = v$ so $b_{min} = 2R/N$. This gives:

$$t_{relax} = \frac{N}{8 \log N/2} t_{cross}. \quad (13)$$

Let's put in some numbers: for a galactic scale halo $t_{cross} \sim 1/(10H_0)$, and for 100GeV WIMP DM particles $N \sim 10^{77}$. This means that the two-body relaxation time for dark matter halos is much larger than the age of the Universe so we can consider DM particles as a collisionless fluid: we can ignore individual and short-range interactions and treat the density distribution as smooth.

2.1 Vlasov-Poisson system of equations

A self-gravitating collisionless fluid of particles of mass m is described in phase space by its distribution function $f(t, \mathbf{x}, \mathbf{p})$, where $\mathbf{p} = m\mathbf{v}$ is the momentum, and its evolution is given by the collisionless Boltzmann equation (CBE) [8]:

$$\frac{df(t, \mathbf{x}, \mathbf{p})}{dt} = \frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m a^2} \cdot \nabla_{\mathbf{x}} f - m \nabla_{\mathbf{x}} \phi \cdot \nabla_{\mathbf{p}} f = 0, \quad (14)$$

with the gravitational potential ϕ obeying the Poisson equation:

$$\nabla^2 \phi = \frac{4\pi G m_{DM}}{a} \left(\int d^3 p f - \bar{n} \right), \quad (15)$$

where m_{DM} is the microphysical dark matter mass and \bar{n} is the mean number density of DM particles in the universe. The system of collisionless Boltzmann and Poisson equations is usually called the Vlasov-Poisson system of

equation. Basically, it simply states that the phase space density is conserved. It is advected in configuration space by velocities, and in velocity space by gravitational forces. One could think of expanding the CBE in its hierarchy of moments similarly to what is done for gases, but in the collisionless case this is not possible because each moment depends on the next order moment and the hierarchy cannot be closed.

In the case of cold dark matter (CDM) the thermal velocity dispersion is much smaller than the bulk velocity due to gravitational interactions, so its velocity distribution has negligible width. This confines the evolution of f in a 3D sub-manifold of phase space (also called the DM sheet) and reduces the dimensionality of the problem, allowing to solve the Vlasov-Poisson system in 3D configuration space. In linear regime the sheet does not fold, so in each point \mathbf{x} the velocity is single-valued and DM can be treated as a perfect fluid with $P \approx 0$ [26], i.e. the hierarchy can be closed and you only need the first two moments (continuity and Euler equations). As soon as shell crossing occurs multiple streams with different velocities pass through the points where caustics form and the fluid can no longer be described by bulk quantities. In the perfect fluid description this is when the velocity dispersion starts to be non-zero and you need the full hierarchy of moments.

2.2 A 1D example

We can have a nice picture of the behavior of the DM sheet with a simple 1D example: the collapse of a plane wave. In Figure 2 you can see the initial conditions and the evolved phase space DM sheet. Shell-crossing occurs at some time a_c . The left panel shows the distribution function at $a = a_c/10$, the central one at shell-crossing $a = a_c$ and the right panel at a much later time $a = 10a_c$.

If we project the sheet in configuration space to compute the density:

$$\rho(x) = m \int f(t, x, p) dp \quad (16)$$

we see that caustics form (see Figure 3). The moment when caustics form is called shell crossing because if you follow the trajectory of elements, at a certain moment they will cross the trajectory of another element and the projection of the distribution function in that point of configuration space will become infinite and then multivalued. See Figure 4.

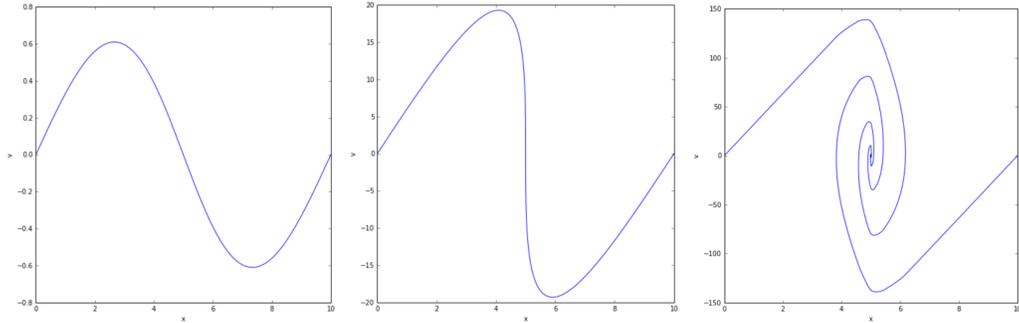


Figure 2: Evolution of a plane wave collapse using the CBE (credit: O. Hahn)

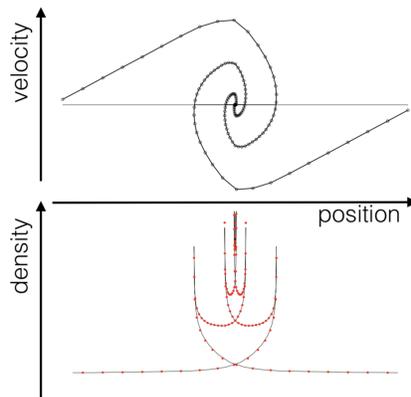


Figure 3: Top panel: phase space distribution function for a collapsed plane wave. Bottom panel: density projected in configuration space. The spikes in the density are the caustics. (Credit: O. Hahn)

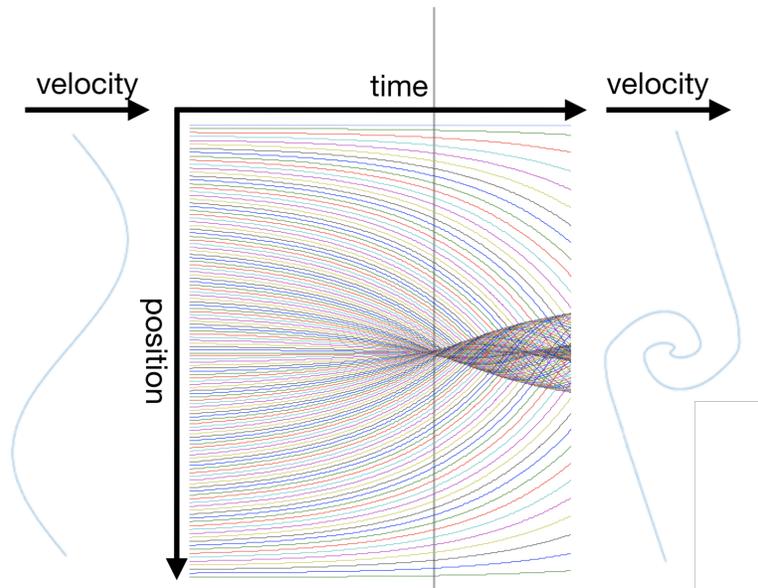


Figure 4: Left panel: initial phase space distribution. Middle panel: trajectories of elements in configuration space. Right panel: final phase space distribution. The grey vertical line indicates the time of shell crossing a_c . (Credit: O. Hahn)

2.3 The N-body technique

The solution of Eq.(14) can be represented by an infinite set of characteristics along which the phase space density is conserved. N-body techniques rely on a coarse-graining of the distribution function, which corresponds to sampling the characteristics with a finite number N of tracers. The distribution function of these tracers is [15, 9]:

$$f_N(t, \mathbf{x}, \mathbf{p}) = \sum_{i=1}^N \delta_D(\mathbf{x} - \mathbf{x}_i(t)) \delta_D(\mathbf{p} - \mathbf{p}_i(t)). \quad (17)$$

Substituting Eq.(17) in (14) we can see that the tracers follow the equations of motion:

$$\dot{\mathbf{x}}_p = \mathbf{v}_p, \quad (18)$$

$$\dot{\mathbf{v}}_p = -\nabla\phi, \quad (19)$$

where \mathbf{x}_p and \mathbf{v}_p are the position and velocity of the tracers. These tracers can be viewed as particles representing a discrete volume δV containing a mass m of microscopic DM particles. Assuming that the points are fairly sampling f , they can also be used to estimate a density, from which then the gravitational force can be calculated through the Poisson equation. A simple way of estimating the density will then be given by:

$$\rho(\mathbf{x}, t) = m \sum_{i=1}^N \delta_D(\mathbf{x} - \mathbf{x}_i(t)). \quad (20)$$

This density estimate will however not be smooth, and in fact, it can lead to arbitrarily high accelerations if two points come arbitrarily close. Such accelerations would however be a pure artefact of the sampling. In the limit $N \rightarrow \infty$ they should never appear. To this end, a better estimate is given by:

$$\rho(\mathbf{x}, t) = m \sum_{i=1}^N \delta_D(\mathbf{x} - \mathbf{x}_i(t)) * W_h(\mathbf{x}) \quad (21)$$

where W_h is a kernel that spreads out each of the points to a scale h so that nearby points overlap and the density field becomes smooth.

2.4 The Lagrangian phase space technique

We have seen before that the dark matter phase space distribution is confined in a 3D sub-manifold of the 6D phase space. One can parametrise this sub-manifold \mathcal{Q} in terms of lagrangian coordinates \mathbf{q} :

$$\mathcal{Q} \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}^3 \otimes \mathbb{R}^3 : \quad \mathbf{q} \mapsto (\mathbf{x}_{\mathbf{q}}, \mathbf{v}_{\mathbf{q}}), \quad (22)$$

whose evolution describes the whole evolution of the fluid. We can decompose the lagrangian manifold into finite volume elements that carry constant mass and define a local mapping between lagrangian and eulerian space, see Figure 5 for a 2D example. This map can be approximated by multi-variate

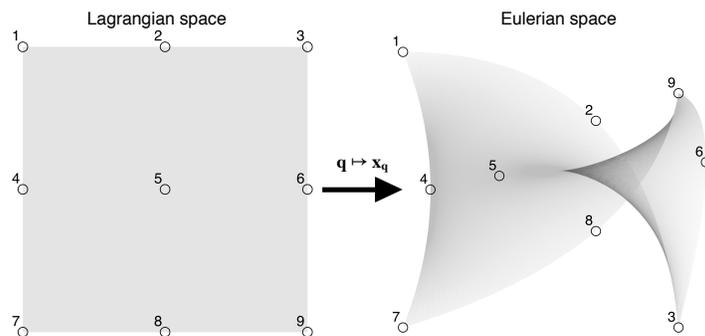


Figure 5: Mapping of one fluid element from lagrangian to eulerian coordinates. The numbered points represent the flow tracers. From [13].

polynomials. The coefficients of these polynomials, each defined on its Lagrangian cubical element, can be represented by a number of supporting points (or flow tracer particles), shown as numbered points in Figure 5. The number of these tracers per element is given by the order of the polynomials, so that a larger number of tracers corresponds to a more accurate representation of the mapping. It can be shown that the equations of motion for these tracers are [13]:

$$\dot{\mathbf{x}}_{\alpha\beta\gamma} = \mathbf{v}_{\alpha\beta\gamma} \quad (23)$$

$$\dot{\mathbf{v}}_{\alpha\beta\gamma} = -J^{-1}\mathbf{f}_{\alpha\beta\gamma} \quad (24)$$

where $J_{ij} = \partial x_i / \partial q_j$ is the Jacobian of the map and $\mathbf{f} = \nabla_q \phi$ is the force in lagrangian space.

Problem: computing the force due to these complexly shaped elements in Eulerian space is complicated, if at all possible. One solution is to replace the elements with massive particles located in their center of mass and then their mass is deposited onto a grid in Eulerian space [13]. In this case the method has two sets of tracers: mass tracers and flow tracers. Since the mass tracers are not interacting with each other there are no spurious two-body interactions like in the N-body case. This method works well for warm dark matter simulations, where N-body techniques fail due to artificial fragmentation [12, 23]. For the cold case one needs to introduce refinements to follow the rapid growth of the element’s volume in the deeply non-linear regime of gravitational collapse (i.e. at the center of halos).

3 Gravity solvers

Once one finds a representation of the density field, the Poisson equation can be solved to find the gravitational potential. The different N-body codes use different approaches for this step. I will describe the basics of each method and discuss pros and cons.

3.1 Particle-particle

The simplest approach to compute the gravitational force acting on particles is called direct summation or Particle-Particle (PP)[2]: the sum

$$\mathbf{f}(\mathbf{x}_j) = m_j \sum_{i \neq j} G m_i \frac{\mathbf{x}_i - \mathbf{x}_j}{|\mathbf{x}_i - \mathbf{x}_j|^3} \quad (25)$$

is performed for each particle j summing over all the other $N - 1$ particles i . This method has a computational cost of the order of N^2 , making it unaffordable for cosmological simulations, where the number of particles ranges from 10^6 to 10^{12} . Moreover, this method diverges when close encounters occur between two particles, so a “smoothing” of the force is required [1]. This is done by substituting the Newtonian force with a force \mathbf{f}_{ij}^{soft} that tends to \mathbf{f}_{ij}^{newt} when $|\mathbf{x}_i - \mathbf{x}_j| \rightarrow \infty$ and tends to 0 when $|\mathbf{x}_i - \mathbf{x}_j| \rightarrow 0$. A simple example of such a smoothed force is:

$$\mathbf{f}_{ij}^{soft} = G m_i m_j \frac{\mathbf{x}_i - \mathbf{x}_j}{(|\mathbf{x}_i - \mathbf{x}_j|^2 - \epsilon^2)^{3/2}}, \quad (26)$$

where ϵ is called softening length and can be interpreted as the size of the particles. Unphysical two-body interactions can arise if the softening length is smaller than the mean inter-particle separation [20], leading to departures from the collisionless nature of the Vlasov-Poisson system.

3.2 Hierarchical Tree

The so called Hierarchical Tree methods [5] alleviate the computational cost of the Poisson solver by using direct summation only for particles that are below a certain distance, while the force due to long range interactions is approximated by expanding the gravitational field in multipoles and truncating the expansion to a given order. The computational cost of these algorithms is of the order of $N \log N$. In these methods, particles are assigned to a tree of computational cells, which is constructed by iterative subdivision of the volume of each cell in 8 parts (octree), halting the procedure when all the cells contain either 0 or 1 particles. See Figure 6 for the representation of a 2D example.

Figure 7 sketches the configuration for the multipole expansion. To compute the force exerted by a collection of particles on a particle at a given distance one can approximate the gravitational potential of the group

$$\phi(\mathbf{r}) = -G \sum_i \frac{m_i}{|\mathbf{r} - \mathbf{x}_i|} = -G \sum_i \frac{m_i}{|\mathbf{r} - \mathbf{s} + \mathbf{s} - \mathbf{x}_i|} \quad (27)$$

by Taylor expanding the denominator for $\mathbf{x}_i - \mathbf{s} \ll \mathbf{r} - \mathbf{s}$. The expansion gives [11]:

$$\frac{1}{|\mathbf{y} + \mathbf{s} - \mathbf{x}_i|} = \frac{1}{|\mathbf{y}|} - \frac{\mathbf{y}(\mathbf{s} - \mathbf{x}_i)}{|\mathbf{y}|^3} + \frac{1}{2} \frac{\mathbf{y}^T [3(\mathbf{s} - \mathbf{x}_i)(\mathbf{s} - \mathbf{x}_i)^T - (\mathbf{s} - \mathbf{x}_i)^2] \mathbf{y}}{|\mathbf{y}|^5} + \dots \quad (28)$$

where $\mathbf{y} = \mathbf{r} - \mathbf{s}$. These terms define the monopole, dipole and quadrupole of the group of particles:

$$M = \sum_i m_i, \quad (29)$$

$$D_i = \sum_j m_j (\mathbf{s} - \mathbf{x}_j) = 0, \quad (30)$$

$$Q_{ij} = \sum_k m_k [3(\mathbf{s} - \mathbf{x}_i)(\mathbf{s} - \mathbf{x}_i)^T - (\mathbf{s} - \mathbf{x}_i)^2]. \quad (31)$$

The dipole vanishes because we carried out the expansion around the center of mass:

$$\mathbf{s} = \frac{1}{M} \sum_i m_i \mathbf{x}_i. \quad (32)$$

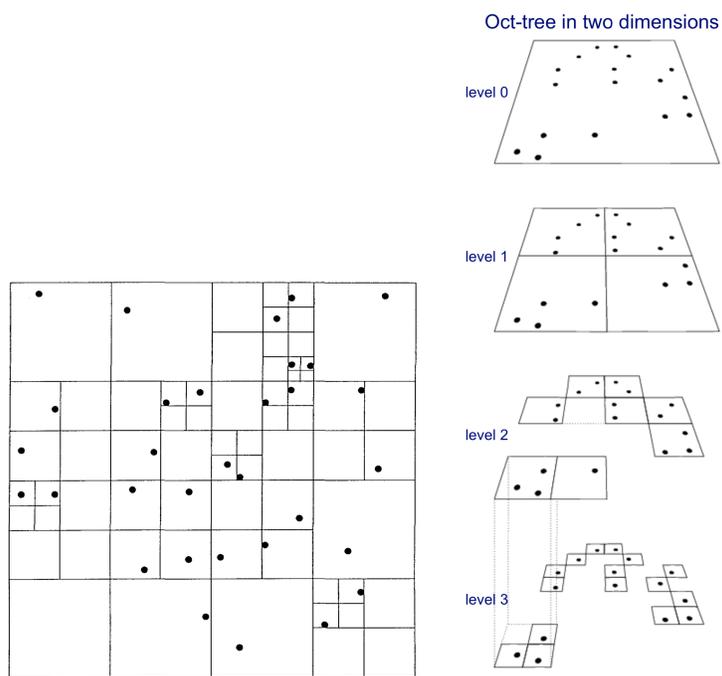


Figure 6: Examples of 2D hierarchical trees.

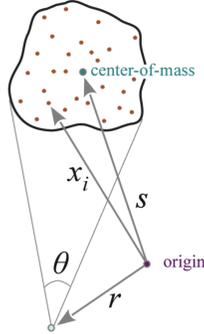


Figure 7: Multipole expansion for a group of distant particles. Provided the reference point \mathbf{r} is sufficiently far away, the particles are seen under a small opening angle θ , and the field created by the particle group can be approximated by the monopole term at its center of mass, augmented with higher order multipole corrections if desired. From [11]

If we stop the expansion at quadrupole order the potential is then given by:

$$\phi(\mathbf{r}) = -G \left(\frac{M}{|\mathbf{y}|} + \frac{1}{2} \frac{\mathbf{y}^T Q \mathbf{y}}{|\mathbf{y}|^5} \right), \quad (33)$$

from which the force can be obtained by differentiation. The opening angle θ is adjusted so that smaller distances correspond to larger opening angles. This procedure assures high resolution of the force in high density regions but is inefficient for low contrast mass distribution.

3.3 Particle-Mesh

Particle-mesh (PM) methods [14] take a different approach to solve the Poisson equation. In these methods a grid is superimposed to the particle distribution and physical quantities like density and velocity are interpolated from particle positions to grid nodes. The Particle Mesh algorithm assumes that particles have certain size, mass, shape, and internal density. This determines the interpolation scheme used to assign densities to grid cells. Let's define the 1D particle shape, $S(x)$, to be mass density at the distance x from the particle for cell size Δx . The common choices are:

Nearest Grid Point (NGP) : particles are point-like and all of particle's mass is assigned to the single grid cell that contains it. The shape function is:

$$S(x) = \frac{1}{\Delta x} \delta\left(\frac{x}{\Delta x}\right) \quad (34)$$

Cloud In Cell (CIC) : particles are cubes (in 3D) of uniform density and of one grid cell size. The shape function is:

$$S(x) = \frac{1}{\Delta x} \begin{cases} 1 & \text{if } |x| < 0.5\Delta x \\ 0 & \text{otherwise} \end{cases} \quad (35)$$

Triangular Shaped Cloud (TSC) : the shape function is:

$$S(x) = \frac{1}{\Delta x} \begin{cases} 1 - \frac{x}{\Delta x} & \text{if } |x| < \Delta x \\ 0 & \text{otherwise} \end{cases} \quad (36)$$

The fraction of the particle's mass assigned to a cell ijk is the shape function averaged over this cell:

$$W(x_p - x_{ijk}) = \int_{x_{ijk}-\Delta x/2}^{x_{ijk}+\Delta x/2} dx' S(x_p - x') \quad (37)$$

The density in a cell ijk is then:

$$\rho_{ijk} = \sum_{p=1}^{N_p} m_p W(\mathbf{r}_p - \mathbf{r}_{ijk}), \quad (38)$$

where:

$$W(\mathbf{r}_p - \mathbf{r}_{ijk}) = W(x_p - x_{ijk})W(y_p - y_{ijk})W(z_p - z_{ijk}). \quad (39)$$

Imposing periodic boundary conditions allows one to use Fast Fourier Transform algorithms [21] to compute the potential. In fact, in Fourier space the Poisson equation is:

$$\phi(\mathbf{k}) = G(\mathbf{k}) \rho(\mathbf{k}), \quad (40)$$

where $G(\mathbf{k})$ is the Green function, whose form depends on the discretisation scheme. For example, for the CIC assignment case, that is the most commonly used, the Green function is given by:

$$G(\mathbf{k}) = \frac{\Omega_m}{3a} \left[\sin^2 \frac{k_x}{2} + \sin^2 \frac{k_y}{2} + \sin^2 \frac{k_z}{2} \right]^{-1}. \quad (41)$$

PM algorithms first transform $\rho(\mathbf{x}) \rightarrow \rho(\mathbf{k})$ using forward FFT, multiply it by the Green function to obtain $\phi(\mathbf{k})$ and then transform $\phi(\mathbf{k}) \rightarrow \phi(\mathbf{x})$ using backwards FFT. The potential is then interpolated back to the particle positions. The advantage of this approach is that grid methods have a “natural” softening length given by the spacing of the grid points, which also sets the resolution scale of the simulation. This also means that the numerical error on the force with this method is more predictable than with PP or Tree methods. In the limit of vanishing spacing ($N_g \rightarrow \infty$) we recover the Vlasov-Poisson system. The computational cost of PM methods is of order N for particle displacements and $N_g \log N_g$ for the Poisson solver, where N_g is the number of grid nodes.

3.4 Hybrid approaches and refinement

Modern cosmological simulation codes use hybrid approaches that combine the methods described above for different regimes or introduce refinement of the grid to have a better resolution inside high density peaks. TreePM hybrid methods use the Tree algorithm for small-range interactions and exploit the speed of FFT to compute long-range gravitational forces with the PM algorithm. Adaptive Mesh Refinement (AMR) methods allow to reach higher accuracy in high density regions with respect to PM methods by introducing sub-grids with finer spacing where the density in the grid cells exceeds a given threshold. The potential at refined levels of the grid is computed with relaxation methods, using the solution interpolated from the previous level as initial guess and as boundary conditions, while at coarse level the potential is computed using standard FFT methods [17, 22]. For the lagrangian phase space method since the discretisation error increase with the size of the elements a natural refinement criterium is to split each element into smaller elements.

4 Cosmological simulations

Cosmological simulations solve the Vlasov-Poisson system of equations for matter perturbations in an expanding universe, where the evolution of the scale factor a is given by the background cosmological model. The main driver of the computational cost of simulations is the number of particles. For a fixed number of particles there is a compromise to be made between

the volume and the mass resolution, i.e. the smallest mass object that can be resolved in the simulation. This is represented in Figure 8 where some state-of-the-art simulations (in 2015) are represented as points in the volume versus mass resolution plane. The diagonal lines show the relation between volume and mass resolution for a fixed number of particles. Ideally for LSS

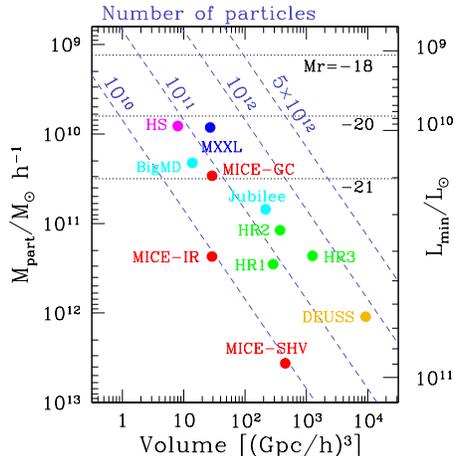


Figure 8: Volume and mass resolution for state-of-the-art cosmological simulations. Diagonal lines show the relation between volume and mass resolution for a fixed number of particles. The vertical right axis shows the faintest galaxy luminosity reached when populating the simulation using HOD (see Section 5.2). From [10].

studies one would want to cover a volume of order tens of Gpc^3 to beat down sample variance and at the same time resolve all the galactic scale halos. These requirements result in a number of particles of the order of tens of trillions, which is on the limit of what can be done today in the most powerful supercomputers. Even with the most efficient codes available today such simulations take millions of cpu-hours on hundreds of CPUs.

4.1 Initial Conditions

From CMB observations we know that at early times the fluctuation field δ is a homogeneous and isotropic Gaussian random field. For such a field, a

particular realisation can be written as:

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3k \sqrt{P(k)} \lambda_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (42)$$

where $\lambda_{\mathbf{k}}$ is a white noise field, i.e. a complex Gaussian random field that satisfies the independence condition:

$$\langle \lambda_{\mathbf{k}_1} \lambda_{\mathbf{k}_2}^* \rangle = (2\pi)^3 \delta_D(\mathbf{k}_2 - \mathbf{k}_1). \quad (43)$$

Since $\delta(\mathbf{x})$ is real the white noise is also subject to the Hermitian condition:

$$\lambda_{\mathbf{k}}^* = \lambda_{-\mathbf{k}}. \quad (44)$$

A realisation of the $\lambda_{\mathbf{k}}$ field can be obtained by drawing Gaussian random numbers with zero mean and unity variance at each grid-point. This field is then multiplied by the square-root of linear power spectrum for the desired cosmology, which can be computed using Boltzmann codes like CAMB [19] or CLASS [18]. Once the density field is constructed, the velocity field is computed using linear theory:

$$\mathbf{v} = -i a \frac{\dot{D}_+}{D_+} \frac{\mathbf{k}}{k^2} \delta. \quad (45)$$

FFT methods are used to Fourier transform the obtained density and velocity fields in configuration space. Once chosen a pre-initial configuration of particles, positions and velocities are evolved to the initial redshift of the simulation using Zel'dovich approximation. The displacement field is computed as:

$$\mathbf{s} = \frac{\mathbf{v}}{a \dot{D}_+}, \quad (46)$$

and particle positions and velocities are given by:

$$\mathbf{x}_p(a) = \mathbf{q} + D_+ \mathbf{s}(\mathbf{q}), \quad (47)$$

$$\mathbf{v}_p(a) = a \dot{D}_+ \mathbf{s}(\mathbf{q}), \quad (48)$$

where \mathbf{q} is the pre-initial position of the particle. In the simplest approach particles are placed at the position of grid-points, but in some cases their pre-initial configuration can be chosen not to coincide with the grid (e.g. glass configurations [25]). In such cases an additional interpolation of the velocity field to particle positions is required to compute the displacement.

Initial conditions for cosmological simulations are generated at high redshift, usually $z \sim 100$, where the Zel’dovich approximation is still valid on the scales of interest. One can also use higher order lagrangian perturbation theory to compute displacements, allowing to start the simulation at lower redshift, $z \leq 70$.

4.2 Identifying Halos

A halo is a gravitationally bound region in which matter has decoupled from the Hubble flow and collapsed. Finding the boundary of a halo is not an easy task since their density profile decays smoothly with radius. Even more complicated is finding a definition that can be measured observationally. There have been multiple definitions used in the field:

Friends of Friends Percolation algorithm. Groups that contain all particles separated by distance less than a given linking length, $b\bar{l}$, where \bar{l} is the mean inter-particle separation in simulations and b is a free parameter of the algorithm. The velocity of the halo is computed as the average of all halo particle velocities. Sub-halos can be found by decreasing the linking length. Pros: simple (only one parameter), doesn’t assume anything about the halo, used in observations. Cons: no analytical predictions, sensitive to resolution, velocities are not very accurate.

Spherical Over-density the radius of the sphere enclosing n times the average/critical density.

$$\bar{\rho}_h = \frac{3M_\Delta}{4\pi r_\Delta^3} = \Delta\rho_{ref} \quad (49)$$

where Δ is the threshold and ρ_{ref} is the reference density. Usual choices for Δ are a constant value ~ 200 or a time-varying Δ_{vir} motivated by the spherical collapse model. ρ_{ref} is usually set to either the mean matter density or the critical density at the redshift of the halo. Pros: can be predicted with analytical models of halo collapse, can be used in observations. Cons: assumes spherical symmetry, leads to pseudo-evolution of mass.

Splashback Radius the radius at which the density profile has a sharp drop (see Figure 9). This seems to coincide with the region where particles accumulate on the first apocenter passage after their infall

onto a halo. These particles pile up due to their low radial velocity, creating a caustic that manifests itself as a sharp drop in the density profile. Pros: more physical. Cons: hard to measure in real data.

Phase space finders uses full 6D phase space information of particles to identify halos. An example of such algorithm is summarised in Figure 10. Pros: does a better job during halo mergers and identifying sub-halos, halo properties are more accurate. Cons: hard to predict analytically.

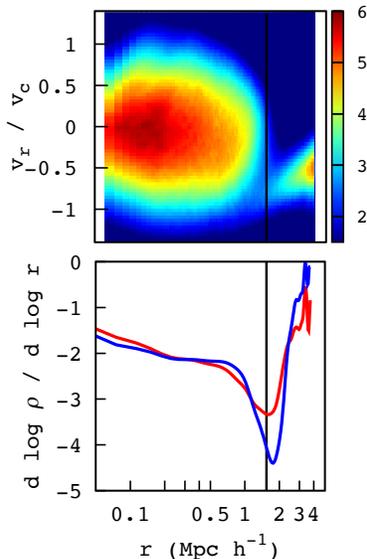


Figure 9: Top: the phase space diagram for halos from the MultiDark simulation in the mass range $M = 1 - 4 \times 10^{14} h^{-1} M_{\odot}$. The colorbar shows the number of particles within each phase space pixel. Bottom: the local slope of the density of all particles (red) and particles with $|v_r| < 0.4 v_{circ}$ (blue), as a function of radius r . The location of the feature in the local slope coincides with the outer caustic at the splashback radius. From [4].

Since a lot of the modelling uses halo masses (cluster mass function, halo occupation distribution) it is important to use a consistent mass definition.

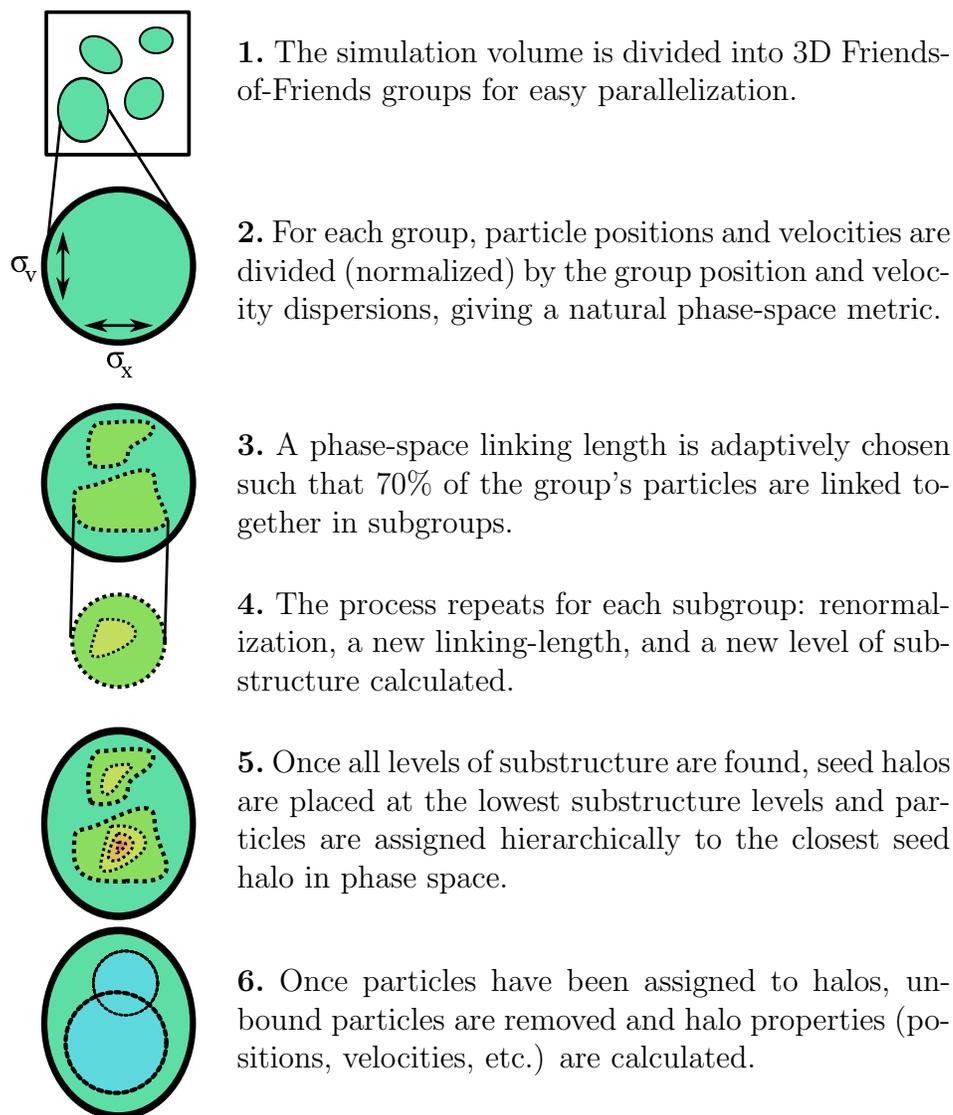


Figure 10: A visual summary of the Rockstar phase space halo finding algorithm [7].

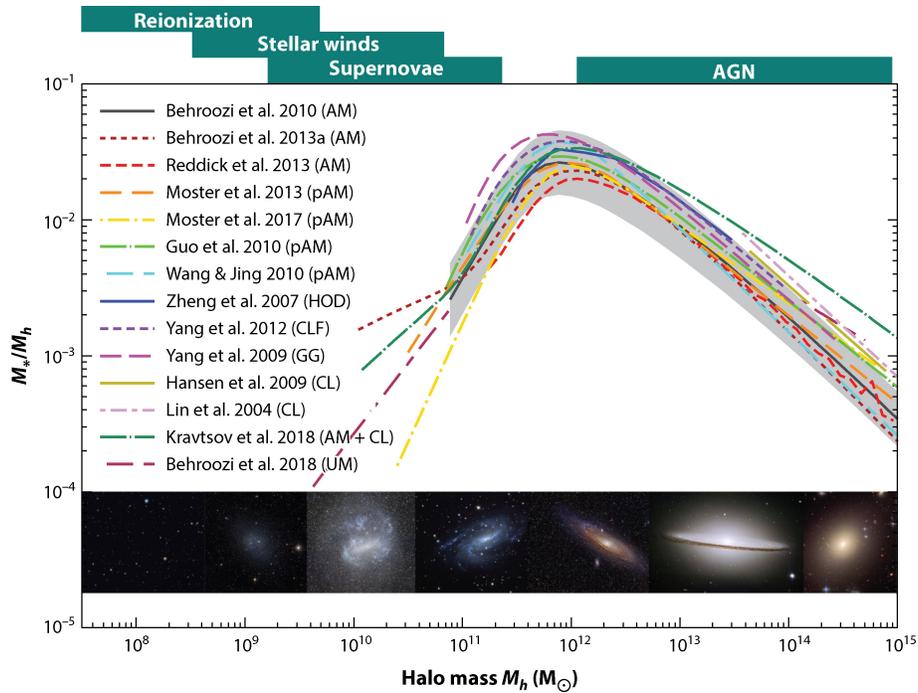
5 Statistical modelling of galaxy distribution

In order to compare results from simulations to observations one needs to relate the statistics of the matter distribution to the visible matter component. Most of the LSS surveys observe galaxies, so we need a prescription to relate the dark matter and the galaxy distribution.

Galaxy formation is a messy process that depends on the detail of the astrophysical processes at galactic and sub-galactic scales. One can use hydrodynamical simulations that include the gas component, but even in that case there is a lot of physics that remains unresolved and needs to be treated effectively. Fortunately, for the purpose of measuring statistical properties of the galaxy density field we can rely on a simplified picture of galaxy formation. To summarise this picture in a few sentences: at early times perturbations in the dark matter component start collapsing and forming halos. The gas falls into the potential well of dark matter halos and if the halo is large enough the gas can cool and form stars and galaxies. We know structure formation is bottom-up, so first the small halos are formed and then they merge to create larger structures. Galaxies inside the halos continue to form stars and accrete new material during mergers.

At first order we can say that the properties of galaxies only depend on halo mass: the larger the halo, the larger the galaxy inside the halo. This means that we can build a simple relation between halo mass and e.g. stellar mass of the galaxy that will tell us about the efficiency of star formation in different halos. Figure 11 summarises the results from different methods that probe the stellar-to-halo mass relation. All available methods now paint a consistent picture of this relation: it peaks at a few per-cent efficiency around L_* galaxies. At high mass AGN feedback heat the gas preventing it to form new stars. At low mass winds from massive stars eject the gas from the galaxy. Below a certain threshold there is not enough gas to form a galaxy. We currently think this happens below $M_h = 10^9 M_\odot$.

We will distinguish between central galaxies, that reside in the center of the halo, and satellite galaxies, that reside in sub-halos that orbit in the main halo. Resolving substructures in simulations can be hard due to the limited mass resolution: on one hand the smallest structures cannot be identified because they are represented only by a handful of particles and on the other



 Wechsler RH, Tinker JL. 2018. *Annu. Rev. Astron. Astrophys.* 56:435–87

Figure 11: Stellar-to-halo mass ratio of central galaxies at $z = 0$. Bottom panel shows example galaxies that are hosted by halos in the specified mass range. Across the top of the figure, we indicate key physical processes that may be responsible for ejecting or heating gas or suppressing star formation at those mass scales. From [24].

hand even if they can be identified their mass is stripped by tidal interactions with the host halo, so that they can fall below the threshold of detectability and disappear during the simulation. For this reason a number of proxies for sub-halo mass is used, like the mass at the time of accretion into the main halo or the maximum circular velocity.

5.1 Abundance Matching

Simplest assumption: the most massive galaxies live in the most massive dark matter halos. We can generate a realisation of a measured stellar mass function and associate each galaxy to a halo in a rank-ordered way. This is called Abundance Matching. One can also use sub-halos (SHAM). Usually stellar mass is used for matching but one can also use luminosity. Since the stellar-to-halo mass relation has some scatter, one can introduce a scatter in the matching. The amount of scatter can be constrained using clustering measurements.

5.2 Halo Occupation Distribution

In the halo occupation distribution (HOD) framework the galaxy-halo connection is modelled using a probability distribution for the number of galaxies (with some selection) in a halo, conditioned on some halo property, usually their mass, $P(N|M)$. This is usually separated into central and satellite galaxies, each with its own distribution. Usually centrals are distributed with a Bernoulli distribution and satellites with a Poisson distribution. We only need to know the average occupation number to specify the distributions. Using hi-res simulations it was determined that the functional shape of the sub-halo abundance is a power-law M^α where $\alpha \sim 1$ [16].

$$\langle N \rangle = \begin{cases} 0 & \text{if } M < M_{min} \\ (M/M_1)^\alpha & \text{otherwise,} \end{cases} \quad (50)$$

This was shown to hold for luminosity or stellar mass threshold selected galaxies. For more complicated selections the functional shape can be more complex.

Acknowledgments

In preparing parts of these notes I relied on private copies of various lecture notes and presentations by Nick Gnedin, Oliver Hahn and Volker Springel.

References

- [1] S. J. Aarseth. Dynamical evolution of clusters of galaxies, I. *Monthly Notices of the Royal Astronomical Society*, 126:223, 1963.
- [2] S. J. Aarseth. *Gravitational N-Body Simulations*. Cambridge University Press, 2003.
- [3] J. Adamek, D. Daverio, R. Durrer, and M. Kunz. gevolution: a cosmological N-body code based on General Relativity. *Journal of Cosmology and Astroparticle Physics*, 2016(07):053–053, July 2016. arXiv: 1604.06065.
- [4] S. Adhikari, N. Dalal, and R. T. Chamberlain. Splashback in accreting dark matter halos. *Journal of Cosmology and Astroparticle Physics*, 2014(11):019–019, Nov. 2014.
- [5] J. Barnes and P. Hut. A hierarchical $O(N \log N)$ force-calculation algorithm. *Nature*, 324:466, 1986.
- [6] C. Barrera-Hinojosa and B. Li. GRAMSES: a new route to general relativistic N-body simulations in cosmology - I. Methodology and code description. *Journal of Cosmology and Astroparticle Physics*, 2020(01):007–007, Jan. 2020. arXiv: 1905.08890.
- [7] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu. The Rockstar Phase-Space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *The Astrophysical Journal*, 762(2):109, Jan. 2013.
- [8] J. Binney and S. Tremaine. *Galactic Dynamics*. Princeton University Press, second edition, 2008.
- [9] T. H. Dupree. Nonlinear theory of drift-wave turbulence and enhanced diffusion. *Physics of Fluids*, 10:1049, 1967.

- [10] P. Fosalba, M. Crocce, E. Gaztañaga, and F. J. Castander. The MICE grand challenge lightcone simulation - I. Dark matter clustering. *Monthly Notices of the Royal Astronomical Society*, 448(4):2987–3000, Apr. 2015.
- [11] N. Y. Gnedin, S. C. O. Glover, R. S. Klessen, and V. Springel. *Star Formation in Galaxy Evolution: Connecting Numerical Models to Reality: Saas-Fee Advanced Course 43*. Swiss Society for Astrophysics and Astronomy. Springer, Sept. 2015.
- [12] M. Götz and J. Sommer-Larsen. Galaxy formation: Warm dark matter, missing satellites, and the angular momentum problem. *Astrophysics and Space Science*, 284:341–344, Apr. 2003.
- [13] O. Hahn and R. E. Angulo. An adaptively refined phase-space element method for cosmological simulations and collisionless dynamics. *Monthly Notices of the Royal Astronomical Society*, 455(1):1115–1133, Jan. 2016.
- [14] R. W. Hockney and J. W. Eastwood. *Computer Simulations Using Particles*. McGraw - Hill, 1981.
- [15] Y. L. Klimontovich. *The Statistical Theory of Non-Equilibrium Processes in a Plasma*. MIT Press, 1967.
- [16] A. V. Kravtsov, A. A. Berlind, R. H. Wechsler, A. A. Klypin, S. Gottloeber, B. Allgood, and J. R. Primack. The Dark Side of the Halo Occupation Distribution. *The Astrophysical Journal*, 609(1):35–49, July 2004.
- [17] A. V. Kravtsov, A. Klypin, and A. M. Khokhlov. Adaptive refinement tree: A new high-resolution N-body code for cosmological simulations. *Astrophysical Journal Supplement Series*, 111:73, 1997.
- [18] J. Lesgourgues. The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *arXiv:1104.2932*, 2011.
- [19] A. Lewis, A. Challinor, and A. Lasenby. Efficient computation of Cosmic Microwave Background anisotropies in closed Friedmann-Robertson-Walker models. *Astrophysical Journal*, 538:473, 2000.
- [20] A. Melott, S. F. Shandarin, J. S. Randall, and Y. Suto. Demonstrating discreteness and collision error in cosmological N-body simulations

- of Dark Matter gravitational clustering. *Astrophysical Journal Letters*, 479:L79, 1997.
- [21] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, third edition, 2007.
- [22] R. Teyssier. Cosmological hydrodynamics with adaptive mesh refinement. *Astronomy and Astrophysics*, 385:337, 2002.
- [23] J. Wang and S. D. M. White. Discreteness effects in simulations of hot/warm dark matter. *Monthly Notices of the Royal Astronomical Society*, 380:93–103, Sept. 2007.
- [24] R. H. Wechsler and J. L. Tinker. The Connection Between Galaxies and Their Dark Matter Halos. *Annu. Rev. Astron. Astrophys.*, 56:435–487, 2018.
- [25] S. D. M. White. Formation and evolution of galaxies. In R. Schaeffer, J. Silk, M. Spiro, and J. Zinn-Justin, editors, *Cosmology and large-scale structure*, page 349. Elsevier, 1996.
- [26] Y. B. Zeldovich. An approximate theory for large density perturbations. *Astronomy and Astrophysics*, 5:84, 1970.