Probing the properties of dark matter particles with astrophysical observations

Magalich, A.

Chapter 4

Structure formation

In this Section we briefly describe the observed *Large-Scale Structure of the Universe* (LSS) and its origin. We define the Large-Scale Structure as those inhomogeneities in the matter distribution of the Universe that are roughly larger than the scale of a galaxy. It appears that galaxies form *groups of galaxies*, individual groups form *clusters of galaxies*, while clusters of galaxies can be a part of *superclusters*. These objects are randomly distributed in the Universe and are connected by quasi-1D structures called *filaments* or quasi-2D structures called *walls*. In addition to these overdense regions there are areas with low density of matter called *voids*. All these features together are called *the Cosmic Web*. We will discuss them in Section 4.1.

It appears that these complicated and non-trivial structures can be described using basic equations of matter dynamics (see Section 4.3) starting from very simple initial conditions (see Section 4.3.4). At early times the Universe is almost homogeneous and evolution of the seeds of the current structures can be describe by a simplified linear equations, see Section 4.3. At later stages the evolution of the overdensities cannot be described analytically because of the non-linear nature of basic equations. The widely used way to overcome this problem is to use numerical simulations, so-called N-body simulations, see Section 4.4.1. In Section 4.6 we describe in some details numerical simulations that will be used in our own work, described in Chapter 5. We conclude that the approach reviewed here and applied for the ΛCDM model can describe the observed LSS with a high precision, see Section 4.7.

4.1 Observations of the Large-Scale Structure

Almost one century ago it has been noted by Edwin Hubble that the distribution of the *nebulae* outside our galaxy is fairly uniform on large scales, but starting from the angular scales $\lesssim 10^\circ$ it becomestak clumpy [112–114]. Let us discuss the main blocks of the LSS. The smallest objects of the LSS that we consider here are individual galaxies. They have different sizes from a few to tens of kpc with the total mass from $10^9 M_\odot$ to $10^{13} - 10^{14} M_\odot$

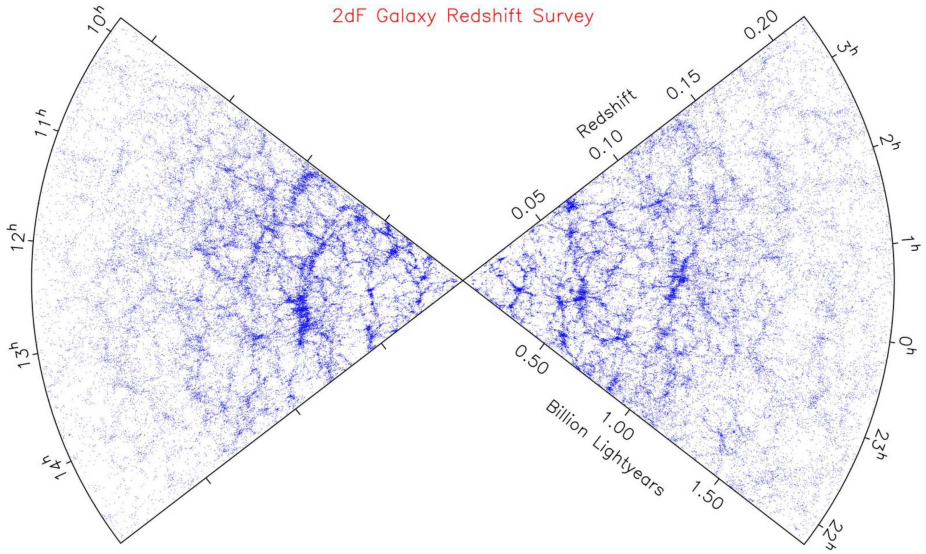

Figure 4.1: The distribution of galaxies found by the 2dF Galaxy Redshift Survey [115].

for the large elliptical galaxies. We expect that there are $\mathcal{O}(10^{11})$ galaxies in the observable Universe [116].

Group of galaxies is a collection of galaxies that consists of $\mathcal{O}(50)$ gravitationally bound members; the collections of galaxies larger than groups in which galaxies are not clustered in some smaller aggregations are called galaxy clusters [117]. The Milky Way galaxy is part of a group of galaxies called the Local Group. A group of galaxies have the typical size of a few Mpc and the typical masses from $10^{13}M_{\odot}$ to $10^{14}M_{\odot}$, while galaxy clusters are from $10^{14}M_{\odot}$ to $10^{15}M_{\odot}$. Typically, galaxy clusters contain $\mathcal{O}(10^3)$ individual galaxies.

The largest known structures in the Universe are superclusters. They are a large group of smaller galaxy clusters or galaxy groups. The large size and low density of superclusters means that they, unlike clusters, expand with the Hubble expansion. The number of superclusters in the observable universe is estimated to be 10 million. The map of the nearby superclusters is shown in Fig. 4.2.

It is known from observations, that small structures such as galaxies, start to form before large structures (see [118] and references therein). The first galaxies are observed at redshifts $z \sim 10$, while the first galaxy clusters start to grow relatively recently, $z \sim 2 - 3$. This is called the bottom-up structure formation and such a scenario exclude, for example, hot dark matter in which large structures (clusters) are formed before galaxies [33].

Because of the stochastic nature of the LSS the main observables that describe it are different distribution and correlation functions: halo mass function, galaxy-galaxy correlation function, matter power spectrum, etc. The interesting feature that can be found in the galaxy-galaxy correlation function is baryon acoustic oscillations (BAO). The same oscillations that are present in CMB are also imprinted in the density of the baryons and can be observed in the distribution of galaxies at cosmological scales, see Fig. 4.3. The possible

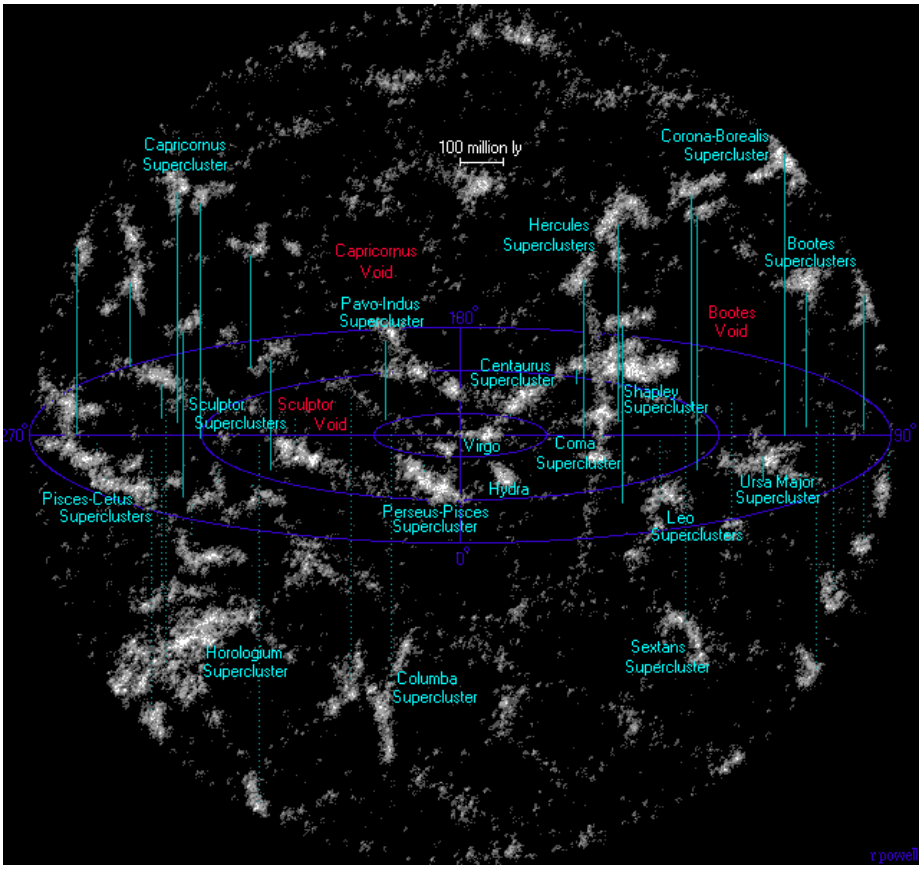

Figure 4.2: A map of superclusters around the Sun.
The figure is from atlasoftheuniverse.com.

application of BAO to cosmology is an independent measurement of the Hubble constant, that can be done using calibration of distances to galaxies using the SN Ia data [120, 121].

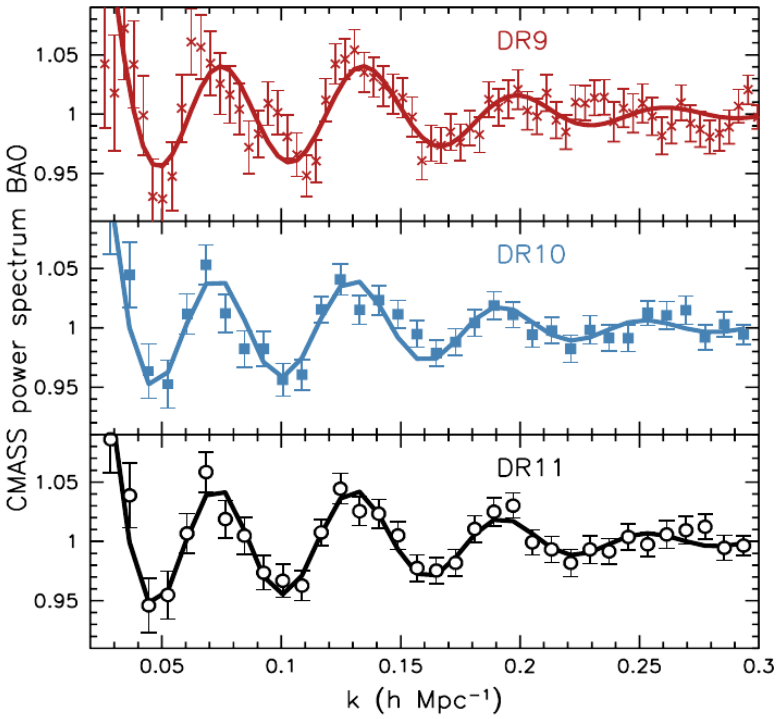

Figure 4.3: The baryon acoustic oscillations in the power spectrum of each of the BOSS data releases, DR9, DR10, and DR11 [119].

4.2 Description of the matter distribution

Matter distribution is defined by its mass density $\rho(\vec{r}, t)$. The inhomogeneities result in appearance of the overdensity $\delta(\vec{x}, t)$ – a relative difference between the local density and the average density of the Universe $\bar{\rho}(t)$:

$$\rho(\vec{x}, t) = \bar{\rho}(t)[1 + \delta(\vec{x}, t)] \quad \Leftrightarrow \quad \delta(\vec{x}, t) = \frac{\rho(\vec{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}. \tag{4.2.1}$$

Structure formation was seeded by a quantum process that led to a random distribution of tiny over- and under-densities at very early times. In this case, the particular values of matter density are not instructive – i.e., observing a galaxy at a particular point by itself does not provide useful information. Only the collective properties of the structures, their positions relative to each other, sizes and numbers are important. Hence, we treat $\delta(\vec{x}, t)$ as a random field, the statistical properties of which we want to study. A random field can be completely described by a series of moments – correlations between 1, 2, 3 to infinitely many points defined as:

$$\langle \delta \rangle = \int \delta(\vec{x}) d^3\vec{x} \tag{4.2.2}$$

$$\langle \delta(\vec{x}_1)\, \delta(\vec{x}_2) \rangle = \int \delta(\vec{x}_1)\, \delta(\vec{x}_2)\, d^3\vec{x}_1 d^3\vec{x}_2 \equiv \xi(\vec{x_1}, \vec{x_2}) \tag{4.2.3}$$

etc. Moments of the distribution characterize the generating process leaving out the information about the particular realization of the field. For simple distributions like Gaussian, the first two moments contain all the information while the others are identically zero. Moreover, in general case, the higher-order moments will decrease in amplitude.

When applied to the observed Universe, where the average overdensity is zero by definition $\langle\delta\rangle = 0$ while primordial[1] three-point and higher-order correlations are measured to be consistent with zero. Hence unless we expect some specific non-gaussianity, the two-point correlation function $\xi(\vec{x_1}, \vec{x_2})$ is the main source of information. According to the Cosmological Principle, the Universe we live in is homogeneous and isotropic at large scales. This also means that $\xi(\vec{x_1}, \vec{x_2})$ is invariant under rotations and translations:

$$\xi(\vec{x_1}, \vec{x_2}) \approx \xi(|\vec{x_1} - \vec{x_2}|) \tag{4.2.4}$$

Power spectrum. At small overdensities the gravitational equations can be linearized. This in turn allows for decoupling of scales and simple solutions for these equations in Fourier space. Hence the Fourier image of the correlation function is a particularly useful quantity:

$$\xi\left(|\vec{x}_1 - \vec{x}_2|\right) = \int \frac{d^3\vec{k}_1 d^3\vec{k}_2}{(2\pi)^6} \left\langle \hat{\delta}\left(\vec{k}_1\right) \hat{\delta}\left(-\vec{k}_2\right) \right\rangle e^{-\left(i\vec{k}_1\cdot\vec{x}_1 - i\vec{k}_2\cdot\vec{x}_2\right)} \tag{4.2.5}$$

where δ being a real field, means $\delta(-\vec{k}) = \delta^*(\vec{k})$

$$\left\langle \delta\left(\vec{k}_1\right) \delta^*\left(\vec{k}_2\right) \right\rangle = (2\pi)^3 P\left(\vec{k}_1\right) \delta_{\mathrm{D}}\left(\vec{k}_1 - \vec{k}_2\right) \tag{4.2.6}$$

$$\xi\left(|\vec{x}_1 - \vec{x}_2|\right) = \xi(r) = \int \frac{d^3\vec{k}}{(2\pi)^3} P(k) e^{-i\vec{k}\cdot\vec{r}} \tag{4.2.7}$$

4.3 Basic equations governing Large Scale Structure and methods of solution

4.3.1 Equations for a self-gravitating fluid

A simplified description of the density perturbations evolution can be obtained from Newtonian theory if we treat a discrete medium of dark matter particles as a perfect fluid over the expanding background satisfying the Friedmann equations. This approximation is valid as long as we are interested in scales that are both sub-horizon and much larger than the mean free path of the particles.

A perfect fluid is characterized by 3 functions: density distribution $\rho(\vec{r}, t)$, entropy per unit mass $S(\vec{r}, t)$ and velocity field $\vec{v}(\vec{r}, t)$. The density distribution satisfies the continuity

[1] Non-gaussianity can be generated during structure formation

equation:

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}(\rho \vec{v}) = 0. \tag{4.3.1}$$

The forces acting on a small matter element with a mass M are gravity and pressure p:

$$\vec{F}_{gr} = -M\vec{\nabla}\phi, \tag{4.3.2}$$

$$\vec{F}_{pr} = -\oint p d\vec{\sigma} = -\int_V \vec{\nabla} p dV \approx -\vec{\nabla} pV, \tag{4.3.3}$$

where ϕ is the gravitational potential and $\vec{\sigma}$ is the surface element.

Acceleration of this mass element on the trajectory is

$$a \equiv \frac{d\vec{v}(\vec{r}(t),t)}{dt} = \left(\frac{\partial \vec{v}}{\partial t}\right)_r + \frac{dr^i(t)}{dt}\left(\frac{\partial \vec{v}}{\partial r^i}\right) = \frac{\partial \vec{v}}{\partial t} + (\vec{v}\cdot\vec{\nabla})\vec{v}. \tag{4.3.4}$$

This leads to the equation

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v}\cdot\vec{\nabla})\vec{v} + \frac{\vec{\nabla} p}{\rho} + \vec{\nabla}\phi = 0. \tag{4.3.5}$$

The self-consistent gravitational potential is given by the Poisson equation

$$\Delta\phi = 4\pi G\rho. \tag{4.3.6}$$

If the dissipation is negligible, the entropy is also conserved:

$$\frac{dS(\vec{r}(t),t)}{dt} = \frac{\partial S}{\partial t} + (\vec{v}\cdot\vec{\nabla})S = 0. \tag{4.3.7}$$

The final equation that closes the system is the equation of state that defines the pressure:

$$p = p(\rho, S) \tag{4.3.8}$$

4.3.2 Linearized theory

The system of equations above is non-linear, but it can be considerably simplified in the case when the perturbations are small: $\rho = \rho_0 + \rho_1, \rho_1/\rho_0 \ll 1$. Expanding density, entropy, velocity, pressure and gravitational potential around the averages, we obtain:

$$\frac{\partial \delta\rho}{\partial t} + \rho_0 \vec{\nabla}\vec{\delta v} = 0 \tag{4.3.9}$$

$$\frac{\partial \vec{\delta v}}{\partial t} + \frac{\vec{\nabla}\delta p}{\rho_0} + \vec{\nabla}\delta\phi = 0 \tag{4.3.10}$$

$$\frac{\partial \delta S}{\partial t} = 0 \tag{4.3.11}$$

$$\Delta \delta\phi = 4\pi G \delta\rho \tag{4.3.12}$$

The first-order perturbation of pressure is constrained by the equation of state to be

$$p = p(\rho_0 + \delta\rho, S_0 + \delta S) = p_0 + \delta p \tag{4.3.13}$$

$$\delta p = \left(\frac{\partial p}{\partial \rho}\right)_S \delta\rho + \left(\frac{\partial p}{\partial S}\right)_\rho \delta S \equiv c_s^2 \delta\rho + \sigma \delta S \tag{4.3.14}$$

where c_s is the speed of sound.

The equation for the entropy perturbation state that the entropy is an arbitrary time-independent function. Since the entropy perturbations are not observed, we can put $\delta S = 0$ without loss of generality.

Then the system becomes

$$\frac{\partial \delta\rho}{\partial t} + \rho_0 \vec{\nabla} \vec{\delta v} = 0 \tag{4.3.15}$$

$$\frac{\partial \vec{\delta v}}{\partial t} + \frac{c_s^2}{\rho_0} \vec{\nabla} \delta\rho + \vec{\nabla} \delta\phi = 0 \tag{4.3.16}$$

$$\Delta \delta\phi = 4\pi G \delta\rho \tag{4.3.17}$$

By taking the divergence of the velocity equation and using other equations to express $\vec{\nabla}\delta\vec{v}$ and $\Delta\delta\phi$ in terms of $\delta\rho$, we obtain:

$$\frac{\partial^2 \delta\rho}{\partial t^2} - c_s^2 \Delta \delta\rho - 4\pi G \rho_0 \delta\rho = 0 \tag{4.3.18}$$

Since the coefficients of equations do not depend on the coordinates, a transition to the Fourier space can considerably simplify it:

$$\delta\rho(\vec{r}, t) = \int \delta\rho_k(t) e^{\imath \vec{k} \cdot \vec{r}} \frac{d^3 k}{(2\pi)^{3/2}} \tag{4.3.19}$$

$$\delta\ddot{\rho}_k + (k^2 c_s^2 - 4\pi G \rho_0) \delta\rho_k = 0 \tag{4.3.20}$$

This equation has two independent solutions

$$\delta\rho_k \propto e^{\pm \imath \sqrt{k^2 c_s^2 - 4\pi G \rho_0} t} \equiv e^{\pm \imath \omega(k) t} \tag{4.3.21}$$

The sign under square root governs the behaviour of $\delta\rho_k$ modes. We define the so-called

Jeans scale k_J for which $\omega(k_J) = 0$:

$$k_J = \sqrt{\frac{4\pi G\rho_0}{c_s^2}} \tag{4.3.22}$$

For $k > k_J$ the density perturbation behaves as a sound wave driven by pressure and for $k < k_J$ the perturbation exponentially grows or decays, reflecting the *gravitational instability*:

$$\delta\rho_k \propto e^{\pm|\omega|t} \tag{4.3.23}$$

When considering large scales $k \to 0$, $|\omega|t \to t/t_{gr} \equiv \sqrt{4\pi G\rho_0}t$. We interpret t_{gr} as the gravitational timescale characteristic to the collapse of the region with density ρ_0.

4.3.3 Perturbations in the expanding Universe

The background density and average velocity fields are subject to Friedmann equations:

$$\dot{\rho}_0 = -3H(t)\rho_0 \tag{4.3.24}$$

$$\vec{v}_0 = H(t)\vec{r} \tag{4.3.25}$$

$$\dot{H} + H^2 = -\frac{4\pi G}{3}\rho_0 \tag{4.3.26}$$

Note that the Hubble flow velocity explicitly depends on the coordinates, which will complicate the form of the equations in the Fourier space. To avoid this, we can switch to the so-called *comoving frame* where coordinates are multiplied by the scale factor $\vec{x} = a(t)\vec{r}$. In this frame the derivatives operators become

$$\vec{\nabla} \to \frac{1}{a}\vec{\nabla} \tag{4.3.27}$$

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} - \left(\vec{v}_0 \cdot \frac{1}{a}\vec{\nabla}\right) \tag{4.3.28}$$

The system in the comoving frame takes form

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\vec{\nabla}\delta\vec{v} = 0, \tag{4.3.29}$$

$$\frac{\partial\delta\vec{v}}{\partial t} + H\delta\vec{v} + \frac{c_s^2}{a}\vec{\nabla}\delta + \frac{1}{a}\vec{\nabla}\delta\phi = 0, \tag{4.3.30}$$

$$\Delta\delta\phi = 4\pi G a^2\rho_0\delta, \tag{4.3.31}$$

where $\delta \equiv \left(\frac{\delta\rho}{\rho}\right)$ is called *density contrast*.

Following the logic of the static background, we find the equation for the density

perturbation:

$$\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2}\Delta\delta - 4\pi G\rho_0\delta = 0. \tag{4.3.32}$$

Finally, after Fourier transform with the respect to the comoving coordinate $\vec{x}$ the equation becomes

$$\ddot{\delta}_k + 2H\dot{\delta}_k + \left(\frac{c_s^2 k^2}{a^2} - 4\pi G\rho_0\right)\delta_k = 0 \tag{4.3.33}$$

On the scales much larger than the Jeans scale, we can neglect the k-term and find the most general solution

$$\delta = C_1 H \int \frac{dt}{a^2 H^2} + C_2 H \tag{4.3.34}$$

For a flat, matter-dominated universe $a \propto t^{2/3}$ and $H \propto t^{-1}$:

$$\delta = C_1 t^{2/3} + C_2 t^{-1} \tag{4.3.35}$$

Note that in expanding Universe, perturbations grow much slower than in a static Universe – only as a power-law of time instead of exponential.

Multicomponent case. When the Universe is filled with multiple kinds of matter that interact mainly through gravity (e.g. dark matter and baryons), the dynamical equations for each component remain unchanged, while the Hubble rate and the gravitational potential are given by the total energy density.

4.3.4 Initial conditions for inhomogeneities

We have discussed above how the inhomogeneities evolve in the expanding Universe. What is the simplest possible power spectrum?

In the case of $P(k) = 0$ the density field is completely homogeneous. Next to that, what power spectrum would contain the least information?

Let's assume a power-law shape of $P(k) = Ak^n$. First of all, observationally $n \sim 1$. And indeed, for $n = 1$ (known as the Harrison-Zeldovich spectrum) the variation of the gravitational potential is scale-independent. Indeed, if we write the Poisson equation in proper coordinates for a perturbation δ giving rise to a perturbation of the potential $\delta\Phi$, we have

$$\nabla^2\delta\Phi = 4\pi G\bar{\rho}\delta \tag{4.3.36}$$

in Fourier space

$$\delta\hat{\Phi}_k = -4\pi G\bar{\rho}\frac{\hat{\delta}_k}{k^2} \tag{4.3.37}$$

meaning that

$$\Delta^2_{\delta\Phi_k} \propto k^3 \left\langle \left| \delta\hat{\Phi}_k \right|^2 \right\rangle \propto k^3 \left\langle |\hat{\delta}(k)|^2 \right\rangle \frac{1}{k^4} \propto k^{n-1} \tag{4.3.38}$$

4.4 From linear theory to N-body simulations

At some point, the growing perturbations achieve level $\delta \sim 1$ breaking the defining assumption of the linear theory. These perturbations have to be described by non-linear gravitational equations for which there are no analytical solutions.

However, it is possible to approximate the solutions by replacing individual DM particles with large super-particles. In this case, we can follow the evolution of each particle using computer simulations. This approach is called N-body simulations.

4.4.1 N-body simulations

Dark matter can be thought of as a collision-less self-gravitating fluid. If we associate a distribution function to it, the equations of motion can be expressed in a form of *collisionless Boltzmann equation*:

$$\frac{\partial f}{\partial t} + \dot{\vec{x}}\frac{\partial f}{\partial \vec{x}} - \vec{\nabla}\phi\frac{\partial f}{\partial \dot{\vec{x}}} = 0 \tag{4.4.1}$$

where $f(\vec{x}, \dot{\vec{x}}, t)$ is defined as number of particles at phase-space point $(\vec{x}, \dot{\vec{x}})$

$$\rho(\vec{x}, t) = \frac{m}{a^3(t)} \int d^3\dot{\vec{x}} f(\vec{x}, \dot{\vec{x}}, t) \tag{4.4.2}$$

We can approximate the solution of the Boltzmann equation by replacing the continuous phase space with a collection of N *pseudoparticles* with mass $M = \rho_0 V/N$. By solving in essence Newtonian equations of motion for the pseudo particles we can recover the resulting distribution of matter up to the resolution scale given by the number of particles.

$$\frac{d\vec{p}_i}{da} = -\frac{\vec{\nabla}\phi}{\dot{a}} \tag{4.4.3}$$

$$\frac{d\vec{x}_i}{da} = \frac{\vec{p}_i}{\dot{a}a^2} \tag{4.4.4}$$

4.4.2 How to put initial conditions?

The starting configuration of DM and gas particles is typically computed at linear stages of the perturbation revolution ($z \sim 100$). The smallest scales to resolve control the precise limit of the starting redshift – since the smallest scales reach the non-linear regime first.

We extract the statistics of the primordial perturbations from the temperature of CMB. The spatial correlations are measured to be Gaussian which means that all information is contained inside the mean and the two-point correlation function (or power spectrum).

Two define the initial conditions for second-order Newton equations, we need to specify the distribution and velocity fields.

Particle velocities depend on the initial distribution function of the matter which is taken to be thermal for baryons but might be different for DM (depending on the model). On larger scales, the gravitational potential causes bulk motion of the matter.

The power spectrum provides the amplitudes for the plane wave decomposition, but not the phases which are considered random. Hence, a single power spectrum corresponds to an infinite number of the density field realizations.

Translation between the density field and the discretized distribution of particles is done using the *Zeldovich approximation* [122].

We represent the proper positions of the particles as a combination of the Hubble flow and some *displacement field*:

$$\vec{x}(t) = a(t)\vec{q} + b(t)\vec{f}(\vec{q}) \tag{4.4.5}$$

The particles are assumed to move along the displacement field $\vec{f}(\vec{q})$ with velocity increasing according to the "growth" function $b(t)$. The Lagrangian coordinates $\vec{q}$ are equal to the initial positions of the particles and have a property that $\rho(\vec{q}) = \text{const}$. The proper density is computed using $\rho(\vec{x})d\vec{x} = \rho(\vec{q})d\vec{q}$:

$$\rho(\vec{x}) = \rho_q \left|\frac{d\vec{q}}{d\vec{x}}\right| = \rho_q \left|\frac{d\vec{x}}{d\vec{q}}\right|^{-1} = \rho_q \left|a(t)\delta_{ij} + b(t)\frac{df_i}{dq_j}\right|^{-1} = \tag{4.4.6}$$

$$= \rho_0 \left|\delta_{ij} + \frac{b(t)}{a(t)}\frac{df_i}{dq_j}\right|^{-1} \approx \rho_0 \left[1 - \frac{b(t)}{a(t)}\vec{\nabla}\vec{f}\right] \tag{4.4.7}$$

$$\delta(\vec{x}) \equiv \frac{\rho(\vec{x}) - \rho_0}{\rho_0} = -\frac{b(t)}{a(t)}\vec{\nabla}\vec{f} \tag{4.4.8}$$

In linear theory the perturbations are non-rotational and displacement field can be written as a gradient of the scalar field:

$$\delta(\vec{x}) = -\frac{b(t)}{a(t)}\Delta\Phi \tag{4.4.9}$$

Notice that this equation becomes the Poisson equation for the density perturbation (4.3.31) if we identify $b(t)\Phi \equiv -\frac{1}{4\pi G a \rho_0}\delta\phi$ at $t = 0$.

Finally, we obtain

$$\vec{x}(t) = a\vec{q} + \vec{\nabla}(b(t)\Phi) = a\vec{q} - \frac{\vec{\nabla}\delta\phi(\vec{q}, 0)}{4\pi G a \rho_0} \tag{4.4.10}$$

This equation provides the displacements of the perturbations at time t using the initial gravitational potential. Velocity field is given by a time derivative $\dot{\vec{x}}(t)$.

So, to define the initial conditions for the simulation, we can define a grid of pseudoparticle positions $\vec{q}_i$ and correct them according to the Zeldovich approximation (4.4.10).

However, it has been shown that Zeldovich approximation might be insufficient to correctly set the initial conditions on the smallest scales. A more precise approach is called the *Second-order Lagrangian Perturbation Theory* and is implemented in a code 2LPT [123].

Glass initial conditions. The choice of a regular grid for initial positions $\vec{q}_i$ can introduce artificial patterns in the simulation that survive until late times (Fig. 4.4).

The way to mitigate this is to choose an explicitly irregular grid. This can be done by applying an N-body code to a regular grid with the direction of the gravity reversed. Particles will attempt to separate as far as possible from each other while still keeping the density more or less unaffected.

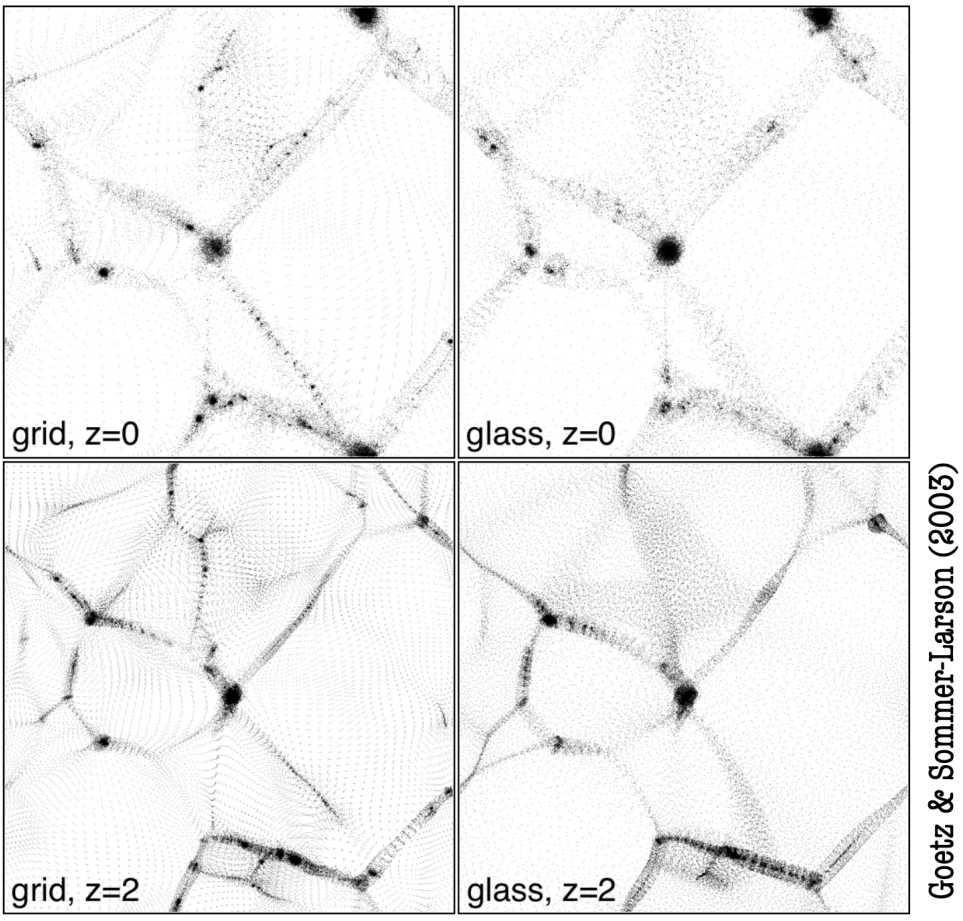

Figure 4.4: Initial conditions built from the evenly spaced *grid* (*left panels*) contain small-scale regularities that survive until late time. *Glass* initial conditions (*right panels*) have the same properties on large scales, but also are disordered on the small scales.

4.5 Numerical methods of N-body simulations

Direct summation approach. The force experienced by each pseudoparticle is the sum of the regular $1/r^2$ Newtonian forces. This poses a difficulty for computation when the particles are close to each other. So to avoid divergences, the gravitational interaction is regularized using *Plummer softening*:

$$\vec{F}_{ij} = GM^2 \frac{\vec{x}_i - \vec{x}_j}{(|\vec{x}_i - \vec{x}_j|^2 + \epsilon^2)^{3/2}} \tag{4.5.1}$$

This corresponds to smearing the particles such that the majority of their mass is contained inside a sphere of radius $\sim 3\epsilon$. This allows the particles to freely pass through each other ($\vec{F}_{ij}(\vec{x}_i = \vec{x}_j) = 0$ without causing numerical problems.

The total force for each particle can be obtained as a *direct sum* over all other particles and the system of equations can be solved by several numerical schemes, including Runge-Kutta, Predictor-Corrector or Leap-Frog (we defer the discussion of them to [45].

Unfortunately, despite the simplicity of this method, it suffers from the $O(N^2)$ complexity which limits its applicability to relatively small particle systems. Even though specialized hardware has been developed to efficiently compute the forces (see GRAPE computer [124], modern simulations do not use this method.

Tree methods. Using the same $1/r^2$ characteristic of the gravitational force, it is possible to make meaningful approximations that reduce the complexity of computation from $O(N^2)$ to $O(N \log N)$.

The main idea lies in separating the nearest neighbors from the particles located far away. Small errors in positions of the latter do not significantly influence the resulting force – hence, we can group them into *clusters* reducing the number of pairs to consider.

One the simplest ways to achieve this is to form a *binary tree* out of particles by grouping together the closest pairs, then pairs of pairs and so on. Each node of the tree is annotated with center-of-mass location, total mass, and its "size". Then, to compute the force one iterates over the tree starting from its root (the largest cluster) considering whether each given node is "far enough" to be considered a point particle. For a node of size L at a distance D this can be expressed as

$$\begin{cases} \text{if } L \ll D & \text{treat node as a point particle} \\ \text{else} & \text{repeat for its descendants} \end{cases} \tag{4.5.2}$$

The precision is controlled by some tolerance parameter equal to the ratio L/D that is "far enough". When tolerance is set to 0, the algorithm becomes the direct summation.

Higher accuracy can be achieved if a low-order multipole expansion of each node is computed in addition to the monopole contribution.

This approach significantly reduces the time of force computation, but at the same time at each simulation step, the tree needs to be rebuilt.

An evolution of this approach is represented by the Barnes-Hut algorithm [125] where the tree is built by successively splitting the simulation volume into octants until each of them contains 1 or 0 particles (Fig. 4.5). This structure allows to efficiently look for nearest neighbors and to only partially rebuild the tree when particles leave their cells.

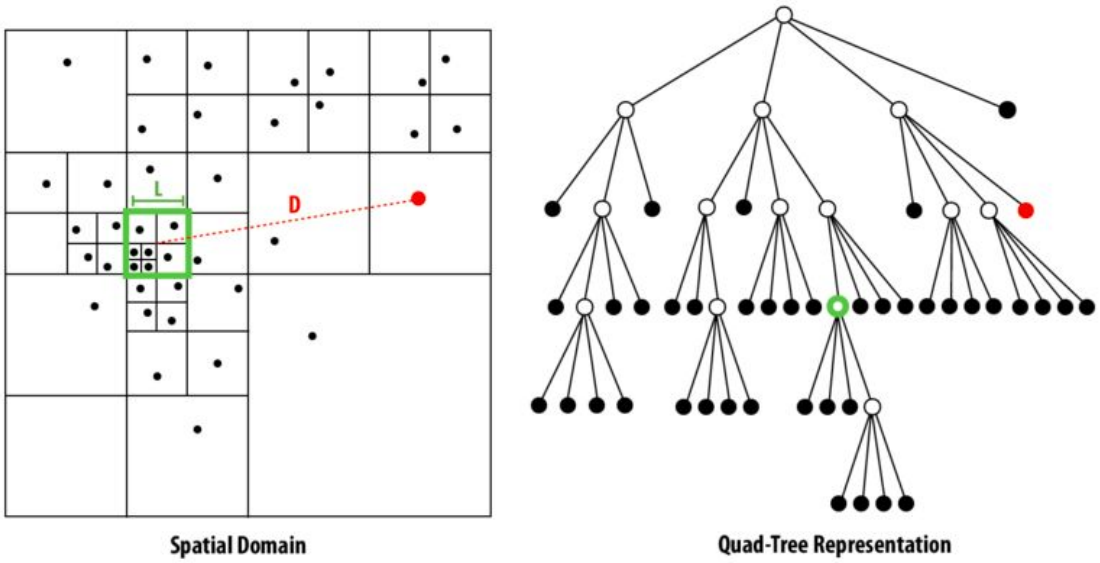

Figure 4.5: In Barnes-Hut algorithm [125], the particles are organized into hierarchical structure (Quad-Tree) that allows for efficient nearest-neighbour search and multipole expansion of the gravitational force.

Particle Mesh methods. An alternative approach to force computation uses the Fast Fourier transform to efficiently solve the Poisson equation for the gravitational potential and compute the resulting force from it.

The first step is to create a regular grid to evaluate the potential. The matter density is interpolated to this grid using some *kernel interpolation function* W:

$$\rho(\vec{g}_i) = \sum_j m_j W(\vec{x}_j - \vec{g}_i) \tag{4.5.3}$$

.

Given a density estimate, the Poisson equation is solved in the Fourier space:

$$\phi_k = -4\pi G a^2 \frac{\rho_k}{k^2} \tag{4.5.4}$$

Then the potential is transformed back to the real space and the force is computed using a discretized derivative operator. Finally, the force is interpolated back to the particles using the same kernel.

This method is very efficient because the Fast Fourier Transform for grid size N_g has the complexity of $O(N_g \log N_g)$. At the same time to achieve a significant speed-up compared to the tree methods, it is desired to have $N_g \ll N$ which significantly limits the force resolution.

P^3M methods. A combination of particle mesh and direct summation (or tree) methods can be used to achieve the best of both worlds. The limited force resolution of the particle mesh is compensated by the direct summation on short distances.

4.5.1 Interacting fluid

In the case of an interacting fluid, the Boltzmann equation governing the particle distribution gains the *collision term*

$$\frac{\partial f}{\partial t} + \dot{\vec{x}}\frac{\partial f}{\partial \vec{x}} - \vec{\nabla}\phi\frac{\partial f}{\partial \dot{\vec{x}}} = \left[\frac{df}{dt}\right]_c \tag{4.5.5}$$

The collision term encompasses all the information about the thermodynamical properties and phenomena occurring in the gas: pressure, viscosity, molecular and atomic processes. In the whole generality, one would need to introduce the distribution functions for all states possible in the system like atoms, ions, free electrons, etc. However, thanks to the separation of scales, we can assume that the quantum phenomena are local and the system can be adequately described by a distribution function of *matter* supplemented by position-dependent fields describing the "subgrid" physics (e.g., ionization fraction $\xi(\vec{x})$). One of the notable exceptions to this is the high energy radiation which can have a large mean-free path.

The equations of motion of gas can be expressed as a system containing the continuity, momentum and energy equations:

$$\frac{\partial \rho}{\partial t} + 3H\rho + \frac{1}{a}\vec{\nabla}\rho\vec{v} = -\Gamma_*\rho \tag{4.5.6}$$

$$\frac{\partial \vec{v}}{\partial t} + \frac{1}{a}(\vec{v}\cdot\vec{\nabla})\vec{v} + H\vec{v} = -\frac{1}{a\rho}\vec{\nabla}P - \frac{1}{a}\vec{\nabla}\phi \tag{4.5.7}$$

$$\frac{\partial E}{\partial t} + 2HE + \frac{1}{a}\vec{\nabla}((E+P)\vec{v}) = \frac{Q - \Lambda(\rho, T)}{\rho} \tag{4.5.8}$$

where Γ_* is the gas destruction rate due to star formation, E is the energy of gas per unit comoving volume, Q and Λ account for heating and cooling of the gas. As before, these equations need to be supplemented by the Poisson equation for the gravitational potential and the gas equation of state.

The approaches to solving this system are divided into two groups: Eulerian and Lagrangian methods. The difference between them lies in the way the fields are discretized.

In Eulerian methods, the equations are solved on a static grid and the interactions such as the flow of matter from point to point are computed. While Lagrangian methods are more similar to particle systems that are mobile themselves.

Eulerian methods. This approach is a standard of computational fluid dynamics. These methods are good at conserving integrals of motion and handling hydrodynamical shocks. However, equations are solved on a grid which limits the resolution. In the context of cosmological simulations, this is a problem because of the huge dynamical range: empty voids versus galaxies, stars, etc.

A modern approach to avoid the computational complexity at high resolution is called *Adaptive Mesh Refinement* (AMR) [126].

Lagrangian methods (Smoothed Particle Hydrodynamics). Since the highest resolution is required in regions of the matter collapse, assigning the fluid variables to the particles moving with the fluid is a natural solution. However, this requires the introduction of *artificial viscosity* to handle shocks.

Most methods are based on the *Smoothed Particle Hydrodynamics* by Gingold and Monaghan [127]. The most important idea of the SPH is the interpolation scheme that connects the "physical" fields with their values at a set of particles.

First, let's introduce an estimate of the physical field $A(\vec{x},t)$ smoothed over scale h:

$$A_S(\vec{x},t) = \int d\vec{y} A(\vec{y},t) W(\vec{x}-\vec{y},h) \tag{4.5.9}$$

$W(\vec{x},h)$ is the kernel interpolation function satisfying conditions

$$\int W(\vec{x}-\vec{y},h) d\vec{y} = 1 \tag{4.5.10}$$

$$\lim_{h\to 0} W(\vec{x}-\vec{y},h) = \delta(\vec{x}-\vec{y}) \tag{4.5.11}$$

The definition of the smoothed field has the advantage that the derivatives of the smoothed quantities are computed through the derivatives of the kernel function

$$\vec{\nabla} A_S(\vec{x},t) = \int d\vec{y} A(\vec{y},t) \vec{\nabla}_y W(\vec{x}-\vec{y},h) \tag{4.5.12}$$

If we discretize the integral by splitting the volume into particles, we can compute the physical quantities using the expression

$$A(\vec{y}) = \sum_j m_j \frac{A_j}{\rho_j} W(\vec{y} - \vec{x}_j, h) \tag{4.5.13}$$

$$\rho(\vec{y}) = \sum_j m_j W(\vec{y} - \vec{x}_j, h) \tag{4.5.14}$$

Example derivation of gas equations can be found e.g. in a SPH review by Monaghan [128].

$$\frac{d\vec{v}_i}{dt} = -\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \vec{\nabla}_i W(\vec{x}_i - \vec{x}_j, h) \vec{\nabla} \phi \tag{4.5.15}$$

$$\frac{d\epsilon_i}{dt} = \sum_j \left(\frac{P_i}{\rho_i^2} + \frac{\Pi_{ij}}{2} \right) (\vec{v}_i - \vec{v}_j) \cdot \vec{\nabla}_i W(\vec{x}_i - \vec{x}_j, h) + \frac{Q_i - \Lambda_i}{\rho_i} \tag{4.5.16}$$

where Π_{ij} represents the mentioned artificial viscosity term.

In this approach, the shocks are spread over several smoothing lengths $\sim 3h$. Because of this, most modern implementations use a spatially-variable smoothing scale $h \propto \rho^{-1/3}$.

4.5.2 Subgrid physics

Equations describing the evolution of baryonic fluid 4.5.6 contain external functions like gas destruction, cooling, and heating rates.

Cooling depends on the thermodynamical quantities as well as chemical composition. It happens through the emission of radiation and redistribution of the kinetic energy between particles during collisions, ionizations, and recombinations. A typical description of the cooling involves a hot cloud of gas that radiates away from the thermal energy. But this cloud is surrounded by medium so ultimately there is no "thermal sink" and energy is just redistributed between the components of the plasma.

Plasma of primordial composition at early times consists from free electrons e^-, chemical elements like H and He and their ionized states $\mathrm{H}^+, \mathrm{He}^+, \mathrm{He}^{++}$. From primordial nucleosynthesis (Sec. 3) we know that Hydrogen and Helium constitute more than 99% of the nuclei in the Universe, so these two species are enough to satisfactorily describe gas at large scales. Other chemical elements could play a role in special conditions like in galaxies or stars. The cooling rates for this very important case is shown in Fig. 4.6.

Note that the cooling function is proportional to the squared density of the gas (since two-body processes are dominating the plasma). This means that cooling is much more effective in the denser regions. An overdensity also represents a minimum of gravitational potential that attracts particles. In the case of collisionless DM, particles are prevented from falling directly into the center of the overdensity because of the angular momentum conserved for all particles individually. On the other hand, radiating particles like baryons

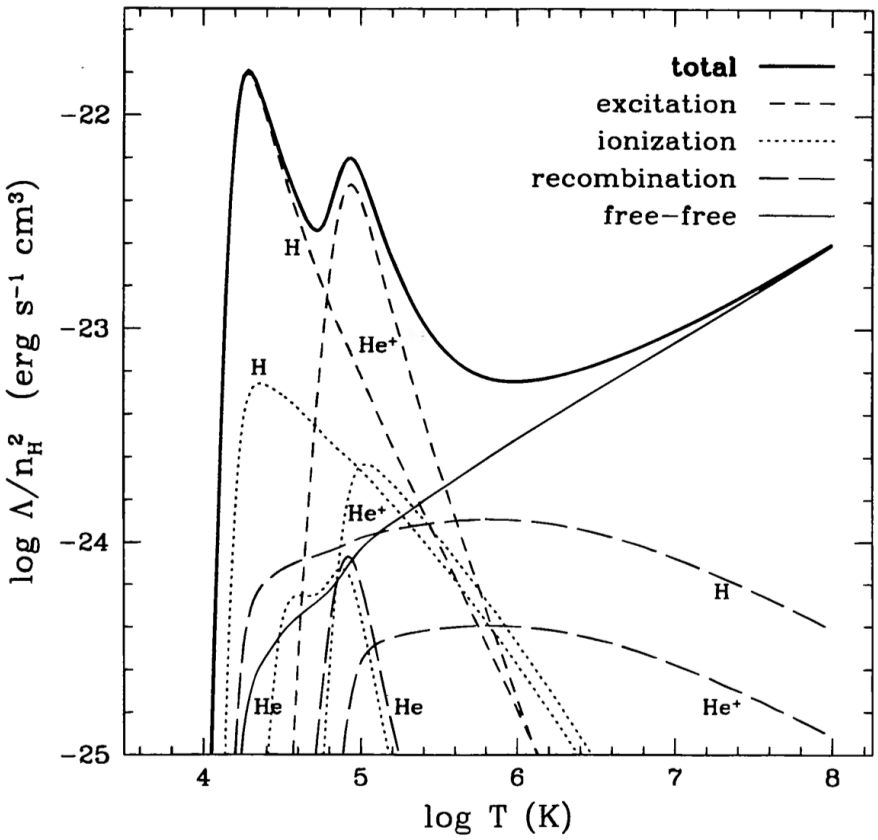

Figure 4.6: Cooling rates as a function of temperature for a primordial composition gas in collisional equilibrium [129]. The heavy solid line shows the total cooling rate. The cooling is dominated by collisional excitation (short-dashed lines) at low temperatures and by free-free emission (thin solid line) at high temperatures. Long-dashed lines and dotted lines show the contributions of recombination and collisional ionization, respectively.

will on average move closer to the center. This, in turn, will increase the density and cooling further.

This mechanism leads to the generation of compact, dense baryonic structures that eventually become stars and galaxies. DM plays an important role in providing the deep potential wells for baryons to fall in.

Star formation. One of the factors playing against the cooling and contraction of baryonic overdensities is the heating due to the increasing pressure. Particles falling in the potential well gain some momentum that compensates for the energy radiated away. At some point, the density and temperature in the center of the cloud might overcome the Coulomb barrier of the nuclei and spark fusion. This would start a chain reaction since fusion is a potent source of energy radiation that will further heat the surrounding plasma.

As a net result, a star is born that on cosmological scales can be treated as a point-like radiation source. The precise details of this process have not been modeled and in any

case, would require tremendous resolution in the context of the cosmological simulation. Typically a single pseudoparticle in simulation is orders of magnitude heavier than stars. Because of this, star formation is treated as a local process happening "inside" a pseudo particle that converts baryons into star particles. Naturally, this requires additional models to specify the conditions and rate of star formation.

For galaxy formation, the feedback processes from the stellar winds and explosions are important and require dedicated modeling. However, they are relevant only to relatively small scales.

Photo-heating and -ionization. Nuclear reactions inside the stars serve as radiation sources capable of heat and ionize the surrounding gas. In general, the treatment of this phenomenon requires solving the radiative transfer equations. However, given a sufficiently homogeneous distribution of sources that is dense in relation to the mean free path of the radiation, we could treat this radiation as a homogeneous field covering all space – the cosmic *Ultraviolet Background* (UVB).

The interaction of the UVB with the gas depends on the cross-section for interaction of a particular atom specie $\sigma_i(\nu)$ and can be encapsulated in photoionization $\Gamma_{\gamma i}$ and photoheating ϵ_i rates:

$$\Gamma_{\gamma i} \equiv \int_{\nu_i}^{\infty} \frac{4\pi J(\nu)}{h\nu} \sigma_i(\nu) d\nu \qquad [\mathrm{s}^{-1}] \tag{4.5.17}$$

$$\epsilon_i \equiv \int_{\nu_i}^{\infty} \frac{4\pi J(\nu)}{h\nu} \sigma_i(\nu) \left(h\nu - \nu_i\right) d\nu \qquad [\mathrm{ergs\,s}^{-1}] \tag{4.5.18}$$

where $J(\nu)$ is the intensity of the UVB (in units of $\mathrm{ergs}^{-1}\mathrm{cm}^{-2}\mathrm{sr}^{-1}\mathrm{Hz}^{-1}$). These rates describe how frequently atoms are ionized by radiation and what excess energy is introduced into plasma by photoelectrons.

Stars might not be solely responsible for the formation of the UVB. Quasars have been long recognized as important sources of ionizing radiation. Other sources like decaying/annihilating DM or evaporating black holes might also participate.

The intensity of the UVB is another quantity that has to be determined from external models. When supplied, it yields the total heating rate of the gas:

$$Q = \sum n_i \epsilon_i \tag{4.5.19}$$

4.6 Simulation codes

Since typical cosmological simulations include both dark matter and gas evolved at the same time, usually the codes layer some kind of an N-body method for DM with either SPH or AMR handling the baryons. Depending on the desired processes, additional numerical

ingredients are added, ranging from subgrid physics for star and black hole formation to radiative transfer schemes for the propagation of radiation.

4.6.1 Code used in this work

In our simulations, we use a modified version of the code P-Gadget3 by Volker Springel et al. [45]. This code uses a TreePM method for collisionless dynamics that complements the accuracy of the tree method on small scales with the efficiency of the particle mesh on large scales. SPH in an explicitly entropy-conserving formulation is used for gas dynamics.

Since we are interested in the physics of the Intergalactic Medium which is either only slightly overdense or underdense, it is not necessary to model precisely the dense regions. This allows us to cut on computing time by using an extremely simple placeholder of the star formation model.

Star formation heuristic. Stars occur only in dense regions and besides their role in the formation of UVB does not participate in structure formation. We use a simple criterion of transforming gas particles into stars: if the temperature of the particle is $< 10^5 K$ while its overdensity is larger than some *critical overdensity* (which we take equal to 10^3).

4.7 Comparison between observations and simulations of Large Scale Structure

As the result of the approach presented in the previous sections it is possible to produce simulations of the LSS at different scales. For example, Fig. 4.7 shows the comparison between the Millennium simulations [130] (red) with the 2dFGRS galaxy survey [115] (blue). We see the remarkable similarity between them, without hint it is not possible to distinguish simulations from observations.

To be more quantitative, let us look at the galaxy-galaxy correlations function produced in the modern IllustrisTNG simulation [132], see Fig. 4.8. Comparing it to the data of the Sloan Digital Sky Survey we see the excellent agreement between them. These, and other comparisons between the simulations and observations convince us that ΛCDM cosmology perfectly fits the data on the LSS.

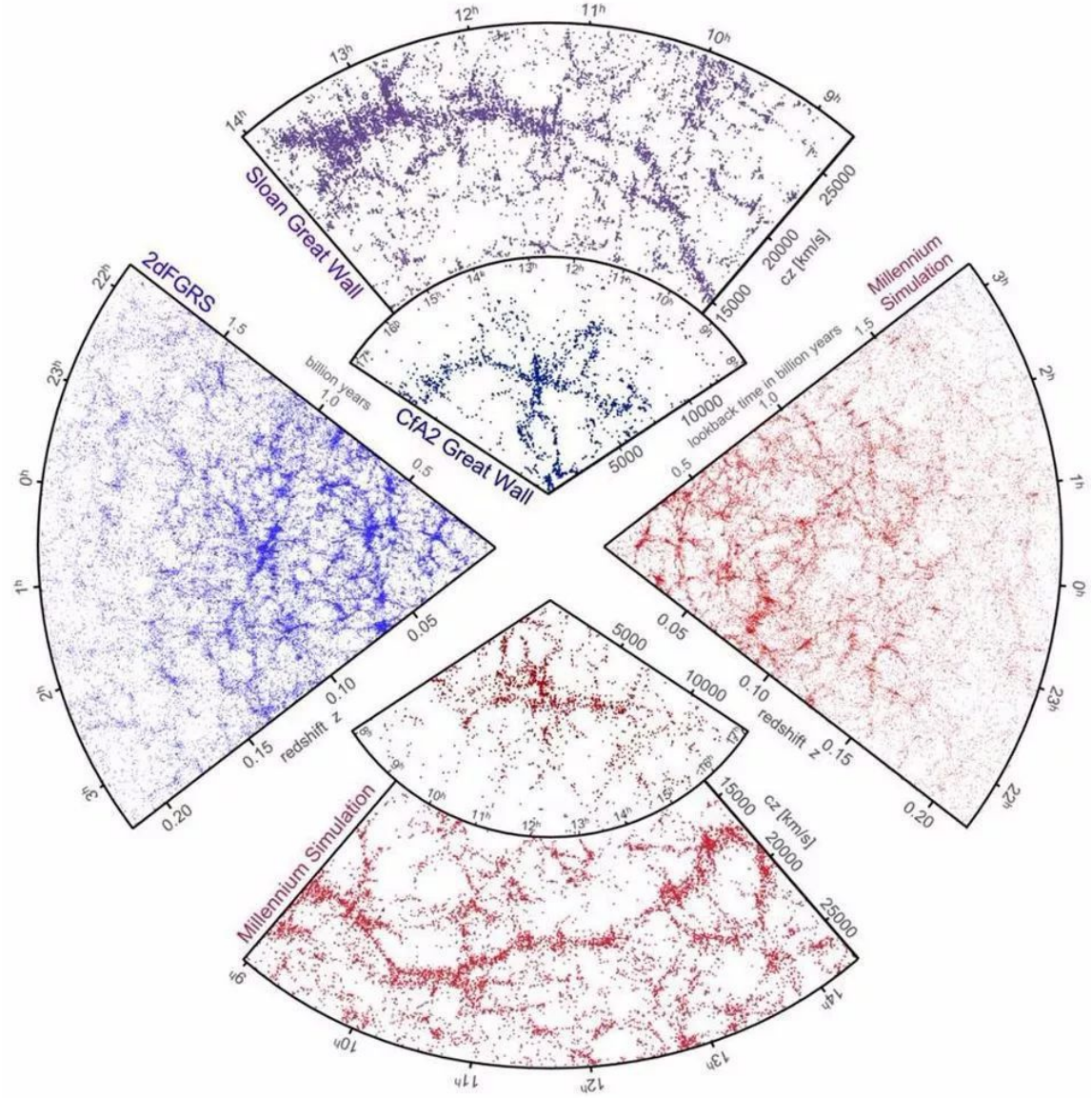

Figure 4.7: Comparison of the galaxy surveys with the simulations. The small slice at the top shows the CfA Great Wall, with the Coma cluster at the centre. Drawn to the same scale is a small section of the SDSS, in which an even larger Sloan Great Wall has been identified. This is one of the largest observed structures in the Universe, containing over 10,000 galaxies and stretching over more than 1.37 billion light years. The cone on the left shows one-half of the 2dFGRS, which determined distances to more than 220,000 galaxies in the southern sky out to a depth of 2 billion light years. The SDSS has a similar depth but a larger solid angle and currently includes over 650,000 observed redshifts in the northern sky. At the bottom and on the right, mock galaxy surveys constructed using semi-analytic techniques to simulate the formation and evolution of galaxies within the evolving dark matter distribution of the Millennium simulation are shown, selected with matching survey geometries and magnitude limits. *Credit: Springel et al.* [131]

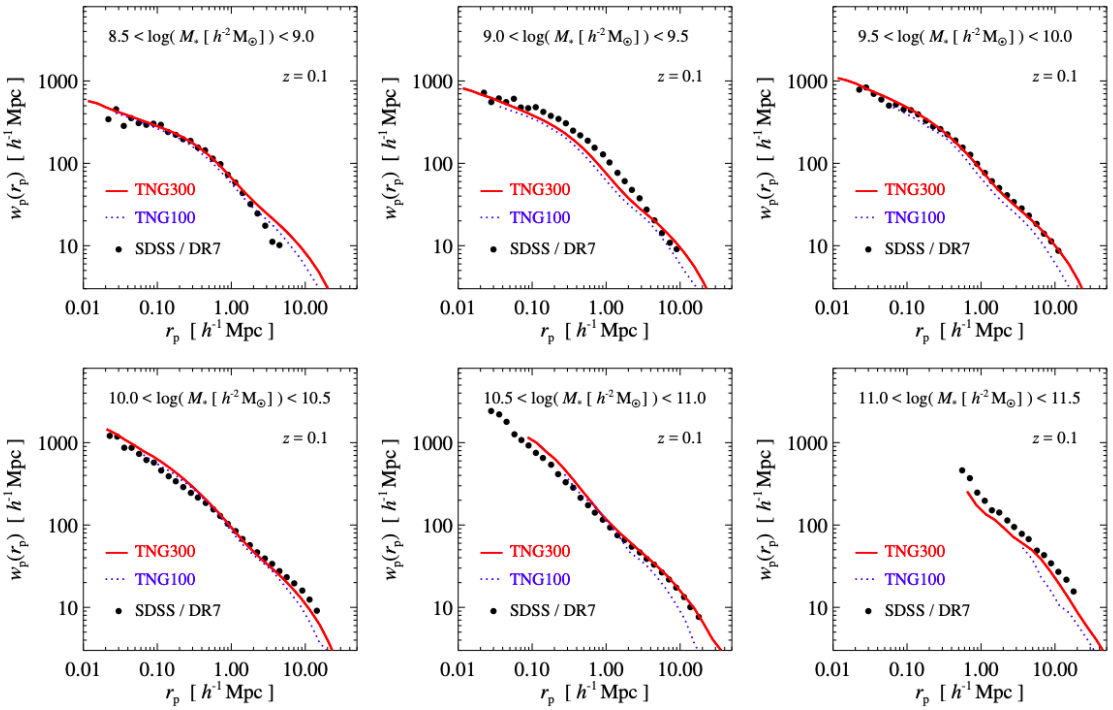

Figure 4.8: Comparison of the projected two-point galaxy correlation functions of the TNG300 simulation (solid) and the TNG100 simulation (dotted) at $z = 0.1$ with the Sloan Digital Sky Survey, in six different stellar mass ranges. Taken from [132].