Cosmology with numerical simulations

Lauro Moscardini and Klaus Dolag

Abstract The birth and growth of cosmic structures is a higly non-linear phenomenon that needs to be investigated with suitable numerical simulations. The main goal of these simulations is to provide robust predictions, which, once compared to the present and future observations, can allow to constrain the main cosmological parameters. Different techniques have been proposed to follow both the gravitational interaction inside cosmological volumes, and the variety of physical processes acting on the baryonic component only. In this paper we review the main characteristics of the numerical schemes most commonly used in the literature, discussing their pros and cons and summarising the results of their comparison.

1 Introduction

Numerical simulations have become in the last years one of the most effective tools to study and to solve astrophysical problems. The computation of the mutual gravitational interaction between a large set of particles is a problem that cannot be investigated with analytical techniques only. Therefore it represents a good example of a problem where computational resources are absolutely fundamental.

Thanks to the enormous technological progress in the recent years, the available facilities allow now to afford the problem of gravitational instability, which is the basis of the accepted model of cosmic structure formation, with a very high mass and space resolution. Moreover the development of suitable numerical techniques permits to include a realistic treatment of the majority of the complex physical pro-

Lauro Moscardini
Dipartimento di Astronomia, Università di Bologna, via Ranzani 1, I-40127 Bologna, Italy
e-mail: lauro.moscardini@unibo.it

Klaus Dolag
Max-Planck-Institut für Astrophysik, P.O. Box 1317, D-85741 Garching, Germany
e-mail: kdolag@mpa-garching.mpg.de

cesses acting on the baryonic component, which is directly related to observations. For this reason, the numerical simulations are now used not only to better understand the general picture of structure formation in the universe, but also as a tool to validate cosmological models and to investigate the possible presence of biases in real data. In some sense they substitute the laboratory experiments which are in practice impossible for cosmology, given the uniqueness of the universe.

The plan of this paper is as follows. In Section 2 we will discuss the schemes proposed to follow the formation and evolution of cosmic structures when the gravity only is in action: in particular, after presenting the model equations in Section 2.1, we will introduce the Particle-Particle method (Section 2.2), the Particle-Mesh method (Section 2.3), the Tree code (Section 2.4), and the so-called Hybrid methods (Section 2.5). Section 3 is devoted to the presentation of the numerical codes used to solve the hydrodynamical equations related to the baryonic component, introduced in Section 3.1. More in detail, Section 3.2 discusses the characteristics of the most used Lagrangian code, the Smoothed Particle Hydrodynamics, while Section 3.3 introduces the bases of the methods based on grids, the Eulerian codes.

2 N-Body codes

2.1 The model equations

In order to write the equations of motion determining the gravitational instability leading to the formation and evolution of cosmic structures, it is necessary to choose the underlying cosmological model, describing the expanding background universe, where $a = 1/(1+z)$. In the framework of General Relativity this means to assume a Friedmann-Lemaître model, with its cosmological parameters, namely the Hubble parameter H_0 and the various contributions coming from baryons, dark matter, dark energy/cosmological constant to the total density parameter Ω_0.

Many different observations are now giving a strong support to the idea that the majority of the matter in the universe is made by cold dark matter (CDM), i.e. non-relativistic collisionless particles, which can be described by their mass m, comoving position $\mathbf{x}$ and momentum $\mathbf{p}$. The time evolution of the phase-space distribution function $f(\mathbf{x},\mathbf{p},t)$ is given by the coupled solution of the Vlasov equation

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{ma^2}\nabla f - m\nabla\Phi\frac{\partial f}{\partial \mathbf{p}} = 0 \tag{1}$$

and of the Poisson equation

$$\nabla^2\Phi(\mathbf{x},t) = 4\pi \mathrm{G} a^2 \left[\rho(\mathbf{x},t) - \bar{\rho}(t)\right]. \tag{2}$$

Here Φ represents the gravitational potential and $\bar{\rho}(t)$ the mean background density. The proper mass density

$$\rho(\mathbf{x},t) = \int f(\mathbf{x},\mathbf{p},t) d^3p \tag{3}$$

is the integral over the momenta $\mathbf{p} = ma^2\dot{\mathbf{x}}$.

The solution of this high-dimension problem is standardly obtained by using a finite set of N_p particles to trace the global matter distribution. For these tracers it is possible to write the usual equations of motion, which in comoving coordinates read:

$$\frac{d\mathbf{p}}{dt} = -m\nabla\Phi \tag{4}$$

and

$$\frac{d\mathbf{x}}{dt} = \frac{\mathbf{p}}{ma^2}. \tag{5}$$

Introducing the proper peculiar velocity $\mathbf{v} = a\dot{\mathbf{x}}$ these equations can be written as

$$\frac{d\mathbf{v}}{dt} + \mathbf{v}\frac{\dot{a}}{a} = -\frac{\nabla\Phi}{a}. \tag{6}$$

The time derivative of the expansion parameter, $\dot{a}$, is given by the Friedmann equation, once the cosmological parameters are assumed.

In the following subsections we will present some of the standard approaches used to solve the N-body problem.

2.2 The Particle-Particle (PP) method

This is certainly the simplest possible method because it makes direct use of the equations of motion plus the Newton's gravitational law. In this approach the forces on each particle are directly computed by accumulating the contributions of all remaining particles.

At each time-step Δt the following operations are repeated:

- clearing of the force accumulators: $F_i = 0$ for $i = 1,...,N_p$, where N_p is the number of particles;
- accumulation of the forces considering all $N_p(N_p-1)$ pairs of particles: $F_{ij} \propto m_i m_j / r_{ij}^2$, where r_{ij} is the distance between two particles having masses m_i and m_j, respectively:
$$F_i = F_i + \sum F_{ij}; \;\; F_j = F_j + \sum F_{ij}; \tag{7}$$
- integration of the equations of motion to obtain the updated velocities v_i and positions x_i of the particles:

$$v_i^{\rm new} = v_i^{\rm old} + F_i \Delta t / m_i \ , \tag{8}$$

$$x_i^{\rm new} = x_i^{\rm old} + v_i \Delta t \ ; \tag{9}$$

- updating of the time counter: $t = t + \Delta t$.

Even if implementing this code is extremely easy from a numerical point of view (see an example in the Appendix of [3]), its computational cost is quite large, being proportional to N_p^2: for this reason its application is in practice forbidden for problems with a very large number of particles.

Notice that in principle this method would compute the exact newtonian force, used then to estimate the particles' acceleration: for this reason the PP method can be considered the most accurate N-body technique. However, in order to avoid the divergence at very small scales, the impact parameter must be reduced by introducing a softening parameter ε in the equation for the gravitational potential Φ:

$$\Phi = -Gm_p/(r^2 + \varepsilon^2)^{1/2} \ : \tag{10}$$

in some sense this corresponds to assign a finite size to each particle, which can be considered as a statistical representation of the total mass distribution. Typical choices for ε range between 0.02-0.05 times the mean interparticle distance. As a consequence, the frequency of strong deflections is also reduced, decreasing the importance of the spurious two-body relaxation, which is generated by the necessarily small number of particles used in the simulations, many orders of magnitude smaller that the number of collisionless dark matter particles really existing in the universe.

As said, the largest limitation of this method is its scaling as N_p^2. An attempt to overcome this problem has been done by building a special-purpose hardware, called GRAPE (GRAvity PipE) [13]. This hardware is based on custom chips that compute the gravitational force with a hardwired force law. Consequently this device can solve the gravitational N-body problem adopting the direct sum with a computational cost which is extremely smaller than for traditional processors.

Few words, which are valid also for most of the following methods, must be spent about time-stepping and integration. In general the accuracy obtained when evolving the system depends on the size of the time step Δt and on the integrator scheme used. Finding the optimum size of time step is not trivial. A possible choice is given by

$$\Delta t = \alpha \sqrt{\varepsilon / |\mathbf{F}/m_p|} \ , \tag{11}$$

where the force is the one obtained at the previous time step, ε is a length scale associated to the gravitational softening, and α is a suitable tolerance parameter. Alternative and more accurate criteria are discussed in [19]. To update velocities and then positions, it is necessary to integrate first-order ordinary differential equations, once the initial conditions are specified: many methods, classified as explicit or implicit, are available, ranging from the simplest Euler's algorithm to the more accurate Runge-Kutta method (see for example [20] for an introduction to these methods).

2.3 The Particle-Mesh (PM) method

In this method a mesh is used to describe the field quantities and to compute their derivatives. Thanks to the structure of the Poisson equation and to the assumption that the considered volume is a fair sample of the whole universe, it is convenient to re-write all relevant equations in the Fourier space, where we can take advantage of the Fast Fourier Techniques: this will allow a strong reduction of the CPU time necessary for each time step. This improvement in the computational cost is, however, paid with a loss of accuracy and resolution: the PM method cannot follow close interactions between particles on small scales. In fact, using a grid to describe the field quantities (like density, potential and force) does not allow a fair representation on scales smaller than the intergrid distance.

Since the introduction of a computational mesh is equivalent to a local smoothing of the field, for the PM method it is not necesssary to adopt the softening parameter in the expression of the force and/or gravitational potential.

Going in more detail, each time step for the PM method is composed of the following operations:

- computation of the density at each grid point starting from the particles' spatial distribution;
- solution of the Poisson equation for the potential;
- computation of the force on the grid points;
- estimation of the force at the positions of each particle using a suitable interpolation scheme;
- integration of the equations of motion.

If the computational volume is a cube of side L and N_p is the number of particles having equal mass m_p, a regular three-dimensional grid with M nodes per direction is built: therefore the grid spacing is $\Delta \equiv L/M$. Each grid point can be identified by a tern of integer numbers (i, j, k), such that its spacial coordinates are $\mathbf{x}_{i,j,k} = (i\Delta, j\Delta, k\Delta)$, with $i, j, k = 1, ..., M$.

2.3.1 Density computation

The particle mass decomposition at the grid nodes represents one of the critical steps for the PM method, both in terms of resolution and computational cost.

The mass density ρ at the grid point $\mathbf{x}_{i,j,k}$ can be written as

$$\rho(\mathbf{x}_{i,j,k}) = m_p M^3 \sum_{l=1}^{N_p} W(\delta\mathbf{x}_l) \,, \tag{12}$$

where W is a suitable interpolation function and $\delta\mathbf{x}_l = \mathbf{x}_l - \mathbf{x}_{i,j,k}$ is the distance between the position of the l-th particle and the considered grid point.

The choice of W is related to the accuracy of the required approximation: of course, the higher the number of grid points involved in the interpolation, the better

the approximation. For problems in higher dimensions, it is possible to write the function W as product of more functions, each of them depending only on the displacement in one dimension: $W_{i,j,k} = w_i w_j w_k$. In order of increasing accuracy, the most commonly adopted interpolation functions are:

- "nearest-grid-point" (NGP): in this case the mass of each particle is totally assigned to the nearest grid point only. As a consequence the density shows a discontinuity every time a particle crosses the grid borders. The NGP interpolation function reads:
$$w_i = 1, \;\; M\,|\delta x_i| \leq 1/2 \; ; \tag{13}$$
- "cloud-in-cell" (CIC): the mass of each particle is assigned to two (i.e. $2^3 = 8$ in the 3D case) nearest points in an inversely proportional way with respect to its distance from the grid point: in this way the density varies with continuity when a particle crosses a cell border, but its gradient is still discontinuous. The CIC interpolation function can be written as:
$$w_i = 1 - M\,|\delta x_i| \,, \;\; M\,|\delta x_i| \leq 1 \; ; \tag{14}$$
- "triangular-shaped-cell" (TSC): the mass decomposition involves three (i.e. $3^3 = 27$ in the 3D case) nearest points. In this way also the density gradient varies smoothly during the cell border crossing; on the contrary the second derivative remains discontinuous. The TSC interpolatation function can be expressed as:
$$w_i = \begin{cases} 3/4 - M^2\,|\delta x_i|^2 & M\,|\delta x_i| \leq 1/2 \\ (1/2)\,(3/2 - M\,|\delta x_i|)^2 & 1/2 \leq M\,|\delta x_i| \leq 3/2 \end{cases} . \tag{15}$$

Of course the interpolating functions can be easily extended to higher orders, providing gradually continuous derivatives of higher orders and better accuracy. However, because of the increasing number of involved grid points, the resulting interpolation would be more computationally demanding.

Notice that, in order to conserve the momentum, the same interpolation schemes W adopted here for the density computation have to be used to obtain the components of the force at the position of each particle, once the force on the mesh will be obtained (see above). In this case:

$$F_x(\mathbf{x}_l) = m_p \sum_{i,j,k=1}^{M} W(\delta \mathbf{x}_l) F_x(\mathbf{x}_{i,j,k}) \; ; \tag{16}$$

$$F_y(\mathbf{x}_l) = m_p \sum_{i,j,k=1}^{M} W(\delta \mathbf{x}_l) F_y(\mathbf{x}_{i,j,k}) \; ; \tag{17}$$

$$F_z(\mathbf{x}_l) = m_p \sum_{i,j,k=1}^{M} W(\delta \mathbf{x}_l) F_z(\mathbf{x}_{i,j,k}) \; . \tag{18}$$

2.3.2 Poisson equation

The application of the Fast Fourier Transforms (FFTs) allows to solve the Poisson equation in a much easier way in the Fourier space, where it becomes:

$$\Phi_k = G_k \delta_k \; ; \tag{19}$$

here Φ_k and δ_k are the Fourier transforms of the gravitational potential and of the density contrast, respectively. In the previous equation, G_k represents a suitable Green function for the Laplacian, for which, in the case of a continuous system, a good expression is given by $G_k \propto k^{-2}$; alternative expressions for the Laplacian giving a better approximation are also available. Using a discrete system on a grid introduces errors and anisotropies in the computation of the force. In order to reduce this problem, it is necessary to find an expression for G_k which corresponds to the desired shape for the particles in the configuration space and which minimizes the errors with respect to a reference force: for this reason the optimal Green function depends on the chosen shape and interpolation scheme.

2.3.3 Force calculation

In a typical PM scheme the computation of the forces at each grid point requires differentiating the potential Φ to derive the i-th force component:

$$F_i(\mathbf{x}) = -m_p \frac{d\Phi}{dx_i} \; . \tag{20}$$

This can be done using the finite difference schemes, largely adopted to numerically solve differential equations. The approximation depends on the number of grid points involved in the computation. As examples, at the lowest order ("two-point centered") the x-component of the force can be obtained as:

$$\frac{F_x(\mathbf{x}_{i,j,k})}{m_p} = \frac{\Phi(\mathbf{x}_{i-1,j,k}) - \Phi(\mathbf{x}_{i+1,j,k})}{2\Delta} \; , \tag{21}$$

while the four-point approximation is given by

$$\frac{F_x(\mathbf{x}_{i,j,k})}{m_p} = \frac{2\left(\Phi(\mathbf{x}_{i+1,j,k}) - \Phi(\mathbf{x}_{i-1,j,k})\right)}{3\Delta} - \frac{\Phi(\mathbf{x}_{i+2,j,k}) - \Phi(\mathbf{x}_{i-2,j,k})}{12\Delta} \; . \tag{22}$$

It is possible to increase the resolution of a PM scheme by computing the force on a grid shifted with respect to the one used for the computation of potential. For example in the case of the "two-point centered" approximation (see eq.21), this "staggered mesh" technique reads:

$$\frac{F_x(\mathbf{x}_{i,j,k})}{m_p} = \frac{\Phi(\mathbf{x}_{i-1/2,j,k}) - \Phi(\mathbf{x}_{i+1/2,j,k})}{\Delta} \; . \tag{23}$$

It is well known that the finite difference schemes introduce truncation errors in the solutions. A way to avoid this problem and to gain a better accuracy is to obtain the forces directly from the gravitational potential in Fourier space: $\mathbf{F}_k = -i\mathbf{k}\Phi_k$. In this case it is necessary to inverse-transform separately each single component of the force, using more frequently the FFT routines.

2.3.4 Pros and cons

The big advantage of the PM method is its high computational speed: in fact the number of operations scales as $N_p + N_g \log(N_g)$, where N_p is the number of particles and $N_g = M^3$ the number of grid points. The small dynamical range, strongly limitated by the number of grid points and by the corresponding memory occupation, represents the largest problem of the method: only adopting hybrid methods (see above), it is possible to reach the sufficient resolution necessary in cosmological simulations.

2.4 Tree codes

Thanks to its computational performance and accuracy, this method represents today the favorite tool for cosmological N-body simulations. The idea to solve the N-body problem is based on the exploitation of a hierarchical multipole expansion, the so-called tree algorithm. The speed up is obtained by using, for sufficiently distant particles, a single multipole force, in spite of computing every single distance, as required for methods based on direct sum. In this way the sum ideally reduces to $N_p \log(N_p)$ operations, even if, for gravitationally evolved structures, the scaling can be less efficient.

In practice, the multiple expansion is based on a hierarchical grouping that is obtained, in the most common algorithms, by subdividing in a recursive way the simulation volume. In the approach suggested by [2] a cubical root node is used to encompass the full mass distribution. This cube is then repeatedly subdivided into eight daughter nodes of half the side-length each, until one ends up with 'leaf' nodes containing single particles. The computation of the force is then done by "descending" the tree. Starting from the root node, the code evaluates, by applying a suitable criterion, whether or not the multipole expansion of the node provides an accurate enough partial force: in case of a positive answer, the tree descent is stopped and the multipole force is used; in the negative case, the daughter nodes are considered in turn. Usually the opening criterion is based on a given fixed angle, whose choice controls the final error: typically one assumes the angle to be ≈ 0.5 rad. Obviously, the smaller and the more distant the particles' groups, the more accurate the assumption of multipole expansion.

2.5 Hybrid methods

Having the previous techniques quite different numerical properties with opposite pros and cons, it is possible to combine them to build new algorithms (called hybrid methods), possibly mantaining the positive aspects only. A first attempt has been done with the P^3M code, which combines the high accuracy of the direct sum implemented by the PP method at small scale with the speed of the PM algorithm to compute the large-scale interactions. An improved version of this code has been also proposed, where spatially adaptive mesh refinements are possible in regions at very high density. This algorithm, called AP^3M [6], has been used to run several cosmological simulations, including the Hubble Volume simulations [9].

In the last-generation codes like GADGET[23], hybrid codes are built by replacing the direct sum with tree algorithms: the so-called TreePM. In this case, the potential is explicitly split in Fourier space into a long-range and a short-range part according to $\Phi_k = \Phi_k^{\rm long} + \Phi_k^{\rm short}$, where

$$\Phi_k^{\rm long} = \Phi_k \exp(-k^2 r_{\rm s}^2) \; ; \tag{24}$$

here $r_{\rm s}$ corresponds to the spatial scale of the force-split. The long-range potential can be computed very efficiently with mesh-based Fourier methods, like in the PM method. The short-range part of the potential can be solved in real space by noting that for $r_{\rm s} \ll L$ the short-range part of the real-space solution of the Poisson equation is given by

$$\Phi^{\rm short}(\mathbf{x}) = -G \sum_i \frac{m_i}{\mathbf{r}_i} {\rm erfc}\left(\frac{\mathbf{r}_i}{2r_{\rm s}}\right). \tag{25}$$

In the previous equation $\mathbf{r}_i$ is the distance of any particle i to the point $\mathbf{x}$. Thus the short-range force can be computed adopting the tree algorithm, except that the force law is modified by a long-range cut-off factor.

This approach allows to largely improve the computational performance compared to ordinary tree methods, maintaining all their advantages: the very large dynamical range, the insensitivity to clustering, and the precise control of the softening scale of the gravitational force.

2.6 Initial conditions and simulation setup

As said, the N-body simulations are a tool often used to produce, once a given cosmological model is assumed, the predictions, which can be directly compared to real observational data to falsify the model itself. For this reason it is necessary to have a robust method to assign to the N_p particles the correct initial conditions, corresponding to a fair realisation of the desired cosmological model. The standard inflationary scenarios predict that the initial density fluctuations are a random field with an almost Gaussian distribution (for non-Gaussian models and the corresponding initial

conditions, see for example [11]). In this case the field is completely defined by its power spectrum $P(k)$, whose shape is related to the choices made about the underlying cosmological model and the nature of dark matter.

The generation of initial conditions with a Gaussian distribution and a given $P(k)$ can be easily obtained exploiting the characteristics of a Gaussian distribution. In fact, as discussed by [1], in order that a generic field $F(x)$ to be strictly Gaussian, all its different Fourier spatial modes F_k have to be reciprocally independent, to have random phases θ_k and to have amplitude distributed according to a Rayleigh distribution:

$$P(|F_k|,\theta_k)d|F_k|d\theta_k = \exp\left(-\frac{|F_k|^2}{2P(k)}\right)\frac{|F_k|d|F_k|}{P(k)}\frac{d\theta_k}{2\pi}\ . \tag{26}$$

The real and imaginary parts of F_k are then reciprocally independent and Gaussian distributed. Consequently, it is sufficient to generate complex numbers with a phase randomly distributed in the range $[0,2\pi]$ and with an amplitude normally distributed with a variance given by the desired spectrum.

To obtain the perturbation field corresponding to this density distribution, it is then necessary to multiply it by a suitable Green function to obtain the potential which can be then differentiated. The following application of the Zel'dovich approximation [26] enables us to find the initial positions and velocities for a given particle distribution. A more detailed description of the algorithms to create cosmological initial conditions can be found in [8]. Notice that the main limitations of this method are essentially due to the use of a finite computational volume: the wavelengths close to the box size are badly sampled while those larger are not present at all!

Two further complications should be mentioned. Using a perfectly regular grid to distribute the unperturbed particles can introduce on the resulting density power spectrum discreteness effects which can be reduced by starting from an amorphous fully relaxed particle distribution. As suggested by [25], this can be constructed by applying negative gravity to a system and evolving it for a long time until it reaches a relaxed state. Second, when interested to studies of individual objects like galaxy clusters, we have to recall that also large-scale tidal forces have an important role in determining their final properties. To include these effects one can apply the so-called "zoom" technique [24]: a high-resolution region is self-consistently embedded in a larger scale cosmological volume at low resolution. This allows an increase of the dynamical range up to two orders of magnitude while keeping the full cosmological context. For galaxy simulations it is even possible to apply this technique on several levels of refinements to further improve the dynamical range of the simulation.

As important as the initial conditions is the problem of the optimal setup for cosmological simulations. For example, the number of particles necessary to reach the numerical convergence in the description of a given region of interest depends on the astrophysical quantity of interest, ranging from $\approx 30-50$ for the halo mass function to some thousands for the density profile in the central regions. Recently,

[19] presented a comprehensive series of convergence tests designed to study the effect of numerical parameters on the structure of simulated dark matter haloes. In particular the paper discusses the influence of the gravitational softening, the time stepping algorithm, the starting redshift, the accuracy of force computations, and the number of particles in the spherically-averaged mass profile of a galaxy-sized halo. The results were summarized in empirical rules for the choice of these parameters and in simple prescriptions to assess the effective convergence of the mass profile of a simulated halo, when computational limitations are present.

In general, it is important to notice that both the size and the dynamical range or resolution of the N-body simulations have been increasing very rapidly over the last decades, in a way that, thanks to improvements in the algorithms, is faster than the underlying growth of the available CPU power.

2.7 Code comparison

Thanks the analysis of very recent data regarding the cosmic microwave background, the galaxy surveys and the high-redshift supernovae, we entered the era of the so-called high-precision cosmology, with very stringent constraints (i.e. at the per cent accuracy level) on the main parameters. In the perspective of even better data as expected in upcoming projects, the theoretical predictions, in order to be a useful and complementary tool, must reach a similar level of precision, that, particularly in the highly non-linear regime, represents a real challenge.

A first step in this direction is the launch of an extensive program of comparison between the different numerical codes available in the community. An example is the work presented in [12], where ten different codes adopting a variety of schemes (tree, APM, tree-PM, etc.) have been tested against the same initial conditions. In general the comparison has been very satisfactory, even if the variance between the results is still too large to make predictions suitable for the next-generation of galaxy surveys. In particular the accuracy in the determination of the halo mass function is always better than 5 per cent, while the agreement for the power spectrum in the nonlinear regime is at the $5-10$ per cent level, even on moderate spatial scales around $k = 10h\,\mathrm{Mpc}^{-1}$. Considering the internal structure of halos in the outer regions of $\sim R_{200}$, it also appears to be very similar between different simulation codes. Larger differences between the codes in the inner region of the halos occur only if the halo is not in a relaxed state.

3 Hydrodynamical codes

3.1 The model equations

Even if the dark matter component represents the dominant contribution to the total matter distribution in the universe, it is quite important to have also a fair representation of baryons: in fact most of the astrophysical signals are originated by physical processed related to them.

In general the description of the baryonic component is based on the assumption of an ideal fluid, for which the time evolution is obtained by solving the following set of hydrodynamical equations: the Euler equation,

$$\frac{d\mathbf{v}}{dt} = -\frac{\nabla P}{\rho} - \nabla\Phi; \tag{27}$$

the continuity equation,

$$\frac{d\rho}{dt} + \rho\nabla\mathbf{v} = 0; \tag{28}$$

the first law of thermodynamics,

$$\frac{du}{dt} = -\frac{P}{\rho}\nabla\cdot\mathbf{v}. \tag{29}$$

In the previous equations, P and u represent the pressure and the internal energy (per unit mass), that under the assumption of an ideal monatomic gas, are related by the equation of state: $P = 2/3\rho u$. Notice that for the moment we neglect the effect produced by radiative losses, usually encrypted by the cooling function $\Lambda(u,\rho)$ (see the discussion in Section 3.5).

In the recent years, different techniques have been proposed to solve the coupled equations of dark and baryonic matter. All of them treat the gravitational term related to $\nabla\Phi$ with the same kinds of technique described in the previous section. For the hydrodynamical part they adopt different strategies which can be classified in two big categories: lagrangian and eulerian techiques. The former are methods based on a finite number of particles used to describe the mass distribution, while the latter are methods that adopt a mesh to discretize the simulation volume. Before reviewing in the following subsections the main characteristics of their most important prototypes, we notice here that considering self-gravity, as necessary in cosmological applications, introduces a much higher level of complexity, with respect to standard hydrodynamical simulations. In fact the formation of very evolved density structures induces motions which are often extremely supersonic, whith the presence of shocks and discontinuities. Moreover, in order to have reliable simulations, the process of structure formation must be accurately followed on a very large range of scales, covering many order of magnitudes in space, in time and in the interesting physical quantities (temperature, pressure, energy, etc.). Finally, we need to include the expansion of the universe, which modifies the previous set of

equations as:

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{1}{a}(\mathbf{v}\cdot\nabla)\mathbf{v} + \frac{\dot{a}}{a}\mathbf{v} = -\frac{1}{a\rho}\nabla P - \frac{1}{a}\nabla\Phi, \tag{30}$$

$$\frac{\partial \rho}{\partial t} + \frac{3\dot{a}}{a}\rho + \frac{1}{a}\nabla\cdot(\rho\mathbf{v}) = 0 \tag{31}$$

and

$$\frac{\partial}{\partial t}(\rho u) + \frac{1}{a}\mathbf{v}\cdot\nabla(\rho u) = -(\rho u + P)\left(\frac{1}{a}\nabla\cdot\mathbf{v} + 3\frac{\dot{a}}{a}\right) . \tag{32}$$

3.2 Smoothed Particle Hydrodynamics (SPH)

The most popular lagrangian method is certainly SPH, that has a very good spatial resolution in high-density regions, while its performance in underdense region is unsatisfactory. Another charcateristic of SPH is the fact that it adopts an artificial viscosity which does not allow to reach a sufficiently high resolution in shocked region. Even if in general it is not able to treat dynamical instabilities with the same accuracy of the eulerian methods presented in the next subsection, thanks to its adaptive nature, SPH is still the preferred method for cosmological hydrodynamical simulations.

Here we introduce the basic idea of the technique. We refer to [16] for a more extended presentation of the method. Unlike in the eulerian algorithms, in SPH the fluid is represented using mass elements, i.e. a finite set of particles. This choice originates the different performance in high- and low-density regions: in fact where the mean interparticle distance is smaller, the fluid is less sparsely sampled, with a consequent better resolution; the opposite holds when the interparticle distance is large.

3.2.1 Basics of SPH

As said, in SPH the fluid is discretized using mass elements (i.e. a finite set of particles), unlike in eulerian codes where the discretization is made using volume elements. Thanks to an adaptive smoothing, the mass resolution is kept fixed. In more detail, a generic fluid quantity A is defined as

$$\langle A(\mathbf{x})\rangle = \int W(\mathbf{x}-\mathbf{x}', h)A(\mathbf{x}')d\mathbf{x}', \tag{33}$$

where the smoothing kernel W is suitably normalised: $\int W(\mathbf{x}, h)d\mathbf{x} = 1$. Moreover W, assumed to depend only on the distance modulus, collapses to a delta function if the smoothing length h tends to zero.

Using a set of particles with mass m_j and position $\mathbf{x}_j$, the previous integral can be done in a discrete way, replacing the volume element of the integration with the

ratio of the mass and density m_j/ρ_j of the particles:

$$\langle A_i \rangle = \sum_j \frac{m_j}{\rho_j} A_j W(\mathbf{x}_i - \mathbf{x}_j, h)\,. \tag{34}$$

If one adopts kernels with a compact support (i.e. $W(\mathbf{x},h) = 0$ for $|\mathbf{x}| > h$), it is not necessary to do the summation over the whole set of particles, but only over the neighbours around the i-th particle under consideration, namely the ones inside a sphere of radius h. The following kernel represents an optimal choice in many cases and it has been used often in the literature:

$$W(x,h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6\left(\frac{x}{h}\right)^2 + 6\left(\frac{x}{h}\right)^3 & 0 \le x/h < 0.5 \\ 2\left(1 - \frac{x}{h}\right)^3 & 0.5 \le x/h < 1 \\ 0 & 1 \le x/h \end{cases} . \tag{35}$$

If the quantity A is the density ρ_i, Eq. 34 becomes simpler:

$$\langle \rho_i \rangle = \sum_j m_j W(\mathbf{x}_i - \mathbf{x}_j, h). \tag{36}$$

The big advantage of SPH is that the derivatives can be easily calculated as

$$\nabla \langle A_i \rangle = \sum_j \frac{m_j}{\rho_j} A_j \nabla_i W(\mathbf{x}_i - \mathbf{x}_j, h), \tag{37}$$

where ∇_i denotes the derivative with respect to $\mathbf{x}_i$. A pairwise symmetric formulation of derivatives can be obtained by exploiting the identity

$$(\rho \nabla) \cdot A = \nabla(\rho \cdot A) - \rho \cdot (\nabla A), \tag{38}$$

which allows one to re-write a derivative as

$$\nabla \langle A_i \rangle = \frac{1}{\rho_i} \sum_j m_j (A_j - A_i) \nabla_i W(\mathbf{x}_i - \mathbf{x}_j, h). \tag{39}$$

Another symmetric representation of the derivative can be alternatively obtained from the identity

$$\frac{\nabla A}{\rho} = \nabla \left(\frac{A}{\rho} \right) + \frac{A}{\rho^2} \nabla \rho, \tag{40}$$

which then leads to:

$$\nabla \langle A_i \rangle = \rho_i \sum_j m_j \left(\frac{A_j}{\rho_j^2} + \frac{A_i}{\rho_i^2} \right) \nabla_i W(\mathbf{x}_i - \mathbf{x}_j, h). \tag{41}$$

By applying the previous identities, one can re-write the Euler equation as

$$\frac{d\mathbf{v}_i}{dt} = -\sum_j m_j \left(\frac{P_j}{\rho_j^2} + \frac{P_i}{\rho_i^2} + \Pi_{ij} \right) \nabla_i W(\mathbf{x}_i - \mathbf{x}_j, h), \tag{42}$$

while the term $-(P/\rho)\nabla \cdot \mathbf{v}$ in the first law of thermodynamics can be written as

$$\frac{du_i}{dt} = \frac{1}{2} \sum_j m_j \left(\frac{P_j}{\rho_j^2} + \frac{P_i}{\rho_i^2} + \Pi_{ij} \right) (\mathbf{v}_j - \mathbf{v}_i) \nabla_i W(\mathbf{x}_i - \mathbf{x}_j, h). \tag{43}$$

In order to have a good reproduction of shocks, in the previous equations it has been necessary to introduce the so-called artificial viscosity Π_{ij}, for which different forms have been proposed in the literature (see, for example, [17]): most of them try to reduce its effects in the regions where there are no shocks, adopting a specific articial viscosity for each particle.

To complete the set of fluid equations, we notice that the continuity equation does not represent a problem for Lagrangian methods like SPH, being automatically satisfied thanks to the conservation of the number of particles.

3.2.2 The smoothing length

Usually, each individual particle i has its own smoothing length h, which is determined by finding the radius h_i of a sphere which contains N_{nei} neighbours. A large value for N_{nei} would allow better estimates for the density field but with larger systematics; viceversa small N_{nei} would lead to larger sample variances. In the literature standard choices for N_{nei} range between 32 and 100. The presence of a variable smoothing length in the hydrodynamical equations can break their conservative form: to avoid it, it is necessary to introduce a symmetric kernel $W(\mathbf{x}_i - \mathbf{x}_j, h_i, h_j) = \bar{W}_{ij}$, for which the two main variants used in the literature are the kernel average,

$$\bar{W}_{ij} = (W(\mathbf{x}_i - \mathbf{x}_j, h_i) + W(\mathbf{x}_i - \mathbf{x}_j, h_j))/2 \ , \tag{44}$$

and the average of the smoothing lengths

$$\bar{W}_{ij} = W(\mathbf{x}_i - \mathbf{x}_j, (h_i + h_j)/2) \ . \tag{45}$$

Note that when writing the derivatives, we assumed that h is independent of the position $\mathbf{x}_j$. Thus, if the smoothing length h_i is variable for each particle, one would formally introduce the correction term $\partial W/\partial h$ in all the derivatives. An elegant way to do that is using the formulation which conserves numerically both entropy and internal energy, described in the next subsection.

3.2.3 Conserving the entropy

In adiabatic flows, the entropic function $A = P/\rho^{\gamma}$ is conserved. The quantity A is related to the internal energy per unit mass:

$$u_i = \frac{A_i}{\gamma - 1}\rho_i^{\gamma-1} \ . \tag{46}$$

Shocks, which can be captured in SPH thanks to the artificial viscosity Π_{ij}, can originate an evolution of A:

$$\frac{dA_i}{dt} = \frac{1}{2}\frac{\gamma-1}{\rho_i^{\gamma-1}}\sum_j m_j \Pi_{ij}\,(\mathbf{v}_j - \mathbf{v}_i)\,\nabla_i \bar{W}_{ij}. \tag{47}$$

The Euler equation can be derived starting by defining the Lagrangian of the fluid as

$$L(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2}\sum_i m_i \dot{\mathbf{x}}_i^2 - \frac{1}{\gamma-1}\sum_i m_i A_i \rho_i^{\gamma-1} \tag{48}$$

which represents the entire fluid and has the coordinates $\mathbf{q} = (\mathbf{x}_1, ..., \mathbf{x}_N, h_1, ..., h_N)$.

The next important step is to define constraints, which allow an unambiguous association of h_i for a chosen number of neighbours. This can be done by requiring that the kernel volume contains a constant mass for the estimated density,

$$\phi_i(\mathbf{q}) = \frac{4\pi}{3} h_i^3 \rho_i - n m_i = 0. \tag{49}$$

The equation of motion can be obtained as the solution of

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{q}}_i} - \frac{\partial L}{\partial \mathbf{q}_i} = \sum_j \lambda_j \frac{\partial \phi_j}{\partial \mathbf{q}_i}, \tag{50}$$

which - as demonstrated by [21] - can be written as

$$\frac{d\mathbf{v}_i}{dt} = -\sum_j m_j \left(f_j \frac{P_j}{\rho_j^2}\nabla_i W(\mathbf{x}_i - \mathbf{x}_j, h_j) + f_i \frac{P_i}{\rho_i^2}\nabla_i W(\mathbf{x}_i - \mathbf{x}_j, h_i) + \Pi_{ij}\nabla_i \bar{W}_{ij} \right). \tag{51}$$

In the previous equation we notice the additional term due to the artificial viscosity Π_{ij}, which is needed to capture shocks.

The coefficients f_i, defined as

$$f_i = \left(1 + \frac{h_i}{3\rho_i}\frac{\partial \rho_i}{\partial h_i}\right)^{-1} \ , \tag{52}$$

incorporate fully the variable smoothing length correction term. For a detailed derivation of this formalism and its conserving capabilities see [21].

3.3 Eulerian methods

These methods solve the hydrodynamical equations adopting a grid, which can be fixed or adaptive in time. The first attempts were based on the so-called central difference schemes, i.e. schemes where the relevant hydrodynamical quantities were only represented by their values at the center of the grid, and the various derivatives were obtained by the finite-difference representation. This approach was not satisfactory: in fact the schemes were only first-order accurate and they made use of some artificial viscosity to treat discontinuities, similarly to SPH. In more modern codes (like for example the piecewise parabolic method, PPM) the shape of the hydrodynamical quantities $f_{n,u}(x)$ is recovered with much higher accuracy thanks to the use of several neighbouring cells. This allows to calculate the total integral of the given quantity over the grid cell, divided by the volume of each cell (e.g. cell average, $\hat{u}_n$), rather than pointwise approximations at the grid centres (e.g. central variables, u_n):

$$\hat{u}_n = \int_{x_{n-0.5}}^{x_{n+0.5}} f_{n,u}(x)dx \; . \tag{53}$$

This global shape is also used to estimate the values at the cell boundaries (e.g. $u^l_{n\pm0.5}$,$u^r_{n\pm0.5}$, the so-called interfaces), which can be used later as initial conditions of a Riemann problem, e.g. the evolution of two constant states separated by a discontinuity. Once the Riemann is analytically solved (see, e.g., [7]), it is possible to compute the fluxes across these boundaries for the time interval and then to update the cell averages $\hat{u}_n$. Notice that for high-dimensional problems, this procedure has to be repeated for each coordinate direction separately.

In general, the eulerian methods suffer from limited spatial resolution, but they work extremely well in both low- and high-density regions, as well as in treating shocks. In cosmological simulations, accretion flows with large Mach numbers are very common. Here, following the total energy in the hydrodynamical equations can produce inaccurate thermal energy, leading to negative pressure, due to discretisation errors when the kinetic energy dominates the total energy. In such cases, the numerical schemes usually switch from formulations solving the total energy to formulations based on solving the internal energy in these hypersonic flow regions.

3.4 Code Comparison

The hydrodynamical schemes presented in the previous sections are based on quite different approaches, having for construction very different characteristics (in practice particles against grid). For this reason, in order to be validated, they must be tested against problems with known solutions and then one has to compare their results, once the same initial conditions are given. The first standard test is the reproduction of the shock tube problem: in this case codes able to capture the presence

of shocks, like the most sophisticated eulerian methods, have certianly better performances. Another possible test is the evolution of an initially spherical perturbation.

In general, a code validation in a more cosmological context is difficult. A first attempt of comparison between grid-based and SPH-based codes can be found in [14]. More recently [18] compared the thermodynamical properties of the intergalactic medium predicted by the lagrangian code *GADGET* and by the eulerian code *ENZO*.

In the case of simulations for single cosmic structures, the most complete code comparison is the one provided by the Santa Barbara Cluster Comparison Project [10], where 12 different groups ran the own codes (including both eulerian and lagrangian schemes), starting from the same identical initial conditions corresponding to a massive galaxy cluster. In general the agreement for the dark matter properties was satisfactory, with a 20 per cent scatter around the mean density and velocity dispersion profiles. A similar agreement was also obtained for many of the gas properties, like the temperature profile or the ratio of the specific dark matter kinetic energy to the gas thermal energy; somewhat larger differences are found in the temperature or entropy profiles in the innermost regions. One of the most worrying discrepancies is certainly when the total X-ray luminosity is considered: in this case the spread can be also a factor 10. Most of the problem originates from a too low resolution in the central core, where the gas density (which enters as squared in the estimate of the X-ray luminosity) reaches its maximum. Another large difference between the results of different hydro-codes is related to the predicted baryon fraction and its profile within the cluster.

In general we can conclude that lagrangian and eulerian schemes are providing compatible results, with some spread due to their specific weaknesses. However, we have to remind that these comparisons have been done in the non-radiative regime, i.e. excluding a long list of complex physical processes acting on baryons (see the following section), processes whose implementation is mandatory for a complete description of the formation of cosmic structures.

3.5 Extra gas physics

We know very well that the formation and evolution of cosmic structures is not determined only by the gravity due to the total matter distribution and by the adiabatic behaviour of gas: many other important processes are in action, in particular on the baryonic component, influencing the final (physical and observational) properties of the structures. Here we briefly list these processes, discussing their importance and pointing out the related numerical issues.

- *Cooling*. Its standard implementation is made by adding the cooling function $\Lambda(u,\rho)$ in the first law of thermodynamics, under the assumptions that the gas is optically thin and in ionisation equilibrium, and that only two-body cooling processes are important. Considering a plasma with primordial composition of H and He, these processes are collisional excitation of H I and He II, collisional

ionisation of H I, He I and He II, standard recombination of H II, He II and He III, dielectric recombination of He II, and free-free emission (Bremsstrahlung). Being the collisional ionisation and recombination rates depending only on the temperature, when the presence of a ionising background radiation is excluded it is possible to solve the resulting rate equation analytically; alternatively the solution is obtained iteratively (see a discussion in [15]). When also the metallicity is implemented in the code, the number of possible processes becomes so large that it is necessary to use pre-computed tabulated cooling function. Finally we notice that for pratical reasons the gas cooling is followed as "sub time step" problem, decoupled from the hydrodynamical treatment.

- *Star formation.* The inclusion of the cooling process originates two numerical problems. First, since the cooling is a runaway process, in the central regions of clusters the typical cooling time becomes significantly shorter than the Hubble time, causing the so-called overcooling problem: the majority of baryons can cool down and condense out of the hot phase. Second, since cooling is proportional to the square of the gas density, its efficiency is strongly related to the number of the first collapsed structures, whose good representation in simulations depends on the assumed numerical resolution. To avoid these problems, it is necessary to include the process of star formation starting from the cold and dense phase of the gas. These stellar objects, once the phase of supernova explosion is reached, can inject a large amount of energy (the so-called feedback) in the gas, increasing its temperature and possibly counteracting the cooling catastrophe. From a numerical point of view, this process is implemented via simple recipes (see, for example, [15]), based on the assumptions that the gas has a convergent flow, it is Jeans unstable and, most important, it has an overdensity larger than the threshold corresponding to collapsed regions. This procedure allows to compute the fraction of gas converted into stars. For computational and numerical reasons, the star formation is done only when a significant fraction of the gas particle mass is interested by the transformation: at this point, a new collisionless "star" particle is created from the parent star-forming gas particle, whose mass is reduced accordingly. This process takes place until the gas particle is entirely transformed into stars. In order to avoid spurious numerical effects, which arise from the gravitational interaction of particles with widely differing masses, one usually restricts the number of star particles spawned by a gas particle to be relatively small, typically $2-3$.
- *Supernova feedback.* Assuming a specific initial mass function, it is possible to estimate the number of stars ending their lifes as type-II supernovae and then to compute the total amount of energy (typically 10^{51} erg per supernova) that each star particle can release to the surrounding gas. Assuming that the typical lifetime of massive stars is smaller than the typical simulation time step, the feedback energy is injected in the surrounding gas in the same step (instantaneous recycling approximation). Improvements with respect to this simple model include an explicit sub-resolution description of the multi-phase nature of the interstellar medium, which provides the reservoir of star formation. Such a sub grid model tries to model the global dynamical behaviour of the interstellar medium in which

cold, star-forming clouds are embedded in a hot medium (see for example the self-regulated model proposed by [22] and a general critical discussion in [4]). .

- *Chemical enrichment.* A possible way to improve the previous models is the inclusion of a more detailed description of stellar evolution, and of the corresponding chemical enrichment. In particular it is possible to follow the release of metals also from type-Ia supernovae and low and intermediate mass stars, avoiding the instantaneous recycling approximation (see the discussion in [5]).
- *Other processes.* The baryons present in the cosmic structures undergo other physical processes that can have an important role in their modelization: a non exaustive list includes the effects related to thermal conduction, radiative transfer, magnetic fields, relativistic particles, black holes and extra sources of feedback (like, for instance, AGN). The first attempts of implementing the corresponding physics in cosmological simulations have been done, even if in some cases the results are not yet convergent. Having a robust implementation of all these phenomena represents the more difficult and changelling frontier for future numerical experiments.

References

1. Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15
2. Barnes, J., & Hut, P. 1986, Nature, 324, 446
3. Binney, J., & Tremaine, S. 1987, Princeton, NJ, Princeton University Press, 1987
4. Borgani, S., et al. 2006, MNRAS, 367, 1641
5. Borgani, S., Fabjan, D., Tornatore, L., Schindler, S., Dolag, K., & Diaferio, A. 2008, Space Science Reviews, 134, 379
6. Couchman, H. M. P. 1991, ApJ, 368, L23
7. Courant, R., & Friedrichs, K. O. 1948, Pure and Applied Mathematics, New York: Interscience
8. Efstathiou, G., Davis, M., White, S. D. M., & Frenk, C. S. 1985, ApJS, 57, 241
9. Evrard, A. E., et al. 2002, ApJ, 573, 7
10. Frenk, C. S., et al. 1999, ApJ, 525, 554
11. Grossi, M., Dolag, K., Branchini, E., Matarrese, S., & Moscardini, L. 2007, MNRAS, 382, 1261
12. Heitmann, K., et al. 2008, Computational Science and Discovery, 1, 015003
13. Ito, T., Makino, J., Fukushige, T., Ebisuzaki, T., Okumura, S. K., & Sugimoto, D. 1993, PASJ, 45, 339
14. Kang, H., Ostriker, J. P., Cen, R., Ryu, D., Hernquist, L., Evrard, A. E., Bryan, G. L., & Norman, M. L. 1994, ApJ, 430, 83
15. Katz, N., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19
16. Monaghan, J. J. 1992, ARA&A, 30, 543
17. Monaghan, J. J. 1997, Journal of Computational Physics, 136, 298
18. O'Shea, B. W., Nagamine, K., Springel, V., Hernquist, L., & Norman, M. L. 2005, ApJS, 160, 1
19. Power, C., Navarro, J. F., Jenkins, A., Frenk, C. S., White, S. D. M., Springel, V., Stadel, J., & Quinn, T. 2003, MNRAS, 338, 14
20. Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Cambridge: University Press
21. Springel, V., & Hernquist, L. 2002, MNRAS, 333, 649

22. Springel, V., & Hernquist, L. 2003, MNRAS, 339, 289
23. Springel, V. 2005, MNRAS, 364, 1105
24. Tormen, G., Bouchet, F. R., & White, S. D. M. 1997, MNRAS, 286, 865
25. White, S. D. M. 1996, Cosmology and Large Scale Structure, 349
26. Zel'Dovich, Y. B. 1970, A&A, 5, 84