

Contents

1	Introduction	1
1.1	Extragalactic sky	1
1.2	Cosmological model	1
1.3	Instruments	3
1.4	Basic concepts	3
2	Cosmological Background	5
2.1	Cosmological Principle	5
2.2	Elements of General Relativity	5
2.3	FRW metric	6
2.4	Friedmann equation	6
2.5	Solutions	7
2.6	Distances and times	8
2.7	Thermal history	10
2.8	Thermodynamics	10
2.9	Particle relics	14
2.10	Primordial nucleosynthesis	15
2.11	Recombination	16
2.12	Λ CDM model	17
2.13	Problems with Λ CDM	18
2.14	Inflation	20
3	Linear Perturbations	23
3.1	Ideal Fluid	23
3.2	Collisionless gas	27
3.3	Solutions	27
3.4	Relativistic perturbations	29
3.5	Transfer function	31
3.6	Primordial perturbations	32
3.7	Cosmological field statistics	33
3.8	Matter power spectrum	35
3.9	Cosmic Microwave Background anisotropies	36
3.10	Galaxy Clustering	39
4	Nonlinear Perturbations	43
4.1	Approaches for non-linear structure formation	43
4.2	Spherical Collapse	43
4.3	Collisionless dynamics	46
4.4	Virial theorem	48
4.5	Steady state solutions	50
4.6	Collisionless relaxation	52
4.7	Halo Mass Function	53
4.8	Halo mergers	56
4.9	Halo clustering	56
4.10	Internal structure of Haloes	57
5	Baryonic Structures	61

5.1	State and distribution of baryons	61
5.2	Stars	62

Source: <https://github.com/thomabir/astrophysics-ii>

1 Introduction

1.1 Extragalactic sky

When looking in a clear night into the sky you can see a whitish band where a lot of stars are located (see figure 1.1). This band is the milky way the galaxy, to which our solar system belongs. If we were able to have a look onto our galaxy from the outside, we would see something like in figure 1.2. We see that the milky way has a galactic center which has the form of a bar, and spiral arms which come from the center. Our solar system is situated quite far away from the galactic center in one of the arms.

But our galaxy is not the only one in our Universe, and there are different types of galaxies. The milky way is a spiral galaxy. In figure 1.3a an other spiral galaxy the M101 is illustrated. A subgroup of spiral galaxies are barred spiral galaxies (see figure 1.3b), which are characterized by the barred shape of their center. Figure 1.3c shows an elliptical galaxy and figure 1.3d is an example of an irregular galaxy, the one shown in the figure is an irregular dwarf galaxy. If a spiral or an irregular galaxy has a really bright galactic center, it is called a Seyfert galaxy (see figure 1.3e). They are a subgroup of active galactic nuclei and the brightness of the galactic center is probably caused by a super massive black hole in the center of the galaxy. Galaxies can gather in groups (see figure 1.4a) or clusters (see figure 1.4b). Galaxy groups contain up to 50 galaxies and, they are the smallest collection of galaxies. Galaxy clusters consist of hundreds to thousands of galaxies. They are the largest gravitationally bound structures in the universe and their mass is around 10^{14} to 10^{15} solar masses.



Figure 1.1: An image of the night sky which shows the milky way.

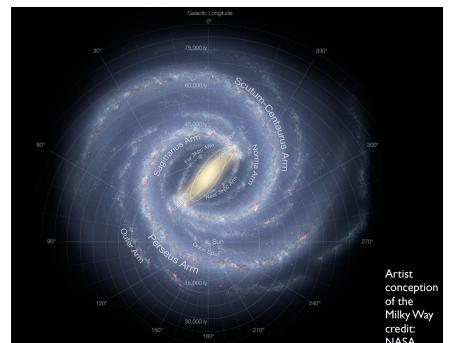


Figure 1.2: An illustration of the milky way, showing it from an outside perspective.

1.2 Cosmological model

The standard cosmological model assumes that the universe was initially created in a Big Bang and underwent different epochs until it became the way it is now. Figure 1.5 shows the history of the universe, when assuming this model.

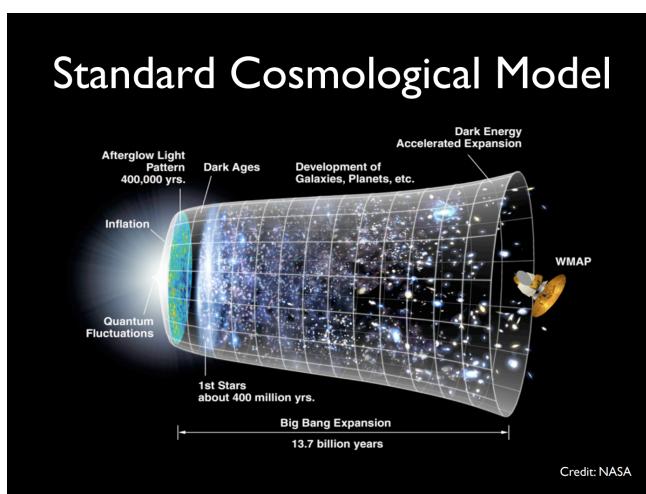
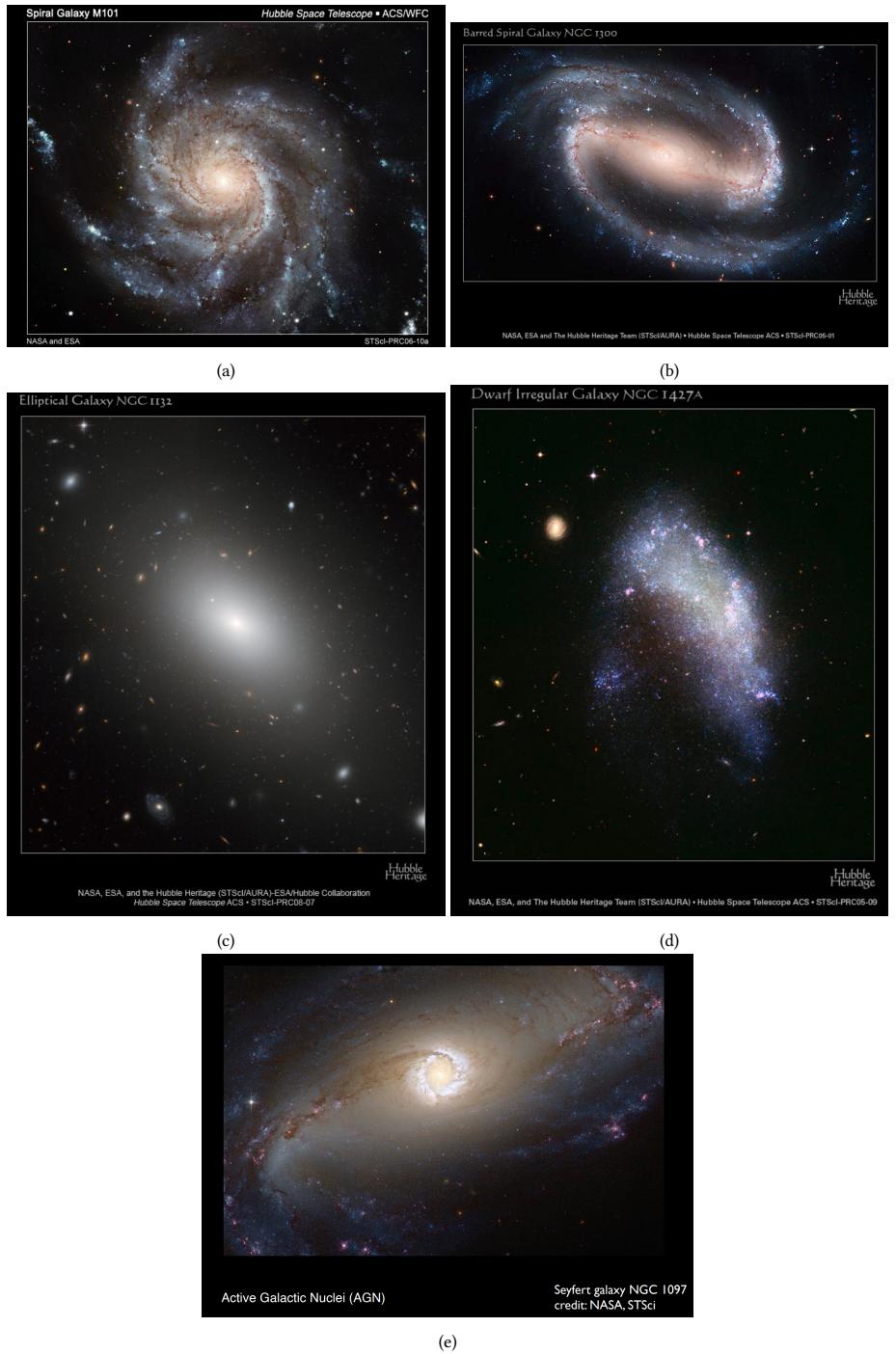


Figure 1.5: The standard cosmological model describes the evolution of our Universe starting with the Big Bang.



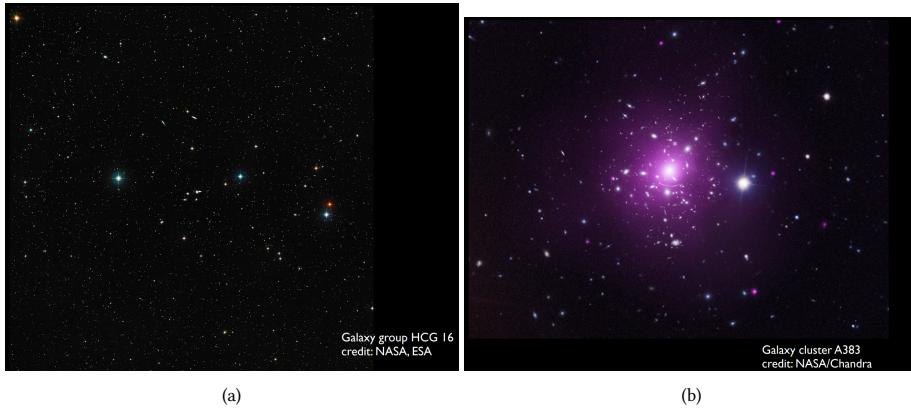


Figure 1.4: An example of a galaxy group (a) and of a galaxy cluster (b).

From this model we also find that the universe nowadays consist of about 4% ordinary matter, 20% dark matter and 76% dark energy. We will look at this closer in chapter 2.

1.3 Instruments

For experimental results, large telescopes are needed. In order for them to be useful, they have to be placed in places with the right conditions. Therefore, a lot of the telescopes are in space, especially if one wants to observe in the ultraviolet. Figure 1.6 shows the electromagnetic spectrum and at which wavelengths it is necessary to do the observations in space. Radio waves can be observed from earth, for those observation we use array telescopes.

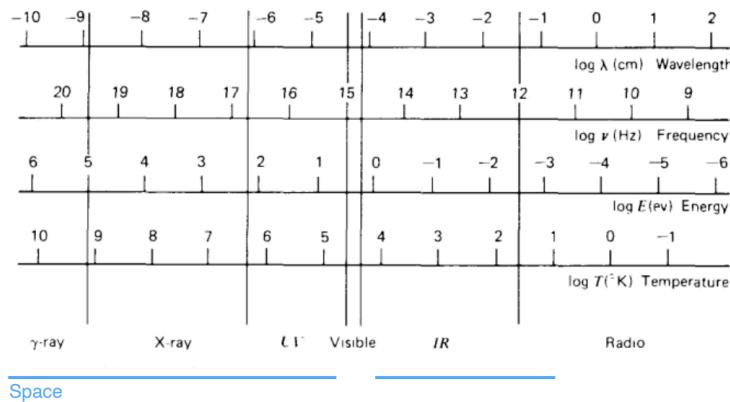


Figure 1.6: The figure shows at which wavelengths it is necessary to put measurement devices in space in order to do observations.

Rybicki & Lightman, 1979

1.4 Basic concepts

Astrophysical units are uniquely used in Astrophysics. In order to talk of the mass of different stars, galaxies and more we use the unit of solar masses, where $1M_{\odot} = 1.99 \cdot 10^{30}$ kg corresponds to the mass of the sun. When talking about time usually years instead of seconds are used, where $1 \text{ yr} = 3.16 \cdot 10^7$. Finally we come to the unit of distance. The most known distance scale used in Astrophysics is probably

Table 1.1: Astrophysical scales.

Mass scales	Distance scales	Time scales
Dwarf galaxies $\approx 10^9 M_\odot$	Galaxy size $\approx 10^4$ pc	Universe age $\approx 14 \cdot 10^9$ yr
Galaxies $\approx 10^{12} M_\odot$	Galaxy cluster size $\approx 10^6$ pc	Sun age $\approx 4.6 \cdot 10^9$ yr
Galaxy groups $\approx 10^{13} M_\odot$	Homogeneity $\approx 10^8$ pc	
Galaxy clusters $\approx 10^{15} M_\odot$	Observable Universe $\approx 10^9$ pc	

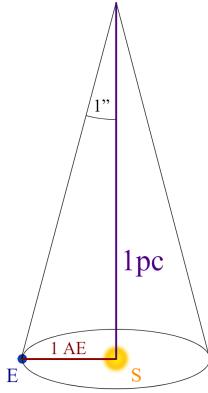


Figure 1.7: One parsec corresponds to the needed distance between sun and object which results into an angle of $1''$, if one looks at the triangle created by the sun, the object and the earth, which rotates around the sun.

light-years, which is the distance light travels in one year. An other unit for distances which we will use mostly is parsec, $1 \text{ pc} = 3.09 \cdot 10^{16} \text{ m} = 3.26 \text{ ly}$. Figure 1.7 shows the meaning of a parsec. It is the correlation between the distance between earth and sun and the respective object (a star in the figure) to which we want to know the distance. 1 parsec corresponds to the angle of $1''$ at the object. In order to get a feeling for scales, some mass, distance and time scales are listed in table 1.1

To talk about stars and other bright objects in the universe we need a way to talk about the luminosity, flux and the magnitude of these objects. The intrinsic luminosity is defined as $L = \text{energy/time}$, whereas the flux at the observer is defined as the luminosity divided by the area, $F = L/4\pi r^2$, where r is the distance between object and observer.

Usually, the flux is with respect to a specific waveband X , therefore we write f_X . An other possibility to say something about the brightness of the object is the magnitude, which is the most common way. There is the apparent magnitude m_X which describes the brightness of the object seen by the observer (which means here on earth). In order to compare the brightness of different objects and to be able to say something about their properties such as mass, one needs the absolute magnitude M_X . The absolute magnitude is the magnitude one would see at a distance of 10 pc to the object. The magnitudes are calculated as follows:

$$m_X = -2.5 \log(f_X/f_{X,0}), \\ M_X = -2.5 \log(L_X) + C_X,$$

where $f_{X,0}$ is the flux zero point and C_X is the zero point. For the zero point, the star Vega is used, which has an apparent magnitude of 0 mag.

The relation between the apparent and the absolute magnitude can be used to find the distance r to the object,

$$m_X - M_X = 5 \log(r/r_0). \quad (1.1)$$

2 Cosmological Background

2.1 Cosmological Principle

The Cosmological principle states that the universe is homogeneous and isotropic on sufficiently large scales (larger than 100 Mpc). This is a generalization of the Copernican principle, according to which there is no special place and no special direction in the universe.

In the following the evolution of the universe will be described, assuming that it is isotropic and homogeneous. Later local perturbations about this uniform background will be taken into account.

2.2 Elements of General Relativity

The theoretical basis of cosmology is General Relativity (GR). In contrary to Newton's laws, where a body is attracted by another one. In GR gravity is seen as a manifestation of the curvature of space-time and this curvature leads among others to the observation that bodies attract each other. The curvature of space-time is described by a metric $g_{\mu\nu}$ which describes the motion of distance on a curved space.

As an illustration we consider a two-dimensional space with a coordinate system x^μ , where $\mu = 1, 2$. The physical square distance is then $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ where dx^μ and dx^ν are the coordinate distances. Assuming that the space is homogeneous and isotropic the distance can be written as

$$ds^2 = a^2 \left(\frac{dr^2}{1 - kr^2} + r^2 d\phi^2 \right)$$

where k describes the curvature and a is the length scale known as the scale factor which can also be time-dependent. If the scale factor is time dependent this means that the surface is uniformly expanding or contracting. The curvature k for a homogeneous and isotropic surface is either $k = 0$ for a plane, $k = 1$ for a sphere or the curvature is negative $k = -1$ for a hyperbolic surface. In order to describe the sphere the coordinate transformation $r = \sin\chi$ is used and to describe the hyperbolic surface $r = \sinh\chi$.

The discussion of the two-dimensional space can be adapted to our universe, which is in four dimensional space-time described by the Minkowsky metric $g_{\mu\nu} = \text{diag}(1, -1, -1, -1)$. In this space time the physical square distance for a flat universe would be given by $ds^2 = c^2dt^2 - (dx^2 + dy^2 + dz^2)$. In general the dynamics of space-time is described by the Einstein field equations:

$$G_{\mu\nu} = \frac{8\pi G}{c^2} T_{\mu\nu}.$$

$G_{\mu\nu}$ is the Einstein tensor, which includes the curvature of the space-time. $T_{\mu\nu}$ is the stress energy tensor and describes the content of the space-time. For an ideal fluid the stress energy tensor is

$$T_{\mu\nu} = \text{diag}(\rho c^2, p, p, p),$$

where ρc^2 is the energy density and p the pressure.

In the space-time the trajectories of particles are in geodesics which means they take the straightest paths possible. As an example for photons (massless particles) $ds^2 = 0$.

2.3 FRW metric

The Friedmann-Robertson-Walker (FRW) metric is the metric of a homogeneous and isotropic universe:

$$ds^2 = c^2 dt^2 - a(t)^2(d\chi^2 + r(\chi)^2 d\Omega^2)$$

- χ : comoving radius
- $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$: solid angle element
- $a(t)$: scale factor
- $r(\chi) = f_K(\chi) = \begin{cases} \sin \chi & \text{closed case, positive curvature} \\ \chi & \text{flat case} \\ \sinh \chi & \text{open case, negative curvature} \end{cases}$

Hubble parameter

- Hubble parameter: $H := \dot{a}/a$
- today's value gets a subscript zero: H_0
- Because it is hard to measure H accurately, we write it as

$$H_0 = 100h \frac{\text{km}}{\text{s Mpc}},$$

where $h \approx 0.7$ is the dimensionless Hubble parameter.

- $H_0^{-1} \approx 10$ Gyr is about the age of the universe
- $cH_0^{-1} \approx 4$ Gpc is about the size of the observable universe

2.4 Friedmann equation

The Friedmann equations are derived by plugging the FRW metric into Einstein's equations:

$$\begin{aligned} H^2 &= \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2} \\ \frac{\ddot{a}}{a} &= \frac{4\pi G}{3} \left(\rho + \frac{3p}{c^2} \right) \end{aligned}$$

The critical density is defined as

$$\rho_{\text{crit}}(t) = \frac{3H(t)^2}{8\pi G}$$

Today, the critical density is about five hydrogen atoms per cubic metre, or one galaxy per Mpc³.

Density parameters:

- The subscript i describes one component of the universe ($i = \text{radiation, dark matter, matter ...}$)
- density parameter: $\Omega_i(t) = \rho_i(t)/\rho_{\text{crit}}(t)$

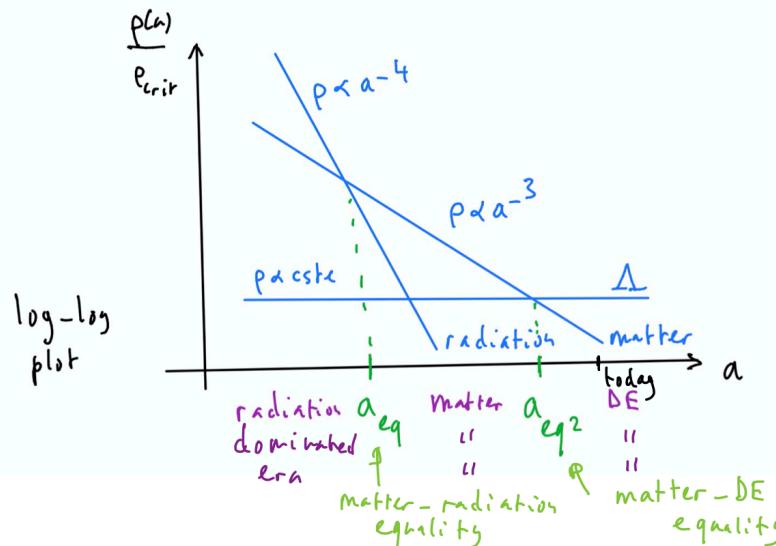


Figure 2.1: Domination of different components at different times

- total energy density: $\rho(t) = \sum_i \rho_i(t)$
- total density parameter: $\Omega(t) = \rho(t)/\rho_{\text{crit}}(t)$
- curvature density parameter: $\Omega_{K,0} = 1 - \Omega_0 = -Kc^2/H_0^2 a_0^2$

With these definitions, the (first) Friedmann equation can be rewritten as

$$\frac{H}{H_0} = \sqrt{\frac{\rho}{\rho_{\text{crit},0}} + \Omega_{K,0} \left(\frac{a_0}{a}\right)^2}$$

2.5 Solutions

To solve the Friedmann equation, $\rho(t)$ or $\rho(a)$ need to be known. It can be calculated as

$$\rho = n\epsilon$$

where n is the particle number per unit volume and ϵ the energy per particle

- Relativistic matter. ϵ is constant with a , while $n \propto a^{-3}$. Thus $\rho \propto a^{-3}$.
- Radiation. $\epsilon = h\nu = hc/\lambda \propto a^{-1}$. Thus $\rho \propto a^{-4}$.
- Vacuum energy is constant in a

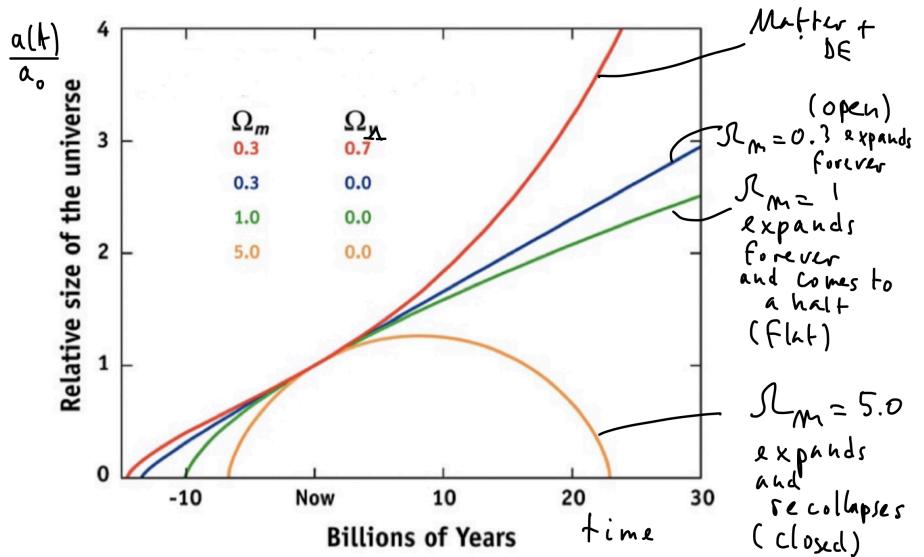
There is a generalization for general fluids:

- equation of state: $p = wpc^2$
- density: $\rho \propto a^{-3(1+w)}$
- $w = \begin{cases} 0 & \text{matter} \\ 1/3 & \text{radiation} \\ -1 & \text{vacuum energy} \end{cases}$

The results can be plugged into the Friedmann equation:

$$\begin{aligned} \frac{H}{H_0} &= \sqrt{\frac{\rho}{\rho_{\text{crit},0}} + \Omega_{K,0} \left(\frac{a_0}{a}\right)^2} \\ &= \sqrt{\Omega_{m,0} \left(\frac{a_0}{a}\right)^3 + \Omega_{r,0} \left(\frac{a_0}{a}\right)^4 + \Omega_{\Lambda,0} + \Omega_{K,0} \left(\frac{a_0}{a}\right)^2} \end{aligned}$$

This is a differential equation with $\Omega_{i,0}$ as parameters.



The standard cosmological model:

- $\Omega_{m,0} \approx 0.3$
- $\Omega_{r,0} \approx 10^{-5}$
- $\Omega_{\Lambda,0} \approx 0.7$
- $\Omega_{K,0} \approx 0$
- $\Omega_0 \approx 1$
- $h \approx 0.7$

At different times, the universe is dominated by different components. Approximations:

- Matter dominated:

$$\frac{H}{H_0} = \sqrt{\Omega_{m,0} \left(\frac{a_0}{a} \right)^3} \implies a \propto t^{2/3}$$

- Radiation dominated:

$$\frac{H}{H_0} = \sqrt{\Omega_{r,0} \left(\frac{a_0}{a} \right)^4} \implies a \propto t^{1/2}$$

- Λ dominated:

$$\frac{H}{H_0} \propto \text{constant} \implies a \propto e^{Ht}$$

- General fluid ($w \neq -1$):

$$\rho \propto a^{-3(1+w)} \implies a \propto t^{\frac{2}{3(1+w)}}$$

2.6 Distances and times

2.6.1 Angular distance & Luminosity distance

The comoving distance χ and the proper distance $a\chi$ to a source are not directly observable. However, the angular size θ and the flux F of an object can be measured directly.

Intrinsic properties of the source:

- its size D
- its luminosity L

Properties of space-time:

- the comoving distance to the source χ
- the proper distance to the source $a(t)\chi$

Measurable quantities for an observer:

- the angular size θ
- the flux F

In Euclidean space, the following relations hold:

$$\theta = \frac{D}{d} \quad F = \frac{L}{4\pi d^2}$$

where d is the distance to the source. In FRW-space, we define the following:

- the angular-diameter distance d_A satisfies $\theta = D/d_A$. One can show $d_A = ar(\chi)$
- the luminosity distance d_L satisfies $F = L/4\pi d_L^2$. One can show that $d_L = r(\chi)/a$

2.6.2 Comoving radius

We measured the redshift of a photon that has travelled to us on a radial trajectory. How far away (in comoving distance) is the source?

$$\begin{aligned} 0 &= ds^2 && \text{photon} \\ &= c dt^2 - a(t)^2 [d\chi^2 + r(\chi)^2 d\Omega^2] && \text{FRW metric} \\ \implies c dt &= a(t) d\chi && d\Omega^2 = 0 \text{ on a radial trajectory} \\ \implies d\chi &= \frac{c dt}{a(t)} \\ &= \frac{c da}{a^2 H(a)} && H = \frac{\dot{a}}{a}, \text{ so } dt = \frac{da}{a H(a)} \\ \implies \chi(a) &= c \int_a^{a_0} \frac{da'}{a'^2 H(a')} && \chi(a_0) = 0 \end{aligned}$$

$H(a)$ has to be obtained from the Friedmann equations. As a result, we will get $\chi(a, a_0)$. We can use $a/a_0 = 1/(1+z)$ to get $\chi(z, a_0)$. Since a_0 can be defined arbitrarily (for example, $a_0 = 1$), we get $\chi(z)$.

2.6.3 Comoving Horizon

Suppose a (hypothetical, non-interacting) photon was emitted at the Big Bang. How far (in comoving distance) could it have travelled until now? This comoving distance is called the comoving horizon. We plug into the previous equation, with $a = 0$ at the start:

$$\chi(a) = c \int_0^{a_0} \frac{da'}{a'^2 H(a')}$$

2.6.4 Age of the Universe

How old is the universe?

$$\begin{aligned} t_0 &= \int_0^{t_0} dt \\ &= \int_0^{a_0} \frac{da}{aH(a)} \quad H(a) = \dot{a}/a \end{aligned}$$

$H(a)$ is again found from the Friedmann equation. With the standard cosmology, $t_0 \approx 14$ Gyr.

2.7 Thermal history

According to the Big Bang paradigm, the universe was once hot and dense, and now it expands and cools down. Today, it is far from thermal equilibrium, but it must have been in thermal equilibrium at some point in the past if it continuously expands.

A system is in thermal equilibrium if $\Gamma \gg H$

- $\Gamma = \text{interactions/time}$ is the interaction rate
- $H = \dot{a}/a$ is the Hubble constant

Similarly, a system is in thermal equilibrium if $\tau_\Gamma \ll \tau_H$

- $\tau_\Gamma = 1/\Gamma$ is the characteristic timescale of interactions
- $\tau_H = 1/H$ is the characteristic timescale of expansion

We already know about H . Γ is defined as

$$\Gamma = nv\sigma$$

- n number density, particles/volume
- v velocity of particles
- σ scattering cross-section, has units of area.

At early times, $\Gamma \gg H$. Particles are in thermal equilibrium with the plasma and coupled to photons. This scenario will be treated in section 2.8.1.

At later times $\Gamma \ll H$. Particles are not in thermal equilibrium and are decoupled from photons. See section 2.8.2

The decoupling or “freeze out” happens when $\Gamma \approx H$. This transition is described by the Boltzmann equation in section 2.8.3.

In table 2.1 and fig. 2.2, an overview of the thermal history of the universe is given.

2.8 Thermodynamics

To describe the evolution of the universe quantitatively, a few definitions are required:

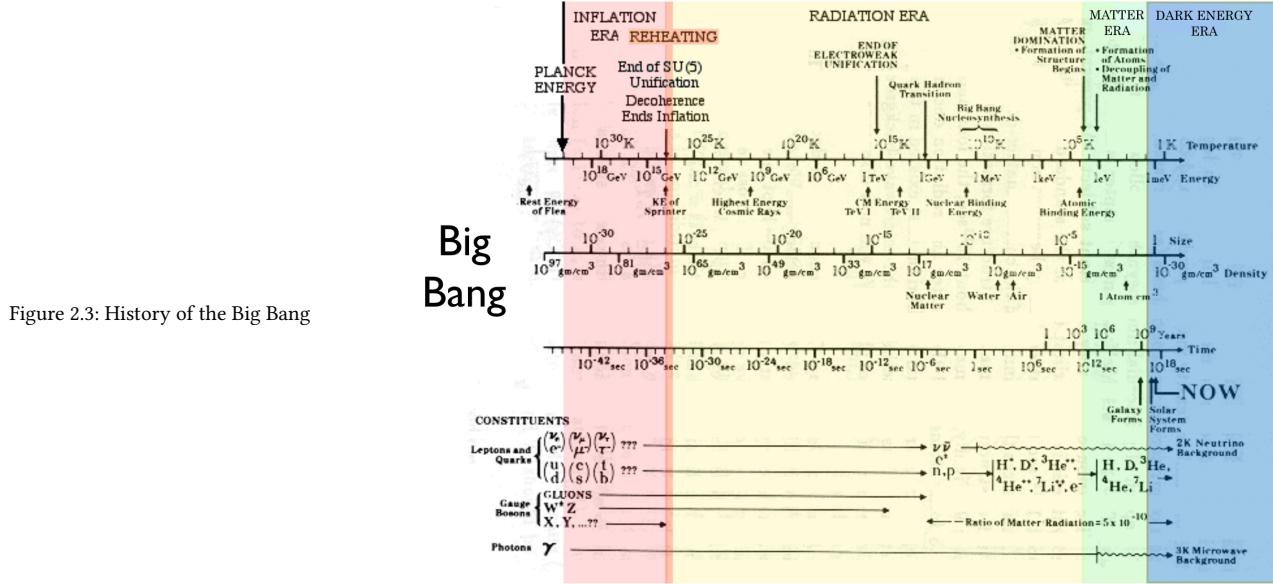
- The probability that a particle is in a volume $d^3x d^3p$ at time t is given by the **(phase space) distribution function**:

$$f(\mathbf{x}, \mathbf{p}, t) d^3x d^3p$$

In a homogeneous and isotropic universe, we have $f(\mathbf{x}, \mathbf{p}, t) = f(p, t)$.

Event		time	redshift	energy temp.
Inflation	A phase of extremely rapid exponential expansion, caused by a phase transitions where the inflaton field emerged. Inflation explains properties of the universe which are difficult to account for without.	?	?	?
Baryogenesis	Baryons (protons, neutrons) are formed from quarks. weirdly, there are way more baryons formed than antibaryons. This is the matter-antimatter asymmetry.	?	?	?
QCD phase transition	The universe has cooled sufficiently such that hadrons (baryons and mesons) can form.	10^{-5} s	10^{12}	200 MeV $3 \cdot 10^{12}$ K
	Pions annihilate and decay, the only hadrons left are nucleons (protons and neutrons).	10^{-4} s		50 MeV 10^{12} K
Dark Matter freeze-out	Dark Matter interacts very weakly with ordinary matter, so it decouples early on.	?	?	?
Electron-positron annihilation	Electrons and positrons annihilate through $e^+ + e^- \rightarrow 2\gamma$. Since the number of charged particles decreases, neutrinos decouple.	4 s	$2 \cdot 10^9$	0.3 MeV $5 \cdot 10^9$ K
Big Bang nucleosynthesis	Light nuclei such as D and He get synthesized. They are still ionized.	3 min	$4 \cdot 10^8$	0.08 MeV 10^9 K
Matter-radiation equality		$6 \cdot 10^4$ yr	3400	0.75 eV 8700 K
Recombination	Formation of neutral atoms through $e^- + p^+ \rightarrow H + \gamma$	$2 \cdot 10^5$ yr	1200	0.34 eV 4000 K
Surface of last scattering	The number density of charged particles has decreased enough for photons to decouple. These photons form the CMB.			
Reionization	Stars form and re-ionize hydrogen.	$2 \cdot 10^8$ yr	20	4 meV 50 K
Dark Energy - Matter equality		9 Gyr	0.4	0.33 meV 3.8 K
Today		13.8 Gyr	0	0.24 eV 2.7 K

Table 2.1: Thermal history of the universe



Adapted from Kolb & Turner 1990

- The number of particles per unit volume is given by the **number density**:

$$n(t) = 4\pi \int f(p, t)p^2 dp$$

- The energy per unit volume is given by the **energy density**:

$$\rho(t) = 4\pi \int E(p)f(p, t)p^2 dp,$$

with $E(p) = \sqrt{p^2 + m^2}$.

- The **pressure** is given by

$$P(t) = 4\pi c^2 \int \frac{p^2}{3E(p)} f(p, t)p^2 dp$$

Since we work in natural units, we drop the c .

2.8.1 At early times ($\Gamma > H$)

The distribution function of particles in thermodynamic equilibrium is the Bose-Einstein or the Fermi-Dirac distribution:

$$f_{\text{eq}}(p, t) = \frac{g}{(2\pi)^3} \left[\exp\left(\frac{E(p) - \mu}{T}\right) \pm 1 \right]^{-1}$$

- $+$ is for fermions and $-$ for bosons
- g is a spin degeneracy factor. Examples: $g_\nu = 1$, $g_\gamma = 2$, $g_{\text{quark}} = 6$
- μ is the chemical potential, which is the response of a thermodynamics system to a change of particle number. Usually, $\mu = 0$ for our purposes.
- T is the temperature of the universe, which is time dependent.

For *non-relativistic particles*, $T \ll m$ and $E \approx m + p^2/2m$. Plugging this in yields

$$n = g \left(\frac{mT}{2\pi} \right)^{3/2} \exp\left(\frac{p - m}{T}\right) \quad \rho = nm \quad P = nT$$

For *relativistic particles*, $T \gg m$ and $E \approx p$. For both fermions and bosons, this yields

$$n = gT^3 \quad \rho = gT^4 \quad P = \frac{\rho}{3}$$

In the case of a relativistic gas, $T \propto a^{-1}$, so $n \propto a^{-3}$ and $\rho \propto a^{-4}$. This is what we have already seen in section 2.4.

2.8.2 At late times ($\Gamma < H$)

The transition between non-equilibrium and equilibrium takes place at the decoupling time of freeze-out time t_f . At $t > t_f$, the transition function is

$$f(p, t) = f\left(p \frac{a(t)}{a(t_f)}, t_f\right)$$

The shape of the function is “frozen in” at the freeze-out time t_f .

For a relativistic particle,

$$f(p, t) = \frac{g}{(2\pi)^3} \left[\exp\left(\frac{pa(t)}{T_f a(t_f)} \pm 1\right) \right]^{-1},$$

which is the same as for an equilibrium particle, but with $T_f := T_f a(t_f)/a(t)$.

2.8.3 Boltzmann equation

The Boltzmann equation

$$\frac{df_i}{dt} = c_i[f_i]$$

describes the time evolution of the distribution function. $c_i[f_i]$ is the collision term.

f only depends on p and t , so we can write

$$\frac{df_i}{dt} = \frac{\partial f_i}{\partial t} + \frac{\partial f_i}{\partial p} \frac{\partial p}{\partial t}$$

The last term can be simplified. Because $p = p_0 a^{-1}$, $dp/dt = -p_0 a^{-1} \dot{a} = -pH$. The Boltzmann equation then becomes

$$\begin{aligned} \frac{\partial f_i}{\partial t} - pH(t) \frac{\partial f_i}{\partial p} &= c_i[f_i] \\ \implies \frac{\partial}{\partial t} \int d^3 p f_i - H(t) \int d^3 p p \frac{\partial f_i}{\partial p} &= \int d^3 p c_i[f_i] \\ \implies \frac{dn_i}{dt} + 3H(t)n_i &= \int d^3 p c_i[f_i] \end{aligned}$$

In the last line, we used partial integration.¹ If the system is collisionless, $c_i[f_i] = 0$, and the solution of the Boltzmann equation is $n_i \propto a^{-3}$. The $3H(t)n_i$ is called the Hubble drag term.

We now look at reactions of the type $i + j \leftrightarrow a + b$. The collision term is then of the form

$$c_i[f_i] = \alpha(T)n_a n_b - \beta(T)n_i n_j$$

- $\alpha(T)$ is the production rate

¹ In detail:

$$\begin{aligned} \int d^3 p p \frac{\partial f_i}{\partial p} &= \int dp p^3 \frac{\partial f_i}{\partial p} \\ &= - \int dp \frac{\partial p^3}{\partial p} f_i \\ &= -3n_i \end{aligned}$$

- $\beta(T)$ is the destruction rate

To simplify the equation, we make a few assumptions:

- a and b are in equilibrium with a general plasma at temperature T
- $n_i = n_j$ (this is the case for antiparticles)
- radiation era: $a \propto t^{1/2} \propto T^{-1}$

The equation is then

$$\frac{dn_i}{dt} + 3H(t)n_i = \beta(T)(n_{i,\text{eq}}^2 - n_i^2)$$

To analyse the equation, we define

- $x = m_i/T$ is used as a time variable
- $y_i = n_i/S$, where S is the entropy
- $\Gamma(x) = n_{i,\text{eq}}(x)\beta(x)$

The equation is then

$$\frac{x}{y_{i,\text{eq}}} \frac{dy_i}{dx} = -\frac{\Gamma(x)}{H(x)} \left[\left(\frac{y_i}{y_{i,\text{eq}}} \right)^2 - 1 \right]$$

Initial conditions: $x \ll 1$ at early times, so $y_i = y_{i,\text{eq}}$.

No analytical solution is known. A set of numerical solutions is shown in fig. 2.4

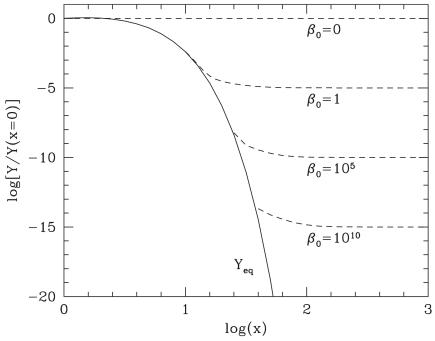


Figure 2.4: Solutions of the Boltzmann equation for different destruction rates β . We assume $\beta(T) = \beta_0$ is constant. The vertical axis is a proxy for abundance, and the horizontal axis for time. First, particles remain in equilibrium (solid line), then they decouple (dashed lines), leaving relic abundances. When β_0 is large, thermal equilibrium is maintained longer, so the relic abundance is lower.

2.9 Particle relics

There are two types of relics:

- Hot relics freeze out when the particles are still relativistic. Since $x = m_i/T$, this means $x_f \ll 1$
- Cold relics freeze out when the particles are already non-relativistic, with $x_f \gg 1$

2.9.1 Hot relics

Hot relics are still relativistic today, so their rest mass is $m_i \ll T_0 = 2.4 \cdot 10^{-4}$ eV. An example would be massless neutrinos.

The solution of the Boltzmann equation is

$$\Omega_{i,0} h^2 = \frac{g_{i,\text{eff}}}{2} \left[\frac{g_{*s}(x_0)}{g_{*s}(x_f)} \right]^{4/3} \Omega_{\gamma,0} h^2$$

The g 's are degeneracy factors and satisfy $g_{*s}(x_0) \leq g_{*s}(x_f)$, and the photon density is $\Omega_{\gamma,0} h^2 = 2.5 \cdot 10^{-5}$. It follows that $\Omega_{i,0}$ is very small, which means that hot relic particles contribute very little to today's energy density.

2.9.2 WIMPs

Next we consider weakly interacting massive particles (WIMPs). Examples are massive neutrinos and stable, light supersymmetric particles. WIMPs can either be hot ($x_f \ll 1$) or cold ($x_f \gg 1$).

Hot WIMPs

The solution of the Boltzmann equation yields

$$\Omega_{i,0} h^2 = 7.64 \cdot 10^{-2} \left(\frac{g_{i,\text{eff}}}{g_{*s}(x_f)} \right) \frac{m_i}{\text{eV}}$$

Since we know that $\Omega_{i,0} < 1$, we can get a constraint $m_i < 100 \text{ eV}$ on the mass of the hot WIMPs, which is possible for neutrinos. However, hot WIMPs are ruled out as dark matter by structure formation arguments.

Cold WIMPs

Boltzmann says

$$\Omega_{i,0} h^2 = \begin{cases} 1.8 \left(\frac{m_i}{\text{GeV}} \right)^{-2} \left[1 + 0.17 \ln \left(\frac{m_i}{\text{GeV}} \right) \right] & \text{if } m_i < 100 \text{ GeV} \\ \left(\frac{m_i}{3 \text{TeV}} \right)^2 & \text{if } m_i > 100 \text{ GeV} \end{cases}$$

To get $\Omega_{i,0} < 1$, we need m_i to be between 1.4 GeV and 3 TeV . Cold Wimps are good candidates for dark matter.

2.10 Primordial nucleosynthesis

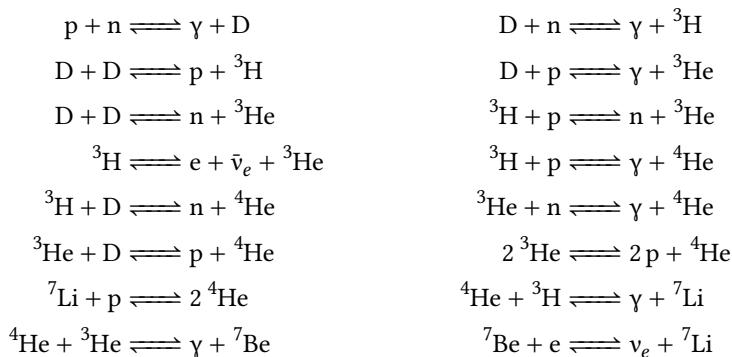
This is the epoch where protons and neutrons first combined to nuclei (not atoms) heavier than hydrogen. Heavier atoms can also be synthesized in stars through nuclear reactions, which has to be differentiated in observations.

Initial conditions

Initially, the temperature is $T < 10^{13} \text{ K}$ and the associated energy scale is $k_B T \approx 0.8 \text{ MeV}$. Protons and neutrons are in thermal equilibrium and interact weakly via the processes $p + e \leftrightarrow n + \nu_e$ and $n + \bar{e} \leftrightarrow p + \bar{\nu}_e$. Since the mass of the nucleons is around 940 MeV , they are already non-relativistic. The masses of neutrons and protons are slightly different, so their abundances after freeze out are different.

Nuclear reactions

Once the temperature drops below 1 MeV , which is the binding energy of a typical nucleus, the nuclei start forming. A host of nuclear reactions can occur then:



For each of these reactions, there is a Boltzmann equation which describes the evolution of the number densities. This coupled system of equations can be solved numerically to find the abundances of nuclei. The solution for a few nuclei is shown in fig. 2.5.

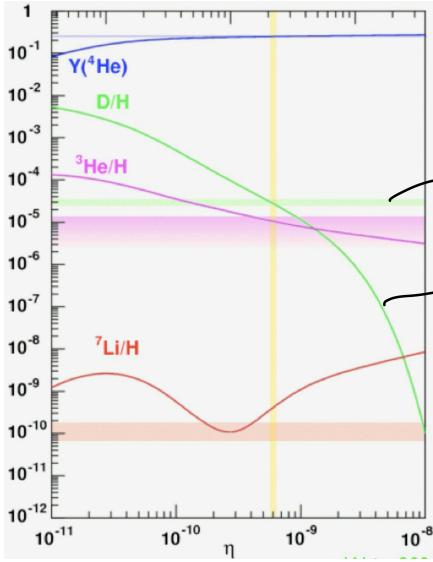


Figure 2.5: The abundances of D, ${}^4\text{He}$, ${}^3\text{He}$, and ${}^7\text{Li}$, as a function of the photon to baryon ration η . The yellow vertical area indicates the measured value of η , while the shaded horizontal bars are the measured abundances. The modelled abundances agree well with the experiments, except for ${}^7\text{Li}$, which is probably due to uncertainties in how much ${}^7\text{Li}$ is destroyed in stars.

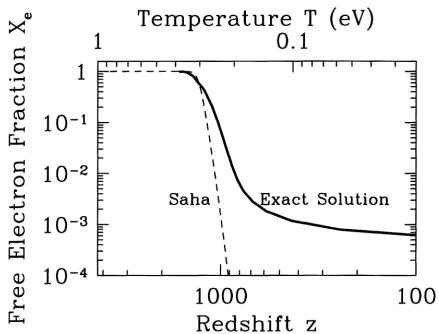


Figure 2.6: The ionization X_e as a function of redshift. Before recombination (at high redshift), the universe is fully ionized with $X_e = 1$. The Saha solution assumes thermal equilibrium, which is only valid until recombination. After the freeze-out, a residual ionization fraction ($X_e \approx 10^{-3}$) remains.

2.11 Recombination

As the universe cools down to a temperature that is lower than the binding energy of hydrogen (13.6 eV), some hydrogen atoms start to form via the reaction $p + e \rightarrow H + \gamma$. This epoch is called recombination, even though electrons and protons combine for the first time. Other atoms also start forming, but we only care about hydrogen for now.s

The ionization fraction x_e is the ratio between the number density of electrons and the number density of baryons (protons, hydrogen):

$$x_e := \frac{n_e}{n_b}$$

For given initial conditions, the ionization fraction can be calculated as a function of z by solving the Boltzmann equation, see fig. 2.6

2.11.1 Recombination

The time of recombination is defined as the time when there are ten times more baryons than electrons: $x_e(z_{\text{rec}}) = 0.1$. This happens at a temperature of $T_{\text{rec}} \approx 0.3 \text{ eV}$ and a redshift of $z \approx 1300$. Note that T_{rec} is smaller than the binding energy of hydrogen. This is because recombination is delayed by the high abundance of photons.

2.11.2 Decoupling

The interaction of electrons and photons (in the relevant energy regime) is described by Thomson scattering, with a scattering cross-section of $\sigma_T = 6.65 \cdot 10^{-25} \text{ cm}^2$. The reaction rate is then (see section 2.7)

$$\Gamma_T = n_e \sigma_T c,$$

where n_e is the number density of electrons, and c the speed of photons. We know that strong coupling occurs as long as $\Gamma_T \gg H$, so we define the moment of decoupling such that

$$\Gamma_T(z_{\text{dec}}) = H(z_{\text{dec}})$$

This happens at $z \approx 1100$, $E \approx 0.26 \text{ eV}$, and $T \approx 3000 \text{ K}$, after the universe is 380 000 yr old. Note that $z_{\text{dec}} < z_{\text{rec}}$, so decoupling occurs soon after recombination.

After decoupling, the universe is transparent to photons. When an observer today stares into empty space, the photons he measures come from the surface of last scattering, where the photons interacted for the last time during decoupling. This is the cosmic microwave background (CMB), which has a temperature of

$$T_{\text{CMB}} = T_{\text{dec}} \frac{a_0}{a_{\text{dec}}} \approx 3 \text{ K}$$

Measurements of the CMB show a blackbody spectrum with remarkably deviations of $\Delta T/T \approx 10^{-5}$, as can be seen in fig. 2.7.

2.12 Λ CDM model

The Λ CDM model is a refinement of the Big Bang model, and it is the standard model of cosmology. Three observational pillars justify this model:

- the expansion of the universe,
- the big bang nucleosynthesis, and
- the cosmic microwave background.

According to Λ CDM, the energy content of the universe today is made up of radiation, matter, dark matter, and dark energy. Their contributions today are:

- Photons and neutrinos: $\Omega_{\gamma,0} \approx \Omega_{\nu,0}/0.68 \approx 2.5 \cdot 10^{-5} h^{-2}$
- Baryons: $\Omega_{b,0} \approx 0.05$
- Dark Matter: $\Omega_{dm,0} \approx 0.25$
- Dark Energy: $\Omega_{\Lambda} \approx 0.7$

We now look at the constituents in more detail.

Radiation

As radiation, we classify those particles that are still relativistic today. Accordingly, they need to have very low mass.

- Most photons are part of the CMB and have a temperature $T \approx 2.73$ K. They contribute only very little to the energy density: $\Omega_{\gamma,0} = 2.5 \cdot 10^{-5}$.
- Massless² neutrinos contribute about the same as photons: $\Omega_{\nu,0} = 0.68\Omega_{\gamma,0} \approx 1.7 \cdot 10^{-5}$.

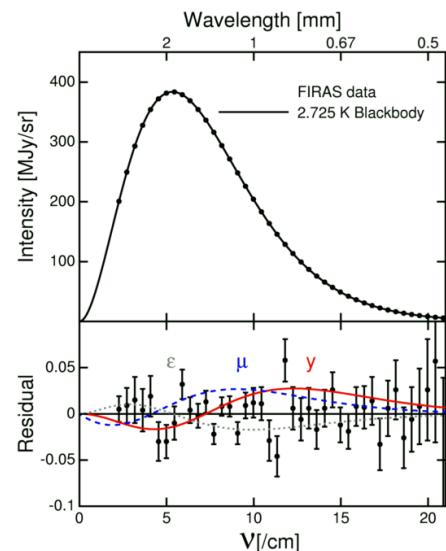


Figure 2.7: The measured spectrum of the cosmic microwave background and a fit with a blackbody spectrum. The residuals show an average deviation of only $\Delta T/T \approx 10^{-5}$.

² Because of neutrino oscillations, we strongly suspect that neutrinos actually have mass, but it is very low.

Baryons

Baryons from the Big Bang nucleosynthesis contribute $\Omega_b \approx 0.05$. They mostly occur in two forms:

- Even though only ten percent of baryons are found in stars, they are the largest fraction of visible matter.
- The rest of the baryons is in the form of various phases, such as cold gas, warm gas, and hot gas, both in the intergalactic and the interstellar medium.

Dark Matter

Dark Matter contributes $\Omega_{dm} \approx 0.25$. We know of its existence through its gravitational effects. There is evidence for dark matter in galaxies, galaxy clusters, large scale structure, and many others.

Dark matter is mostly non-baryonic, because baryons are constrained to $\Omega_b \approx 0.05$ by Big Bang nucleosynthesis. Also, the structure formation constraints below rule out baryons as dark matter candidates.

- Dark matter is cold (non-relativistic). It is thus called Cold Dark Matter (CDM)
- It interacts very weakly, so we can approximate it as non-collisional.

A good candidate for dark matter are particles beyond the standard model, such as WIMPs.

Massive Compact Halo Objects (MACHOs), such as black holes, are not good candidates, since objects with masses from 10^{-6} to $15 M_{\odot}$ are ruled out.

Dark Energy

Dark Energy contributes $\Omega_{\text{de}} \approx 0.7$. It is needed to get a flat geometry for the universe, and to explain the recent acceleration of the expansion of the universe.

From the first Friedmann equation, we know that a constant energy density results in a scale factor $a \propto e^{Ht}$ that is exponentially growing. Consider an effective fluid with equation of state $p_{\text{DE}} = w\rho_{\text{DE}}c^2$. We demand $\ddot{a} > 0$, and the second Friedmann equation then yields $w < -1/3$.

There are a few possible candidates:

- The cosmological constant Λ corresponds to $w = -1$, independent of time.
- A model called *quintessence*, which is a dynamical scalar field, would give rise to another form of energy. Since it is dynamical, you can map it to an effective fluid, with a w that varies in time.
- There could be a theory of gravity that can improve upon or replace general relativity.

Current constraints demand $w = -1$ with an uncertainty of about 5 %. The cosmological constant is thus a model that is consistent with observations, and it is part of the ΛCDM model.

Summary

In total, $\Omega_0 \approx 1$, so the universe has a flat geometry. The exact nature of dark energy and dark matter are some of the most pressing questions in fundamental physics.

There are other ingredients to ΛCDM , such as the model of gravity and the choice of initial conditions. A process called inflation is also added, as we will see later.

The ΛCMD model is very successful to fit current observations, but there are some tensions that are starting to emerge with more detailed measurements from cosmological probes.

2.13 Problems with ΛCDM

The ΛCDM model without inflation has a few intrinsic fundamental problems.

2.13.1 The horizon problem

We have defined the particle horizon

$$\chi_h(t) = \int_0^t \frac{c dt'}{a(t')},$$

which is the maximal comoving distance that a photon could have travelled from the big bang ($t = 0$) until time t . Let's look at the size of the horizon at decoupling time, where photons scattered for the last time. This corresponds to a redshift $z_{\text{dec}} \approx 1100$. The integral evaluates to

$$\chi_h(t_{\text{dec}}) \approx 180h^{-1} \text{ Mpc}$$

The corresponding CMB angular scale is

$$\theta_h = \frac{\chi_h(t_{\text{dec}})}{r(\chi_{\text{dec}})} \approx 1.8 \text{ deg},$$

where r is the comoving angular diameter distance. This is unexpected, because the CMB is extremely homogeneous across the whole sky, even though the photons from different areas of the sky could not have been in causal contact when they scattered! Something must be wrong here.

2.13.2 Flatness problem

We have observationally determined $\Omega_0 \approx 1$, so we have a flat geometry. Let's look at how Ω varies with time. For this, we can use the Friedmann equation:

$$H(a)^2 = \frac{8\pi G}{3} \rho(a) - \frac{Kc^2}{a^2}$$

We look at the deviation of Ω from 1:

$$\begin{aligned} \frac{1 - \Omega(a)}{\Omega(a)} &= \Omega(a)^{-1} - 1 \\ &= -\frac{3Kc^2}{8\pi G\rho(a)a^2} \\ &\propto \begin{cases} a^2 & \text{in the radiation era } \rho \propto a^{-4} \\ a & \text{in the matter era } \rho \propto a^{-3} \end{cases} \end{aligned}$$

Consider time t_i in the radiation era. Then

$$\begin{aligned} \frac{\Omega_i^{-1} - 1}{\Omega_0^{-1} - 1} &= \frac{\Omega_i^{-1} - 1}{\Omega_{\text{eq}}^{-1} - 1} \frac{\Omega_{\text{eq}}^{-1} - 1}{\Omega_0^{-1} - 1} \\ &= \left(\frac{a_i}{a_{\text{eq}}} \right)^2 \left(\frac{a_{\text{eq}}}{a_0} \right) \\ &= \left(\frac{T_{\text{eq}}}{T_i} \right)^2 \frac{T_0}{T_{\text{eq}}} \quad T \propto a^{-1} \text{ for a relativistic gas} \end{aligned}$$

We choose the Planck time as our starting point:

$$t_{\text{init}} = t_{\text{Planck}} := \left(\frac{\hbar G}{c^5} \right)^{1/2}$$

$$T_{\text{Planck}} = 10^{32} \text{ K}$$

$$T_0 = 3 \text{ K}$$

$$T_{\text{eq}} = 10^4 \text{ K}$$

Then

$$\frac{\Omega_{\text{Planck}}^{-1} - 1}{\Omega_0^{-1} - 1} \approx 10^{-60}$$

This means that any deviation of Ω from 1 at the Planck time has been amplified by 60 orders of magnitude until today. This is also called the fine tuning problem, which requires $\Omega = 1$ with extreme accuracy at early times. What is the physical explanation of this?

2.13.3 Monopole problem

Consider $T \approx 10^{14}$ GeV, which was the case during the grand unification (GUT) epoch, where the fundamental forces were unified. At these temperatures, there is a phase

transition. Before the transition, the forces are unified, and during the transition, there is a spontaneous symmetry breaking. A prediction for topological defects that happen during this transition requires that there should be magnetic monopoles with an enormous density $\Omega_{\text{monopole},0} \approx 10^{11}$. If you still exist, this scenario is obviously ruled out by observations.

2.13.4 Structure formation problem

Today, even though the cosmological principle holds for large distance scales (a few 100 Mpc), we still see many structures on various smaller scales in the universe. What were the “seeds”, the initial perturbations that are responsible for the creation of these structures?

2.13.5 Initial condition problem

All of these problems can be resolved by postulating very special and finely tuned initial conditions. These could arise from physics in the quantum gravity era. The benefit of the inflation model is that all these problems could be resolved without having to resort to quantum gravity.

2.14 Inflation

Inflation is a possible solution for the disparities that we have seen in the previous section. We will tackle them one by one from here.

2.14.1 Horizon problem

We define two things to reformulate the horizon problem:

- The size of the forward light cone:

$$\chi_f = \int_0^{t_{ls}} \frac{c dt'}{a(t')}$$

- The size of the past light cone:

$$\chi_p = \int_{t_{ls}}^{t_0} \frac{c dt'}{a(t')}$$

This is the maximum comoving distance that photons could have travelled since last scattering.

The horizon problem can be stated mathematically as $\chi'_f < \chi_p$: Regions which are observable today to have similar temperatures were apparently not in causal contact at last scattering. A solution is to increase χ'_f through a period of accelerated expansion called inflation. A sketch of this is shown in fig. 2.8.

Let t_i and t_e be the start and end time of inflation, with $\Delta t = t_e - t_i$. We know

$$\chi_f = \chi_h(t_{ls}) = \int_0^{t_{ls}} \frac{c dt'}{a(t')}$$

For vacuum energy, we know that $\rho_{\text{vac}}(a)$ is constant. We assume that, at early times, there was some kind of vacuum energy which gives us

$$a \propto e^{Ht} \text{ with } H = \sqrt{8\pi G \rho_{\text{vac}}/3} = \text{constant},$$

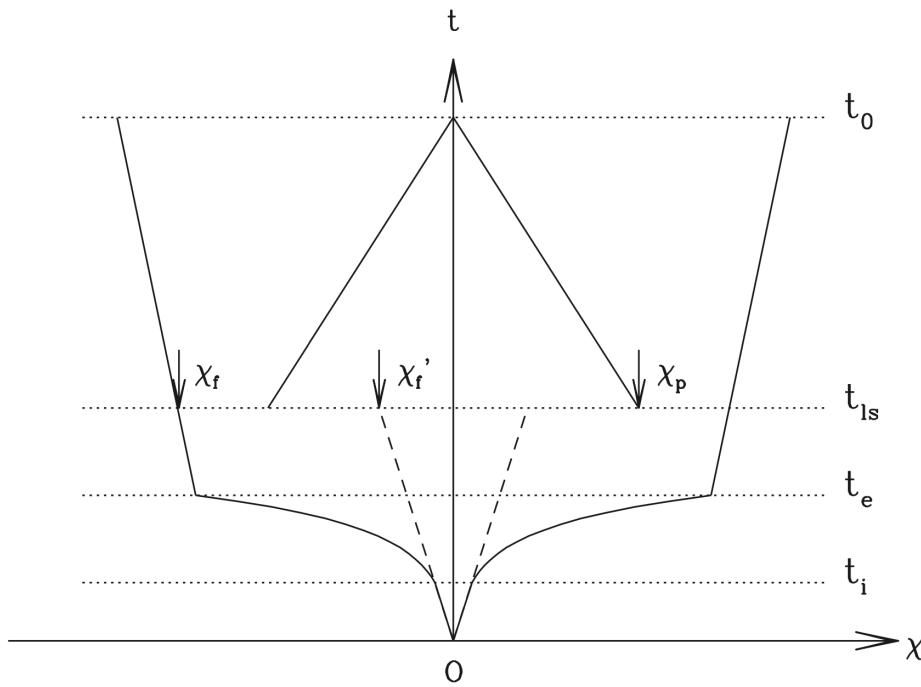


Figure 2.8: The light cone structure in an inflationary universe. Without inflation, the forward light cone (dashed line) would be smaller than our past light cone, χ_p , at the last scattering surface, resulting in causality problems. With a period of inflation from t_i to t_e , however, the forward light cone can be larger than the past light cone at t_{ls} .

exactly like our derivations for the cosmological constant. The contribution to χ_f during inflation is then

$$\begin{aligned}\chi_f(t_i, t_e) &= \int_{t_i}^{t_e} \frac{c dt'}{a(t')} \\ &\propto \frac{1}{Ha(t_e)} (e^{H\Delta t} - 1)\end{aligned}$$

We see that χ_f grows exponentially during inflation. For how long does inflation have to last to solve our problems? We need $\Delta t > 60H^{-1}$, or $e^{H\Delta t} < 10^{25}$, if we assume $t_e \approx t_{GUT}$. We thus require $a_e/a_i > e^{60}$, or in other words, we need 60 “e-foldings”.

2.14.2 Flatness problem

One can show that

$$\frac{\Omega^{-1}(t_e) - 1}{\Omega^{-1}(t_i) - 1} = \frac{a(t_i)}{a(t_e)} < 10^{-52}$$

assuming 60 e-foldings. As a result, any curvature that was originally there gets flattened by inflation.

2.14.3 Monopole problem

The monopoles are diluted by the expansion during inflation by a factor

$$\left(\frac{a_e}{a_i}\right)^3 \approx 10^{78},$$

so the monopole density after inflation is practically zero.

2.14.4 Structure formation problem

As we will see later, inflation provides a mechanism to generate primordial fluctuations. Microscopic quantum fluctuations are turned into macroscopic classical fluctuations by the rapid expansion.

2.14.5 Initial condition problem

Inflation avoids having to set finely tuned initial conditions, since they will be diluted by the expansion. Quantum gravity is not required to solve this problem.

2.14.6 Realization

Inflation requires vacuum energy, which is realized as a scalar field $\phi(x, t)$, called the *inflaton*. The potential of the scalar field is sketched in fig. 2.9.

During inflation, ϕ is in a false vacuum state at a local minimum or a flat part of the potential. During the end of inflation, ϕ reaches the true vacuum state, which has a lower potential energy. This generates vacuum energy, which in turn drives accelerated expansion.

We impose a *slow roll condition*, which impedes the field from changing too quickly:

$$\dot{\phi}^2 \ll V(\phi)$$

There are several models, some of which are old inflation, new inflation, and chaotic inflation. There are many other models proposed, indicating that inflation is still an ongoing field of research.

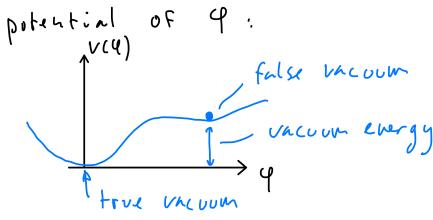


Figure 2.9: The potential of the inflaton field. During inflation, the field moves from its initial false vacuum state to a true vacuum state, which generates vacuum energy.

3 Linear Perturbations

We now drop the assumption that the universe is homogeneous and isotropic. In this chapter, we analyse small perturbations about the background, which can be handled with linear perturbation theory. These perturbations will lead to the structures that we observe today.

We are going to make a few approximations:

- The perturbations are sufficiently small to be treated by linear perturbation theory.
- We ignore relativistic effects and use a Newtonian approximation.

There are several ways to model this structure.

3.1 Ideal Fluid

First, we model the content of the universe as an expanding, self-gravitating, ideal fluid in the Newtonian approximation. We consider the following parameters:

- density $\rho(\mathbf{x}, t)$
- pressure $p(\mathbf{x}, t)$
- velocity $\mathbf{u}(\mathbf{x}, t) = \dot{\mathbf{r}}$
- gravitational potential $\phi(\mathbf{x}, t)$

The time evolution of these parameters is described by the *fluid equations*:

$$\begin{aligned}\frac{D\rho}{Dt} &= -\rho \nabla_r \cdot \mathbf{u} && \text{continuity equation, conservation of mass} \\ \frac{D\mathbf{u}}{Dt} &= -\frac{\nabla_r p}{\rho} - \nabla_r \phi && \text{acceleration} = \text{pressure force} + \text{gravitational force} \\ \nabla_r^2 \phi &= 4\pi G \rho && \text{Poisson equation}\end{aligned}$$

We define the *convective derivative*

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_r,$$

which is the time derivative as one moves along fluid elements.

Assumptions

Let's consider the case where the fluid is expanding. In order to describe this, we can write the position vector as $\mathbf{r} = a(t)\mathbf{x}$, where \mathbf{x} is a comoving coordinate. In the Newtonian treatment, we assume that the scale factor is a known input for the calculation. Then the (total) velocity is

$$\mathbf{u} = \dot{a}(t)\mathbf{x} + \mathbf{v},$$

where the first term is the Hubble expansion, and $\mathbf{v} = a\dot{\mathbf{x}}$ is the peculiar velocity.

We write the density as

$$\rho(\mathbf{x}, t) = \bar{\rho}(t)[1 + \delta(\mathbf{x}, t)],$$

where $\bar{\rho}(t) \propto a^{-3}$ is the mean background density, and $\delta(\mathbf{x}, t)$ is a small density perturbation.

We can plug these assumptions and definitions into the fluid equations to get

$$\begin{aligned}\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1 + \delta)\mathbf{v}] &= 0 \\ \frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a} \mathbf{v} + \frac{1}{a} [\mathbf{v} \cdot \nabla] \mathbf{v} &= -\frac{\nabla \Phi}{a} - \frac{\nabla p}{a \bar{\rho}(1 + \delta)} \\ \nabla^2 \Phi &= 4\pi G \bar{\rho} a^2 \delta\end{aligned}$$

where

$$\Phi = \phi + \frac{1}{2} a \ddot{a} x^2$$

and $\nabla = \nabla_{\mathbf{x}}$.

Thermodynamics

We have three equations in four unknowns, so we still need the equation of state, which relates pressure to other variables, such as density and entropy.

The first law of thermodynamics states that internal energy can either be added by the means of heat or work:

$$dU = dQ + dW$$

We can use $dW = -p dV$ and $dQ = T dS$. From now on, we assume we are dealing with an ideal gas. The equation of state for an ideal gas is

$$\begin{aligned}p &= nk_B T \\ &= \frac{\rho}{\mu m_p} k_B T\end{aligned}$$

where μ is the mean molecular weight and $n = N/V$ is the number density. The internal energy is

$$\begin{aligned}U &= \frac{3}{2} N k_B T \\ &= \frac{3}{2} \frac{\rho}{\mu m_p} V k_B T\end{aligned}$$

Then

$$\frac{\nabla p}{\bar{\rho}} = \frac{1}{\bar{\rho}} \left[\left(\frac{\partial p}{\partial \rho} \right)_S \nabla \rho + \left(\frac{\partial p}{\partial S} \right)_\rho \nabla S \right]$$

Let $c_s^2 = (\partial p / \partial \rho)_S$ be the adiabatic sound speed squared. Then we get

$$\left(\frac{\partial p}{\partial S} \right)_\rho = \frac{2}{3} \rho T$$

and thus

$$\frac{\nabla p}{\bar{\rho}} = c_s^2 \nabla \delta + \frac{2}{3} (1 + \delta) T \nabla S$$

The Euler equation can then be rewritten as

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a} \mathbf{v} + \frac{1}{a} (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{\nabla \Phi}{a} - \frac{c_s^2}{a} \frac{\nabla \delta}{1 + \delta} - \frac{2T}{3a} \nabla S$$

Solving the fluid equation

For small density perturbations δ and small peculiar velocities \mathbf{v} , we can drop terms that are quadratic in these variables, such as $(\mathbf{v} \cdot \nabla)\mathbf{v}$. We get the linearized fluid equations:

$$\begin{aligned}\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot \mathbf{v} &= 0 \\ \frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a} \mathbf{v} &= -\frac{\nabla \Phi}{a} - \frac{c_s^2}{a} \nabla \delta - \frac{2\bar{T}}{3a} \nabla S\end{aligned}$$

We can combine them to get a single equation. First, take the derivative of the continuity equation. Then, use the Euler equation and the Poisson equation to get the following:

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a} \frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho} \delta + \frac{c_s^2}{a^2} \nabla^2 \delta + \frac{2}{3} \frac{\bar{T}}{a^2} \nabla^2 S$$

This is a second order differential equation. We take the Fourier transform:

$$\begin{aligned}\delta(\mathbf{x}, t) &= \sum_{\mathbf{k}} \delta_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{x}) \\ \delta_{\mathbf{k}}(t) &= \frac{1}{V} \int \delta(\mathbf{x}, t) \exp(-i\mathbf{k} \cdot \mathbf{x}) d^3x,\end{aligned}$$

where V is the volume of a sufficiently large box. The fluid equation then becomes³

$$\frac{d^2 \delta_{\mathbf{k}}}{dt^2} + 2\frac{\dot{a}}{a} \frac{d\delta_{\mathbf{k}}}{dt} = \left[4\pi G \bar{\rho} - \frac{k^2 c_s^2}{a^2} \right] \delta_{\mathbf{k}} - \frac{2}{3} \frac{\bar{T}}{a^2} S_{\mathbf{k}}$$

³ $\nabla \rightarrow i\mathbf{k}$ and $\nabla^2 \rightarrow -k^2$

The Poisson equation in Fourier space is

$$-k^2 \Phi_{\mathbf{k}} = 4\pi G \bar{\rho} a^2 \delta_{\mathbf{k}}$$

The fluid equation can now be solved for every \mathbf{k} -mode independently.

We need initial conditions to solve the fluid equation. There are two pressure terms, $\delta_{\mathbf{k}}$ and $S_{\mathbf{k}}$. We distinguish two types of initial conditions:

- $\delta \neq 0, \delta S = 0$: isentropic or adiabatic or curvature perturbations
- $\delta = 0, \delta S \neq 0$: isocurvature perturbations

Adiabatic perturbations are naturally generated by inflation, so we only consider those and set $\delta S = 0$.

The fluid equation can now be simplified further:

$$\frac{d^2 \delta_{\mathbf{k}}}{dt^2} + 2\frac{\dot{a}}{a} \frac{d\delta_{\mathbf{k}}}{dt} = \left[4\pi G \bar{\rho} - \frac{k^2 c_s^2}{a^2} \right] \delta_{\mathbf{k}}$$

Consider the case where the expansion of the fluid can be ignored, so a is a constant. Then $\dot{a} = 0$, so a further term drops out:

$$\begin{aligned}\frac{d^2 \delta_{\mathbf{k}}}{dt^2} &= \left[4\pi G \bar{\rho} - \frac{k^2 c_s^2}{a^2} \right] \delta_{\mathbf{k}} \\ &= -\omega^2 \delta_{\mathbf{k}}\end{aligned}$$

This is simply the equation of motion of a harmonic oscillator. We can rewrite

$$\omega^2 = \left(\frac{c_s}{a} \right)^2 [k^2 - k_J^2]$$

where k_J is the JEANS wavenumber

$$k_J = \frac{a}{c_s} \sqrt{4\pi G \bar{\rho}}.$$

One can also define the Jeans wavelength

$$\lambda_J = \frac{2\pi a}{k_J},$$

and the Jeans mass

$$M_J = \frac{\pi}{6} \bar{\rho} \lambda_J^3,$$

which is the mass of a sphere with a radius of $\lambda_J/2$ and density $\bar{\rho}$.

Solutions

We first consider the case where $k > k_J$, or equivalently $\lambda < \lambda_J$, or $M < M_J$. Then $\omega^2 > 0$, and the equation of motion gives us an oscillatory solution:

$$\delta_k \propto \exp(\pm i\omega t) \quad \omega \in \mathbb{R}$$

This means that density perturbations don't grow, but just oscillate.

When $k < k_J$, the solution is

$$\delta_k \propto \exp(\pm \alpha t) \quad \alpha \in \mathbb{R},$$

which indicates exponentially decaying or growing modes. The growing modes lead to the growth of structure, which is called gravitational or Jeans instability. Once the perturbations become large, our assumptions are not valid any more, and we have to switch to non-linear perturbation theory.

Jeans mass

The Jeans mass is

$$\begin{aligned} M_J &= \frac{\pi}{6} \bar{\rho} \lambda_J^3 \\ &= \frac{\pi^{5/2}}{6} \frac{c_s^3}{G^{3/2} \bar{\rho}^{1/2}} \end{aligned}$$

Before recombination, photons and baryons are tightly coupled, and they act as a single fluid. We get $M_J \approx 10^{16} (\Omega_{b,0} h^2)^{-2} M_\odot$, which means that no baryonic perturbations smaller than a supercluster can grow before recombination.

After recombination, the baryons and the photons are decoupled. They have much smaller pressure, and $M_J \approx 10^5 (\Omega_{b,0} h^2)^{-1/2} M_\odot$. Perturbations with masses larger than a globular cluster can grow.

We also have to take into account the expansion of the universe, which we neglected up to now. Furthermore, these results do not consider dark matter, but only baryons.

3.2 Collisionless gas

We now consider a collisionless gas, such as dark matter. In this case, there is no reason to impose thermodynamic equilibrium, not even locally. As a result, we cannot describe the gas with the thermodynamic variables we used before, since they are only defined in (at least local) thermodynamic equilibrium. We thus need to take one step back and consider distribution functions $f(\mathbf{x}, \mathbf{p}, t)$, whose dynamics are given by the collisionless Boltzmann equation,

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial p^i} \frac{\partial p^i}{\partial t} + \frac{\partial f}{\partial x^i} \frac{\partial x^i}{\partial t} = 0.$$

As before, we can take the moments of the distribution functions:

$$\langle Q \rangle = \frac{1}{n} \int d^3 p Q f(\mathbf{x}, \mathbf{p}, t)$$

with the comoving number density $n(\mathbf{x}, t) = \int d^3 p f(\mathbf{x}, \mathbf{p}, t)$. Some moments are:

- The density $\rho(\mathbf{x}, t)$ can be obtained with $Q = m$, where m is the mass of a particle.
- The bulk velocity $\langle v_i \rangle$ can be obtained with $Q = v_i = p_i / (mn)$

We take moments of the Boltzmann equation. One can show⁴ that one gets the same equations as the fluid equations: the continuity equation and the Euler equation, with the pressure gradient term replaced as follows:

$$\frac{\nabla_i p}{\rho} \rightarrow \sum_{j=1}^3 \frac{\partial}{\partial x_j} [(1 + \delta) \sigma_{ij}^2]$$

⁴ For more details, have a look at ??.

where

$$\sigma_{ij} = \langle v_i v_j \rangle - \langle v_i \rangle \langle v_j \rangle$$

The stress tensor is defined as $\rho \sigma_{ij}^2$.

If the stress tensor is small, we get the fluid equations with $\nabla p = 0$ and $c_s^2 = 0$, so a pressureless fluid.

On small scales, the stress tensor can be important. The large random velocities of the particles can make these particles diffuse and dampen the perturbations. This phenomenon is called *free streaming*. This is not important for cold dark matter, but it is important for hot dark matter.

3.3 Solutions

3.3.1 Collisionless gas

We assume adiabatic perturbations, neglect free streaming, and only consider cold dark matter. In the ideal fluid equations, we can set $p = 0$ and $c_s^2 = 0$. We get

$$\frac{d^2 \delta_k}{dt^2} + 2 \frac{\dot{a}}{a} \frac{d \delta_k}{dt} = 4\pi \bar{\rho}_m \delta_k$$

We consider solutions for an expanding universe and find the following solutions (homework):

- The solution $\delta_- \propto H(t)$ is decaying, since H becomes smaller over time.

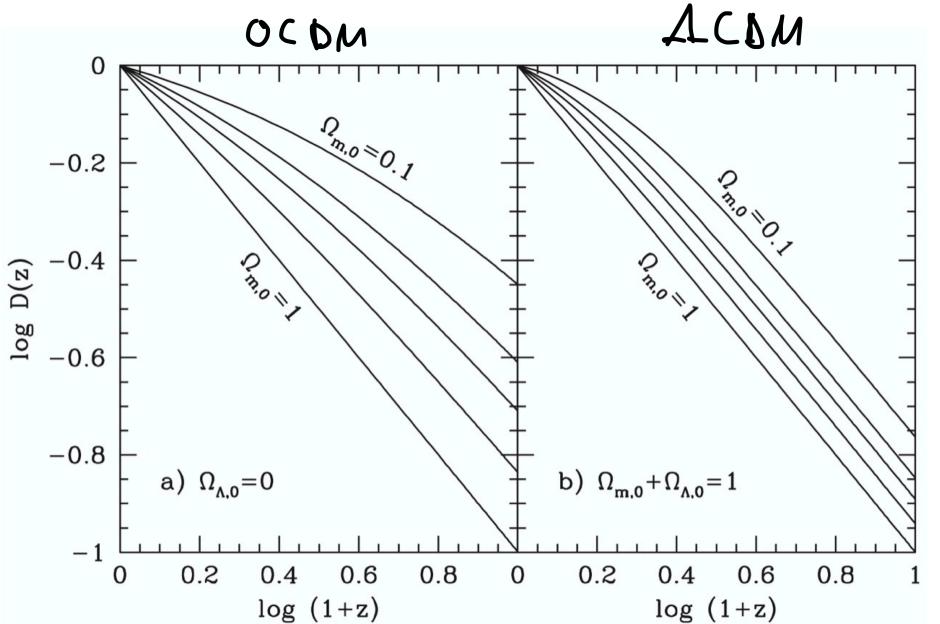


Figure 3.1: Collisionless gas

- The growing solution is

$$\delta_+ \propto H(t) \int_0^t \frac{dt'}{a(t)^2 H(t')},$$

which is a result of gravitational instability. This is similar to what we found before, but this time we did not neglect expansion. We define $D(z) \propto \delta_+$ the linear growth factor.

Examples are shown in fig. 3.1.

- Matter dominated case ($\Omega_0 = \Omega_{m,0} = 1$): $a \propto t^{2/3}$ and $H \propto t^{-1}$. This yields

$$D \propto t^{2/3} \propto a$$

- OCDM ($\Omega_{m,0} < 1, \Omega_{\Lambda,0} = 0$) and Λ CDM ($\Omega_{m,0} + \Omega_{\Lambda,0} = 1$): We get an analytical solution where D grows slower than a . The faster expansion from Λ CDM slows down the growth of structure.
- In the general case, there is no analytical solution. The growth factor usually grows slower than the exponential growth we found in the non-expanding gravitational instability case. In other words, the universe expansion slows down with the growth of structure.

Let's look at how the gravitational potential grows, because we will need that result later. The Poisson equation in Fourier space is

$$-k^2 \Phi_k = 4\pi G \bar{\rho} \delta_k,$$

which can be rearranged to

$$\begin{aligned} \Phi_k &\propto a^2 \bar{\rho}_m \delta_k \\ &\propto a^2 a^{-3} D(a) \\ &\propto \frac{D(a)}{a} \end{aligned}$$

This means that Φ_k is constant in the matter dominated case, and Φ_k decays in the OCDM and Λ CDM cases at late times.

3.3.2 Two non-relativistic components

We consider collisionless dark matter and collisionfull baryons. We assume that the pressureless dark matter dominates the matter density, so $\rho_{\text{tot}} \approx \rho_{\text{dm}}$.

There are now two differential equations:

$$\begin{aligned}\frac{d^2\delta_{\text{dm},k}}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta_{\text{dm},k}}{dt} &= 4\pi G\bar{\rho}_m\delta_{\text{dm},k} \\ \frac{d^2\delta_{\text{b},k}}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta_{\text{b},k}}{dt} + \frac{k^2 c_s^2}{a^2} \delta_{\text{b},k} &= 4\pi G\bar{\rho}_m\delta_{\text{dm},k}\end{aligned}$$

The first equation is the same as in the previous section, because the low-density baryons don't influence the dark matter much. To get analytical solutions, we assume that we are in the matter dominated case, and that $c_s^2 a$ is constant.⁵

Since the dark matter equation is the same, it has the same solution, with a growing mode $D(t) \propto a$ in a matter dominated universe.

For the baryons

$$\delta_{\text{b},k}(t) = \frac{\delta_{\text{dm},k}(t)}{1 + (k/k_J)^2},$$

where $k_J = a/c_s \sqrt{4\pi G \bar{\rho}}$ is the Jeans wavenumber. From this equation, we see that the baryon perturbations are equal to the dark matter perturbations for small k . For large k , pressure becomes more important, and the baryon perturbations will be smaller.⁶

⁵ The latter is equivalent to the condition that the baryons are a specific kind of polytropic fluid, which obeys $p \propto \rho^{4/3}$. For different polytropic indices $\gamma \neq 4/3$, the behaviour is similar, but the equations have to be solved numerically.

⁶ Keep in mind that large k means small length scales, and vice versa. When particles are close together, pressure becomes important.

Acoustic oscillations We have just seen that baryon behaviour becomes a lot more interesting at smaller length scales, where pressure is higher. This leads to oscillations and damping of the amplitude of the perturbations. One can show that the baryon density of a given mode k oscillates like

$$\delta_{\text{b},k} \propto \exp\left(\frac{ikc_s t}{a}\right) + \text{constant},$$

which corresponds to sound waves, which is why they are called *acoustic oscillations*.

Collisional damping Before recombination on small scales, the perturbations are damped by imperfect coupling between baryons and photons. This leads to diffusion of the photons, which suppresses perturbations on small scales. The effect is called *collisional damping* or *Silk⁷ damping*.

⁷ After Joseph Silk, who first discussed photon diffusion in a paper in 1968.

3.4 Relativistic perturbations

For a full treatment of cosmological perturbations, we need a relativistic treatment. We will only give a sketch of the formalism here.

The Einstein equations give the time evolution of space-time. The perturbed version of these equations is

$$\bar{G}_{\mu\nu} + \delta G_{\mu\nu} = 8\pi G(\bar{T}_{\mu\nu} + \delta T_{\mu\nu}),$$

where the variables with a bar are the background solutions from the FRW metric, and the δ -terms are sufficiently small perturbations, so that we can treat them with linear perturbation theory.

In order to calculate G , we need a metric, which we assume to be of the form

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu},$$

where \bar{g} is the FRW metric, and h is again a small perturbation. The perturbation h can again be decomposed, according to the behaviour of their components under coordinate changes.

- Tensor perturbations correspond to gravitational waves, which are important for the polarization of the CMB. However, they are too small to have been detected directly from the CMB.
- Vector perturbations are vorticity modes, which decay. They are thus not very important in cosmology.
- Scalar perturbations can be thought of perturbations in the gravitational potential. They play a central role in structure formation, so we will focus solely on them.

The stress energy tensor can be related to the distribution function of different components i of the universe,

$$\begin{aligned} T_{\mu\nu} &= \sum_i T_{\mu\nu}^i \\ T_{\mu\nu}^i &= \int \frac{d^3 p}{(2\pi)^3 E} p_\mu p_\nu f^i(\mathbf{x}, \mathbf{p}, t), \end{aligned}$$

where f can be found from the Boltzmann equation. The distribution function can also be split into a background and a small perturbation:

$$f^i = \bar{f}^i + \delta f^i.$$

These species i are dark matter, baryons, photons, neutrinos, and potentials of the metric perturbations. We only consider scalar perturbations. All these species interact with each other:

- All species interact with the metric.
- Photons and electrons interact through Compton scattering.
- Electrons and protons interact through Coulomb scattering.

Other forces, such as the weak interaction, are neglected, since they are too rare or weak.

The resulting set of linear differential equations are called the *linearised Einstein-Boltzmann equations*. They have no analytical solution in general, but there are analytical solutions in asymptotic limits.

Solutions in asymptotic limits

For each \mathbf{k} mode, the equations can be solved separately. There are several important scales:

- k^{-1} is the comoving wavelength scale of the \mathbf{k} -mode
- χ_h is the comoving horizon
- a_{eq} is the scale factor at matter-radiation equality, or a_{eq2} at matter-dark energy equality
- $\chi_h(a_{eq})$ is the size of the horizon at matter-radiation equality

The solution of the Einstein-Boltzmann equations is the potential Φ , shown in fig. 3.2. We consider the evolution of a very large and a very small mode:

- Large modes ($k \ll k_{eq}$) are constant in the radiation era. When entering the matter era, they lose about 10 % of their amplitude, but then remain constant.

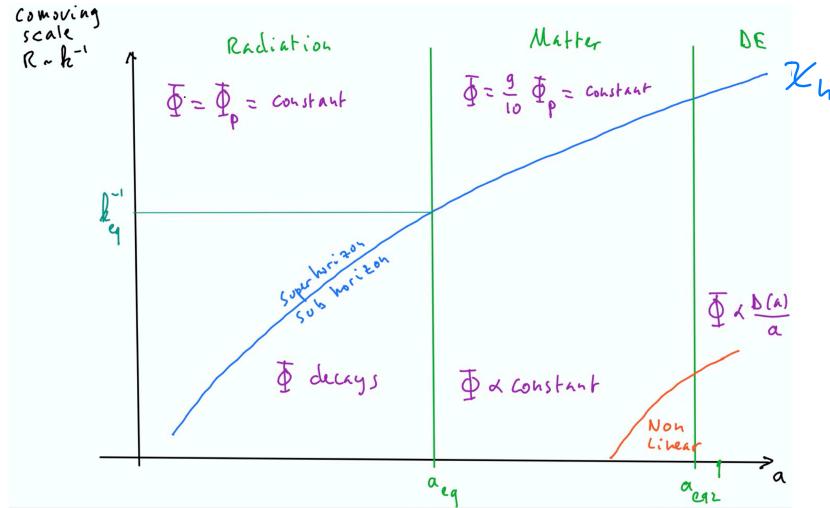


Figure 3.2: Solutions in asymptotic limits. Time is represented by a on the horizontal axis, with early times on the left and today on the right. The cosmological eras are separated by green lines. The comoving wavelength k^{-1} is shown on the vertical axis. The horizon scale χ (blue) determines whether the scale of a mode is inside (sub-horizon) or outside the horizon (superhorizon). As time increases, larger modes enter the horizon. At late time, small modes become non-linear.

- Small modes ($k \gg k_{\text{eq}}$) decay in the radiation era, but remain constant in the matter era. In the dark energy era, they evolve as $D(a)/a$, which we have already seen in the Newtonian treatment.

In summary, small modes are suppressed, while large modes stay mostly constant.

3.5 Transfer function

We can describe the transition from early times to the matter dominated era by a *transfer function*. We write the potential at late times ($a \gg a_{\text{eq}}$) as

$$\Phi_k(a) = \frac{9}{10} \Phi_{p,k} T(k) \frac{D(a)}{a}$$

- $\Phi_{p,k}$ is the primordial potential, which is the potential at initial conditions.
- $T(k)$ is the transfer functions, which obeys $T \rightarrow 1$ as $k \rightarrow 0$
- $D(a)$ is the growth factor, normalized such that $D(a) = a$ in the matter era.⁸
- The 9/10 factor is the large scale suppression which we have seen in the previous diagram.

⁸ We have derived $D(a) \propto a$ before, so we can do this.

The transfer function describes how much of the primordial potential is still left at a time corresponding to scale factor a . The solution for different cases is shown in fig. 3.3.

At small k , $T = 1$ as required.

- In the CDM model, $T(k)$ is suppressed at $k \gg k_{\text{eq}}$, as we have seen before. We find $T(k) \propto (k/k_{\text{eq}})^2 \ln(k/k_{\text{eq}})$.
- In the HDM model⁹, the perturbations are damped strongly due to free streaming when $k > k_\nu$.
- In the baryon model, dark matter is made up of baryons, which is ruled out for different reasons, but we can still look at the solution. Here, two processes act in addition to k_{eq}^{-1} suppression:
 - Jeans instability for $k > k_J$ inhibits growth, and instead leads to oscillations. We have seen these as oscillations in time at constant k before, while here they are shown as oscillations in k at constant time.
 - Silk damping.

⁹ For example, dark matter which is made up of neutrinos

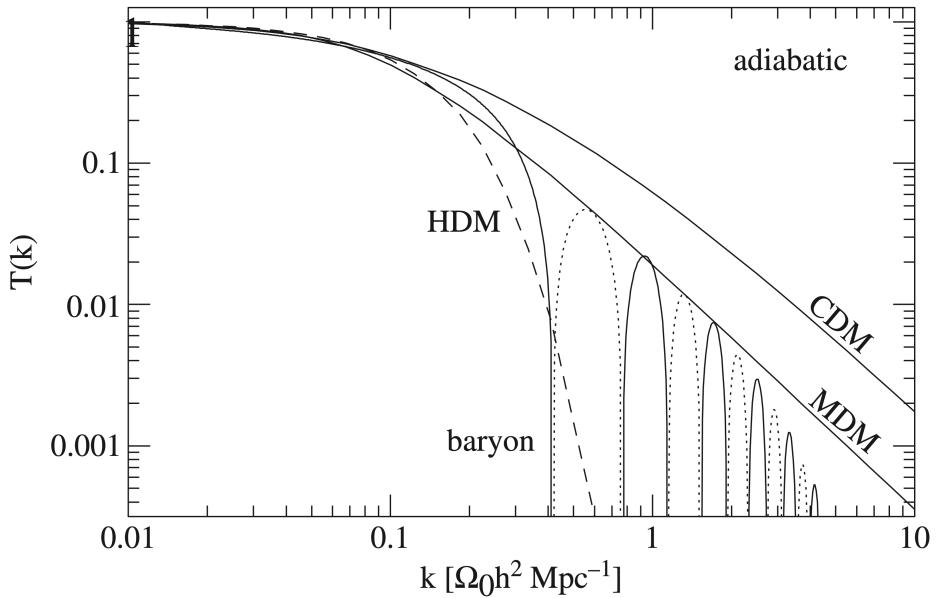


Figure 3.3: The transfer function for adiabatic perturbations. Results are shown for a purely baryonic model, for a cold dark matter model, a hot dark matter model, and a mixed dark matter model. The dotted parts indicate negative values.

As a result, $T(k)$ is suppressed and oscillates on small scales.

- In a mixed model¹⁰, nothing very interesting happens.
- In CDM + baryons (blue line), small oscillations occur around the CDM curve.

3.6 Primordial perturbations

Inflation also gives a mechanism to generate primordial perturbations. Remember that we proposed an inflaton field ϕ , which is in a false vacuum configuration at the beginning, leading to vacuum energy. This drives an accelerated expansion. We required that the scale factor grew by $a_e/a_i > e^{60}$ during inflation, so small features are blown up to macroscopic features. These act as the initial perturbations for formations of structure.

The quantum harmonic oscillator is described by the Hamiltonian

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2x^2,$$

which gives rise to equally spaced energy levels

$$E_n = \frac{\hbar}{m\omega} \left(m + \frac{1}{2} \right),$$

for $n \in \{0, 1, 2, \dots\}$. We consider a state $|n\rangle$, which deviates from its average position by

$$\langle x^2 \rangle = \langle n|x^2|n \rangle = \frac{\hbar}{m\omega} \left(n + \frac{1}{2} \right).$$

We can see that even for the ground state $n = 0$, fluctuations exist. For inflation, each k -mode $\Phi_{p,k}$ acts like a quantum harmonic oscillator.

¹¹ What this means is explained in the next section.

We define the power spectrum¹¹ as the second moment of the Fourier-transformed primordial potential, $P(k) \propto \langle |\Phi_{p,k}|^2 \rangle$. One can show that $P(k) \propto k^{n-4}$, where n is arbitrarily close to 1, and depends on the slow roll parameters $dV/d\phi$ and $d^2V/d\phi^2$. The case $n = 1$ is called *scale invariant* or *Harrison-Zel'dovich*.

Inflation generically predicts the following:

- The perturbations are adiabatic, which was explained in the previous chapter.
- The perturbations are Gaussian, which will be explained in the next section.

3.7 Cosmological field statistics

The density contrast $\delta(\mathbf{x}, t)$ can be measured more-or-less accurately by observations. We can also get a simulated density contrast from a model. There is, however, a problem when trying to compare the measurements with the prediction of the model, since we are attempting to compare two different realizations of a random quantum process.

One way to solve this problem is to compare not the particular density contrasts, but the underlying probability distributions. For a given model, these do not suffer from any randomness, and can therefore be compared directly.

We divide the universe into n cells, which are centred at positions $\mathbf{x}_1, \dots, \mathbf{x}_n$. The random perturbation field is then characterized by the probability distribution function,

$$\text{Prob}_x(\delta_1, \dots, \delta_n) d\delta_1 \cdots d\delta_n,$$

which is the probability that the field δ has values in the range δ_i to $\delta_i + d\delta_i$ at positions \mathbf{x}_i .

As we see, this function Prob_x could be very complicated, potentially taking a huge number of arguments as input. To simplify the situation, we look only at some statistical moments of Prob_x .

3.7.1 Statistical moments

A statistical moment is a quantity that describes the shape of a function. If the function is a probability distribution, some well-known examples of moments are the expected value and the variance.

The moments of a probability distribution Prob_x are defined as¹²

$$\langle \delta_1^{\ell_1} \cdots \delta_n^{\ell_n} \rangle = \int \delta_1^{\ell_1} \cdots \delta_n^{\ell_n} \text{Prob}_x(\delta_1, \dots, \delta_n) d\delta_1 \cdots d\delta_n,$$

where $\ell_i \in \{0, 1, 2, \dots\}$. Because of the cosmological principle, all moments are invariant under spatial translation and rotation.

The first moment is the expected value, $\langle \delta \rangle = 0$, because the density distribution is expected to be zero by assumption.¹³

Some more moments are the variance $\sigma^2 = \langle \delta^2 \rangle$, and the two-point correlation function

$$\xi(x) = \langle \delta_1 \delta_2 \rangle, \quad \text{with } x = |\mathbf{x}_1 - \mathbf{x}_2|.$$

Note that $\xi(0) = \sigma^2$, and that $\xi(x)$ only depends on the distance between \mathbf{x}_1 and \mathbf{x}_2 .

The same analysis can be performed in Fourier space,¹⁴ where we define the power spectrum,

$$P(k) := V \langle |\delta_k|^2 \rangle,$$

¹² If this looks unfamiliar, compare it with the notation that is used in mathematics for the n -th moment μ_n of a single-variate probability distribution P ,

$$\mu_n = \int x^n P(x) dx.$$

¹³ Compare this with a perhaps more familiar notation of the expected value, $\mu = \int x P(x) dx$.

¹⁴ Remember our convention for the Fourier transform,

$$\begin{aligned} \delta(\mathbf{x}, t) &= \sum_k \delta_k(t) \exp(i\mathbf{k} \cdot \mathbf{x}) \\ \delta_k(t) &= \frac{1}{V} \int \delta(\mathbf{x}, t) \exp(-i\mathbf{k} \cdot \mathbf{x}) d^3x. \end{aligned}$$

Also, be aware that we have to convert the probability distribution to Fourier space, written as Prob_k .

which is the variance in Fourier space. It can be related to the two-point correlation function:

$$P(k) = \int \xi(|\mathbf{x}|) \exp(-i\mathbf{k} \cdot \mathbf{x}) d^3x,$$

$$\xi(r) = \frac{1}{(2\pi)^3} \int P(k) \exp(i\mathbf{k} \cdot \mathbf{r}) d^3k.$$

The variance in position space can then be written in terms of the power function,

$$\begin{aligned} \langle \delta^2 \rangle &= \frac{1}{(2\pi)^3} \int P(k) d^3k \\ &= \frac{1}{2\pi^2} \int P(k) k^2 dk \\ &= \frac{1}{2\pi^2} \int P(k) k^3 d \ln(k). \end{aligned}$$

We define $\Delta^2(k) = k^3 P(k)/2\pi^2$, which is the contribution to the variance per $\ln(k)$ interval.

3.7.2 Smoothing

If we only want to analyse features larger than a certain size limit, we convolve the density contrast with a smoothing kernel. An example for such a kernel is the top hat sphere,

$$W_R(r) = \begin{cases} \frac{3}{4\pi R^3} & \text{if } r \leq R, \\ 0 & \text{otherwise.} \end{cases}$$

The smoothed field δ_R is then defined as a convolution of the original field and the kernel:

$$\delta_R(\mathbf{x}, t) = \int d^3y \delta(\mathbf{y}, t) W_R(\mathbf{x} - \mathbf{y}),$$

which is commonly written $\delta_R = \delta * W_R$.

The variance of the smoothed field is then¹⁵

$$\langle \delta_R^2 \rangle = \frac{1}{2\pi^2} \int P(k) \tilde{W}_R^2 k^2 dk,$$

where \tilde{W}_R is the Fourier transform of W_R , which is

$$\tilde{W}_R(k) = \frac{3}{kR^2} [\sin(kR) - kR \cos(kR)].$$

3.7.3 Gaussian fields

The probability distribution can be fully recovered if all of its statistical moments are known, as long as the distribution is homogeneous and isotropic. *Gaussian fields*, where $\text{Prob}(\delta(\mathbf{x}_1) \dots \delta(\mathbf{x}_n))$ is proportional to a multivariate Gaussian function, can still be fully recovered if only the two-point correlation function is known. Consequently, higher order moments can be expressed in terms of the two-point correlation function.

In the standard model of cosmology, we assume that the perturbation fields are initially Gaussian, but they can evolve to non-Gaussian fields later. Then, the initial power spectrum of the perturbation field is all we need to know to fully characterize the field.

¹⁵ We can simplify the calculation by using the convolution theorem,

$$\begin{aligned} \widetilde{f * g} &= \widetilde{f} \cdot \widetilde{g}, \\ \widetilde{f \cdot g} &= \widetilde{f} * \widetilde{g}. \end{aligned}$$

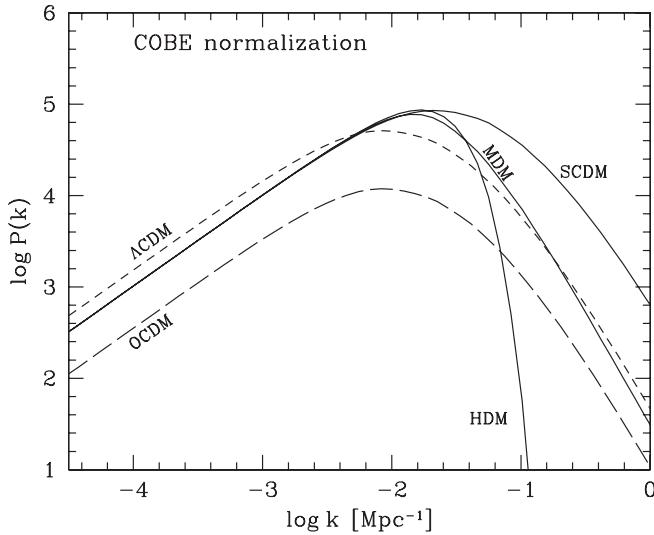


Figure 3.4: Matter power spectrum

3.8 Matter power spectrum

We can now combine the results of the previous three sections. The Poisson equation is

$$-k^2 \Phi_{\mathbf{k}}(a) = 4\pi G \bar{\rho}_m a^2 \delta_{\mathbf{k}}(a),$$

which can be rearranged and combined with the definition of the transfer function to

$$\delta_{\mathbf{k}}(a) = \frac{2}{5} \frac{k^2}{\Omega_m H_0^2} \Phi_{p,\mathbf{k}} T(k) D(a),$$

where $a \gg a_{\text{eq}}$, as we required when we introduced the transfer function. The primordial power spectrum obeys $P(k) \propto k^{n-4}$, and it is commonly written as

$$P_{\Phi_p}(k) = \frac{50\pi^2}{9k^3} \left(\frac{k}{H_0} \right)^{n-1} \delta_H^2 \left(\frac{\Omega_m}{D(a=1)} \right)^2,$$

where δ_H is a normalization parameter.¹⁶ Combining all these, we get

$$\begin{aligned} P(k, a) &= V \langle |\delta(k, a)|^2 \rangle \\ &= 2\pi^2 \delta_H^2 \frac{k^n}{H_0^{n+3}} T(k)^2 \left(\frac{D(a)}{D(a=1)} \right)^2. \end{aligned}$$

The matter power spectrum is plotted as a function of k for different cosmological models in fig. 3.4.

- In the ΛCDM model, the turnover is at $k_{\text{eq}} \propto \Omega_m h^2$.
- In the SCDM model (no cosmological constant, compensated with larger Ω_m), the turnover is at a higher k .
- In the HDM model, free streaming suppresses perturbations on small scales, so the power at large k falls off quickly.

As we can see, by measuring the power spectrum on small scales, we can distinguish the models fairly well.

We look at the matter power spectrum for ΛCDM in more detail in fig. 3.5.

- At large scales, $P(k) \propto k^n$, which is essentially the same as the primordial power spectrum.

¹⁶ The normalization can be done in two ways:

- With δ_H , convenient to relate to the CMB anisotropy power spectrum
- With σ_8 , equal to σ_R with $R = 8h^{-1}\text{Mpc}$ at $z = 0$. This is convenient for large scale structure measurements at low redshifts.

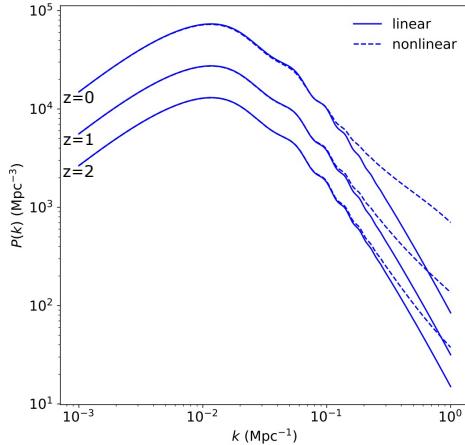


Figure 3.5: Matter power spectrum for Λ CDM. Redshift z is a proxy for time.

- The peak is at $k_{\text{eq}} \approx 0.0073 \text{ Mpc}^{-1} \Omega_m h^2$.
- At intermediate scales, there are baryon acoustic oscillations (BAO). These are at a scale of 0.04 Mpc^{-1} , corresponding to a size of the horizon of 150 Mpc, which is called the sound horizon.
- At small scales,

$$\begin{aligned} P(k) &\propto k^n T(k)^2 \\ &\propto k^{n-4} \ln\left(\frac{k}{k_{\text{eq}}}\right)^2 \end{aligned}$$

At small scales, the amplitude of the perturbations become large, and the linear model no longer holds. As a result, $P(k)$ needs non-linear corrections. We define k_{NL} such that $\Delta(k_{\text{NL}})^2 = 1$. We find that k_{NL} decreases as t increases, so small scales become non-linear first, and large scales later. Small objects form earlier, and large objects form later by merging of smaller ones. This is called a *hierarchical model* for structure formation.

First, galaxies, and then groups were starting to form. Today, galaxy clusters, with a mass of $10^{14} M_\odot$, are starting to form. Superclusters are not collapsing yet, but should do so later.

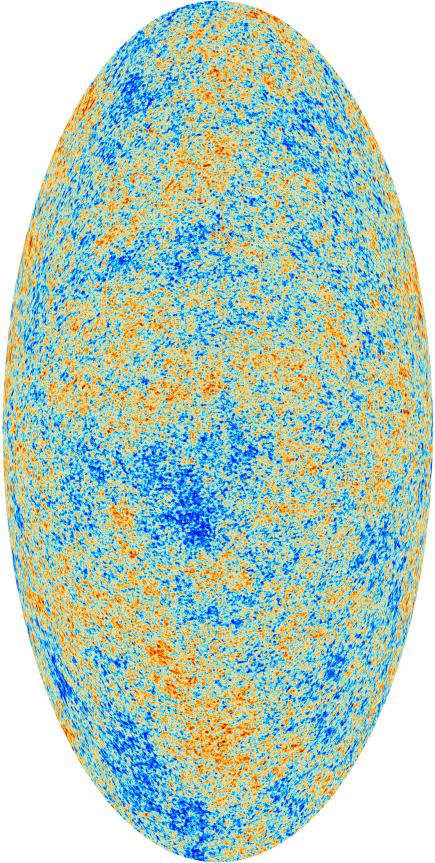


Figure 3.6: The anisotropy of the CMB in galactic coordinates, as measured by the Planck mission. The galactic foreground, which are microwave emission of the Milky Way, have been removed.

3.9 Cosmic Microwave Background anisotropies

The CMB gives us a picture of the universe at last scattering, which happened at $z_* \approx 1100$. While the spectrum is very close to a blackbody spectrum, with a temperature of $T = 2.728(2) \text{ K}$ today, there are small anisotropies with an amplitude of $\Delta T/T \approx 10^{-4}$, as shown in fig. 3.6.

During recombination, protons and electrons combined to hydrogen. We have defined the fraction of free electrons X_e , which decreases during recombination, see ???. The interaction strength between photons and electrons can be characterized by the optical depth,

$$\tau_T(t) = \int_0^t n_e \sigma_T dt,$$

where σ_T is the Thomson scattering cross-section. The optical depth is a measure of opacity for the photons.

For convenience, we define the conformal time τ ,

$$c dt = a d\tau \implies \tau(t) = \int_0^t \frac{c dt'}{a(t')}.$$

We write the temperature as a $T(\mathbf{x}, \hat{\mathbf{q}}, \tau)$, where \mathbf{x} is our position in space, $\hat{\mathbf{q}}$ is the direction in which the photons are travelling, and τ is the conformal time. We define the small temperature deviation θ from the uniform background temperature \bar{T} as

$$T(\mathbf{x}, \hat{\mathbf{q}}, \tau) = \bar{T}(\tau)[1 + \Theta(\mathbf{x}, \hat{\mathbf{q}}, \tau)].$$

First, we decompose the temperature deviation into spherical harmonics,

$$\Theta(\mathbf{x}, \hat{\mathbf{q}}, \tau) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m}(\mathbf{x}, \tau) Y_{\ell m}(\hat{\mathbf{q}}),$$

where

$$a_{\ell m}(\mathbf{x}, \tau) = \int d\Omega \Theta(\mathbf{x}, \hat{\mathbf{q}}, \tau) Y_{\ell m}^*(\hat{\mathbf{q}})$$

are the basis coefficients, and $Y_{\ell m}$ are the spherical harmonics.

Since θ is just one possible outcome of a random quantum process, we again have to fall back to statistical quantities in order to compare measurements with models. We thus define the angular power spectrum $c_\ell = \langle |a_{\ell m}|^2 \rangle$ as a moment of the spherical basis coefficients, where we set $\tau = \tau_0$.

With a Fourier transform, we get $\Theta(\mathbf{x}, \hat{\mathbf{q}}, \tau) \rightarrow \Theta(\mathbf{k}, \hat{\mathbf{q}}, \tau)$. We define $\mu = \hat{\mathbf{q}} \cdot \mathbf{k}$, and

$$\Theta_\ell(k, z) = \frac{1}{(-i)^\ell} \int_{-1}^1 \frac{d\mu}{\tau} \mathcal{P}_\ell(\mu) \Theta(\mathbf{k}, \hat{\mathbf{q}}, \tau),$$

where \mathcal{P}_ℓ are Legendre polynomials. Then

$$c_\ell = \frac{2V}{\pi} \int dk k^2 \langle |\Theta_\ell(\mathbf{k}, \tau_0)|^2 \rangle.$$

We would like to relate this quantity, which we defined at z_0 , to its value at an earlier time z_* . Using the solution to the Einstein-Boltzmann equation, and assuming instantaneous decoupling, we get

$$\begin{aligned} \Theta_\ell(k, \tau_0) &= [\Theta_0(k, \tau_*) + \Psi(k, \tau_*)] j_\ell[k(\tau_0 - \tau_*)] \\ &\quad + 3\Theta_1(k, \tau_*) \left[j_{\ell-1}[k(\tau_0 - \tau_*)] - \frac{(\ell+1)j_\ell[k(\tau_0 - \tau_*)]}{k(\tau_0 - \tau_*)} \right] \\ &\quad + \int_0^{\tau_0} d\tau e^{-\tau_T} [\dot{\Psi}(k, \tau) - \dot{\Phi}(k, \tau)] j_\ell[k(\tau_0 - \tau)], \end{aligned}$$

where j_ℓ are Bessel functions. We now analyse the terms.

The first term is the *Sachs-Wolfe term*, $\Theta_0 + \Psi$. Θ_0 is the monopole, which is the mean temperature at decoupling. Ψ is the gravitational potential at decoupling. This term describes that photons have to climb out of the gravitational potential, which redshifts them. Fluctuations in the potential thus influence the temperature of the photons we see today.

The second term is a function of the dipole Θ_1 , which is related to the bulk velocity of photons at last scattering. This is called the *Doppler term* along the line of sight.

The third term involves $\dot{\Phi}$ and $\dot{\Psi}$, which are time derivatives of gravitational potentials. We saw that, in the matter era, both derivatives are zero. However, they can be non-zero if dark matter and curvature are present. This term is called the *integrated Sachs-Wolfe term* (ISW).

The theoretical model of the angular power spectrum and several measurements are shown in fig. 3.7. We can see that the most powerful mode has a scale of around 1° , which can also be seen in fig. 3.6.

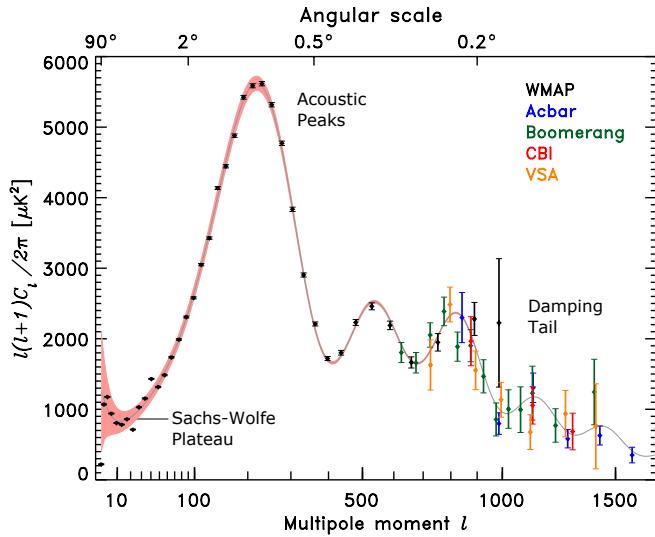


Figure 3.7: The power spectrum of the cosmic microwave background radiation temperature anisotropy in terms of the angular scale (or multipole moment). The measurements and the theoretical model are shown.

Asymptotic limits

We now analyse the different areas in the angular power spectrum.

At large scales, where $k \ll \chi_h^{-1}(z_*)$, we get for the Sachs-Wolfe term

$$C_\ell^{\text{SW}} \propto \delta_H^2 \frac{\Gamma\left(\ell + \frac{n-1}{2}\right) \Gamma\left(\frac{9-n}{2}\right)}{\Gamma\left(\ell + \frac{5-n}{2}\right) \Gamma\left(\frac{3+n}{2}\right)},$$

where δ_H is a normalization factor, and n is the spectral index in the power law. For $n = 1$, we get

$$C_\ell^{\text{SW}} \propto \delta_H^2 \frac{1}{\ell(\ell+1)},$$

so $\ell(\ell+1)C_\ell$ is constant, which is why the vertical axis in the plot was chosen in such a way.¹⁷

At intermediate scales, we see acoustic oscillations. We have already seen this in the ideal fluid equations, where we found baryon-photon oscillations proportional to

$$\exp\left(\frac{ikc_s t}{a}\right) = \exp(ikc_s \tau)$$

at decoupling z_* . The peaks are found at

$$k_m = \frac{m\pi}{L_s(\tau_*)},$$

where $L_s(\tau_*) \approx c_s(\tau_*)\tau_*$ is the sound horizon at decoupling, and $m \in \{1, 2, \dots\}$. Remember that $\Theta = L_{\text{comoving}}/r(\chi)$, where r is the comoving angular diameter distance. The comoving distance which is relevant here is $L_{\text{comoving}} \approx k_m^{-1}$, and $\ell \approx 1/\theta$. We thus find

$$\begin{aligned} \ell_m &\approx k_m r(\chi_*) \\ &= m\pi \frac{r(\chi_*)}{L_s(\chi_*)}. \end{aligned}$$

We find the first peak at $\ell \approx 200$ for the Λ CDM model.

On small scales, we see Silk damping, where oscillations are damped because of imperfect coupling between photons and baryons. This occurs at $\ell > 2000$.

¹⁷ It is not exactly flat for $n = 1$, because the Doppler term and the ISW term have been neglected here.

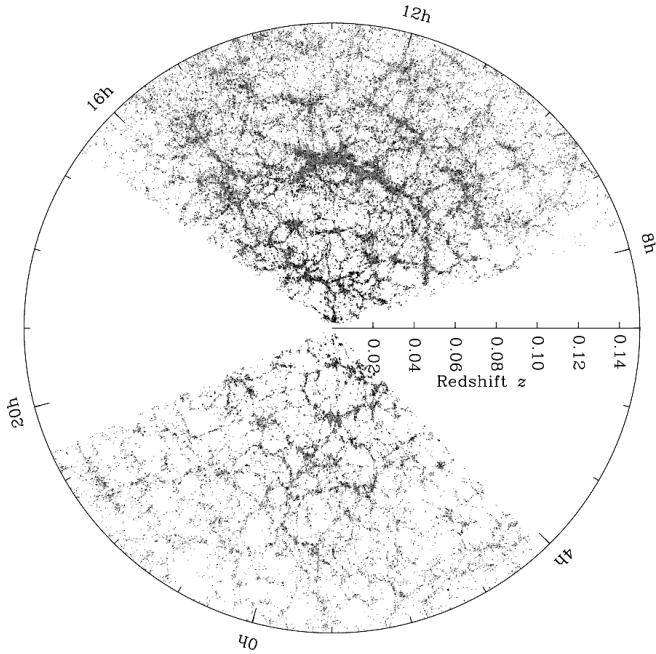


Figure 3.8: A slice of the sky from the SDSS. Some galaxy clusters seem to be stretched in the radial direction, which is caused by the different peculiar velocities of the galaxies. This phenomenon is known as redshift space distortions.

In summary, we see that measurements of the CMB power spectrum can be used to put constraints on cosmological models. The results are shown in ???. Similarly, we can also measure the statistics of the polarization of the CMB.

As an example, we look at the measurement of curvature. The first peak is at a position $\ell_{m=1} \approx k_p r(\chi_*)$. We assume that $L_s(z_*)$ is constant, then we have $\ell_{m=1} \propto r(\chi_*)$. This depends on Ω .

3.10 Galaxy Clustering

We can find density perturbations at low redshift by measuring the clustering of galaxies. Unfortunately, we can only see visible matter, such as galaxies, whereas models predict the total mass distribution, which includes dark matter. To correct for this galaxy bias, we introduce the simple relation

$$\delta_g = b(k, z, \text{galaxy type})\delta,$$

where $\delta_g = (n_g - \bar{n}_g)/\bar{n}_g$ is the galaxy overdensity, $\delta = (\rho - \bar{\rho})/\bar{\rho}$ is the dark matter overdensity, and b is the bias. The bias depends on the galaxy sample, the scale k , and redshift z . Fortunately, on large scales, the bias is mostly scale independent.

There are two kinds of galaxy surveys:

- Imaging surveys, such as POSS/APM, DES, and SDSS, measure the position and shape of galaxies.
- Spectroscopy surveys, such as SDSS, 2DF, and Wiggle-z, measure the position and redshift of galaxies. A slice of the galaxy map measured by SDSS is shown in fig. 3.8.

In fig. 3.9, the power spectrum found by the SDSS is shown. From the residual plot, we see that even the baryon acoustic oscillations are well resolved. These are fixed in comoving coordinates, so they can function as a standard ruler.

We can measure the two-dimensional clustering with an imaging survey such as APM, as shown in fig. 3.10. This can be used to rule out HDM or bayron-only models. The

Figure 3.9: The SDSS power spectrum.

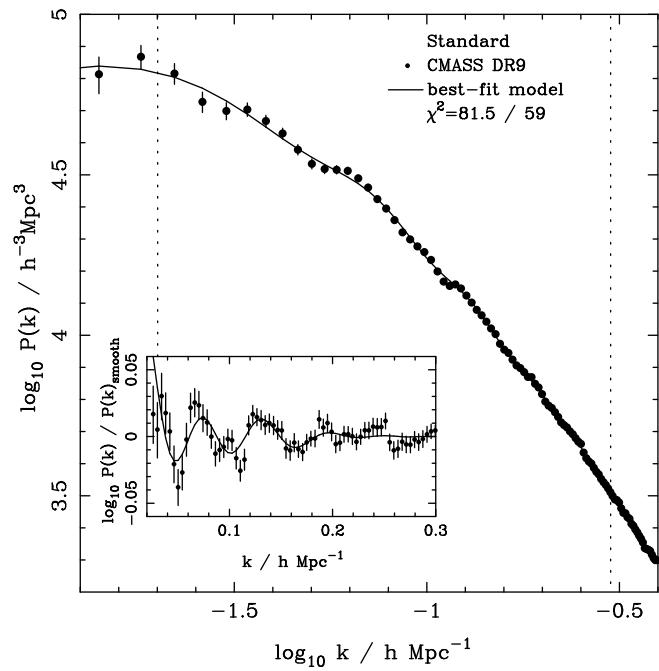
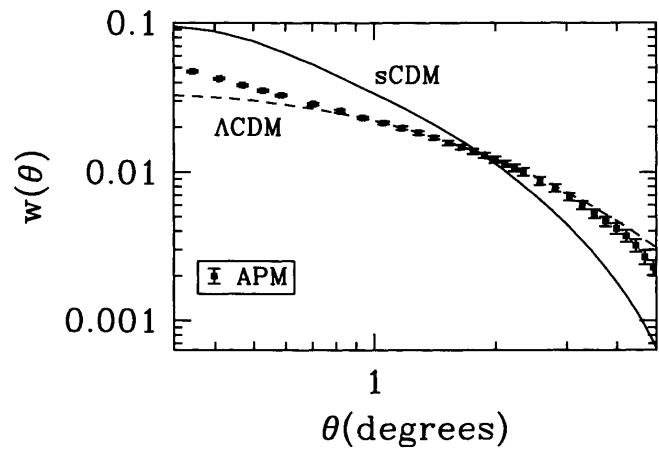


Figure 3.10: Results of the APM survey.



current measurements confirm the Λ CDM model, even though there are still small deviations, which come from the scale dependence of the bias factor.

For the further interpretation of galaxy clustering measurements, we need to study the physics of galaxy formation and non-linear structure formation.

4 Nonlinear Perturbations

4.1 Approaches for non-linear structure formation

Remember that at late times, perturbations become non-linear, and that small-scale perturbations become non-linear first. Perturbations collapse to form dark halos, filaments, and pancakes. This structure is called the cosmic web, which is a signature of non-linear gravitational instability in an expanding universe.

Studying the evolution of non-linear perturbations is very difficult, and no exact analytic solutions have yet been found in the fully non-linear regime. There are different approaches to tackle the problem:

- Numerical simulations
- Higher order perturbation theory is only a valid approach for mildly non-linear perturbations and scales. On small scales, this does not work.
- The Halo model, which is a simplified model, is very successful in describing the formation, evolution, and statistics of collapsed structures.

We will study mainly the Halo model. We start with a model that only contains dark matter, and later introduce baryons.

4.2 Spherical Collapse

We consider a spherical overdensity with in an otherwise homogeneous expanding universe. Outside the sphere, the density is $\bar{\rho}$, while inside it is $\rho = \bar{\rho}(1 + \delta)$.

We make a few simplifying assumptions:

- The universe is flat and matter-dominated, such that $\Omega_0 = \Omega_{m,0} = 1$.
- There is only collisionless dark matter.

We now look at a thin shell with radius r inside the top-hat sphere, and write down its equation of motion. Newton's second law states

$$\frac{d^2r}{dt^2} = -\frac{GM}{r^2},$$

where M is the mass inside the radius r ,

$$M = M(r) = \frac{4\pi r^3}{3} \bar{\rho}(1 + \delta).$$

Note that $M(r)$ is constant as long as the different shells don't cross each other, which we assume for now. The energy per unit mass of the shell, also called its specific energy ϵ , is the sum of a kinetic and a potential term:

$$\epsilon = \frac{1}{2} \left(\frac{dr}{dt} \right)^2 - \frac{GM}{r},$$

which is constant because of conservation of energy. If $\epsilon \geq 0$, the shell expands forever, and if $\epsilon < 0$, the shell might expand for a while, but will eventually contract and collapse. We will focus on the latter case.

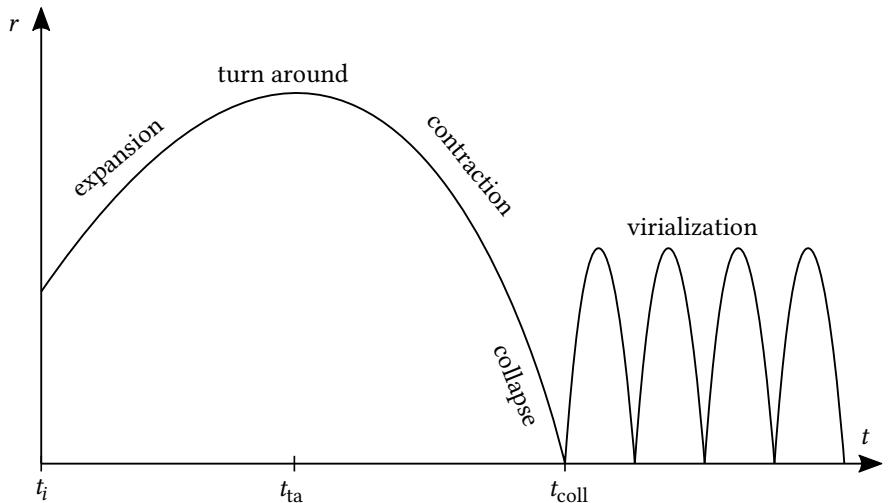


Figure 4.1: The evolution of the shell radius of a spherical overdensity, with negative total energy. The shell expands, reaches a maximum radius, and collapses, which is described by the analytical solution derived in the text. Shortly before collapse, the solution becomes invalid as shells start crossing.

The solution of the equation of motion for $\epsilon < 0$ is

$$r = A(1 - \cos \theta) \quad t = B(\theta - \sin \theta) \quad A = \frac{GM}{2|\epsilon|} \quad B = \frac{GM}{(2|\epsilon|)^{2/3}}$$

The solution $r(t)$ is plotted in fig. 4.1. The shell expands, reaches a maximum radius $r_{\text{ta}} = 2A$ at $t_{\text{ta}} = \pi B$, and collapses at $t_{\text{coll}} = 2t_{\text{ta}}$.¹⁸ However, the model will be invalid shortly before t_{coll} , since shells start crossing and virialize, which will be discussed later.

We would also like to derive an expression for $\rho(t)$ inside the sphere, and find the critical density which is required for collapse. To do this, we use a few more of our assumptions and perform some approximations.

At an early time t_i , we define $r = r_i$ and $v = v_i$. We assume that at t_i , the shell follows the Hubble flow:

$$\begin{aligned} v_i &= \frac{d(ax_i)}{dt} = \dot{a}x_i && \text{no peculiar velocity, and } r = ax \\ &= Hr_i. && H = \dot{a}/a \end{aligned}$$

Plugging in at $t = t_i$, we find

$$M = \frac{4\pi r_i^3}{3} \bar{\rho}(t_i)(1 + \delta_i), \quad \epsilon = \frac{v_i^2}{2} - \frac{GM}{r_i}.$$

Since we assumed a matter-dominated universe, $a \propto t^{2/3}$, and $H = 2/(3t)$. We also assumed a flat universe, so

$$\bar{\rho} = \rho_{\text{crit}} = \frac{1}{6\pi G t^2}.$$

We can plug these expressions into the definitions of A and B to get

$$A = \frac{3}{10} \frac{r_i}{\delta_i}, \quad B = \frac{9}{20} \frac{t_i}{\delta_i}.$$

The density inside the top hat sphere is then

$$\rho = \frac{M}{\frac{4\pi}{3} r^3} = \frac{3M}{4\pi A^3} (1 - \cos \theta)^{-3} \quad \text{because } r = A(1 - \cos \theta),$$

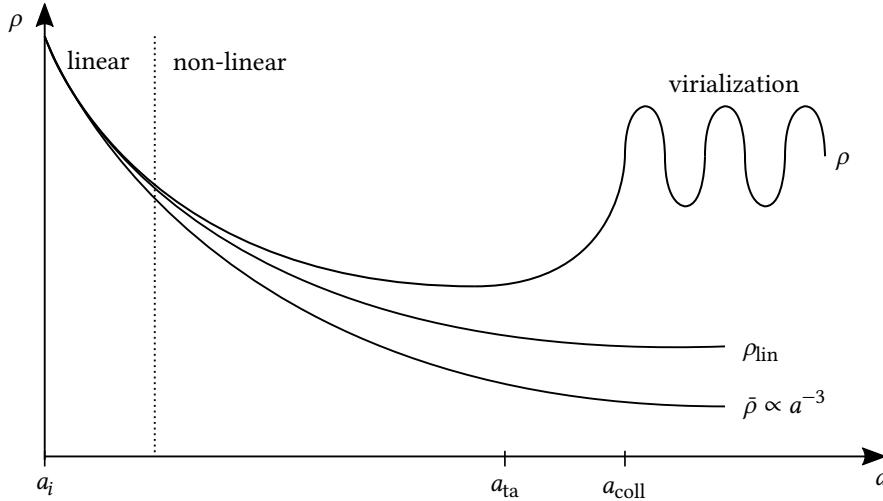


Figure 4.2: The evolution of the density of a spherical overdensity with negative total energy. The exact model is ρ , which agrees well with the linearised approximation ρ_{lin} at early times, but only until the turn-around time. For reference, the average density $\bar{\rho}$ of the universe outside the sphere is also shown.

while outside it is

$$\bar{\rho} = \frac{1}{6\pi G t^2} = \frac{1}{6\pi G B^2} (\theta - \sin \theta)^{-2} \quad \text{because } t = B(\theta - \sin \theta).$$

It follows that

$$1 + \delta = \frac{\rho}{\bar{\rho}} = \frac{9}{2} \frac{(\theta - \sin \theta)^2}{(1 - \cos \theta)^3}.$$

Since we want to find $\rho(t)$ instead of $\rho(\theta)$, we have a little more work to do. An analytical solution is nowhere to be seen, so we try to find at least an approximation. At early times, $t \ll t_{\text{ta}}$, we perform a Taylor expansion in t and θ to get

$$\delta \approx \frac{3}{20} \theta^2 \quad t \approx \frac{B}{6} \theta^3.$$

These two expressions can be combined to find

$$\delta = \frac{3}{20} \left(\frac{6t}{B} \right)^{2/3} = \frac{3}{20} (6\pi)^{2/3} \left(\frac{t}{t_{\text{ta}}} \right)^{2/3},$$

which is fortunately¹⁹ in agreement with what we found in linear perturbation theory for a matter-dominated universe, where we derived $D(z) \propto a \propto t^{2/3}$. Note that this result is only valid at early times, and the actual value of δ might be quite different at later times. To make things clearer, we define

$$\delta_{\text{lin}} = \frac{3}{20} (6\pi)^{2/3} \left(\frac{t}{t_{\text{ta}}} \right)^{2/3},$$

and expect $\delta = \delta_{\text{lin}}$ at early times, but not at late times. We can find ρ from δ , and ρ_{lin} from δ_{lin} , which is a more intuitive value. The relationship between ρ , ρ_{lin} , and the density outside the sphere, $\bar{\rho}$, is shown in fig. 4.2.

The overdensity from the linearized model and the exact model at different times are compared in table 4.1. We notice that the overdensities are larger than one, which in hindsight clarifies why we require non-linear perturbation theory. In the previous chapter, we always assumed $\delta \ll 1$, which is obviously not valid any more.

Depending on the cosmological models, the overdensity δ_c required for collapse is different. We have so far worked with a simplified CDM model (sCDM), where we assumed that the universe is flat and dominated by collisionless dark matter. For other

¹⁹ Since we assumed early times, linear perturbation theory should still be valid, so comparing our new result to what we found in the previous chapter gives us some reassurance that everything still works correctly.

model	t_{ta}	t_{coll}
δ	4.55	∞
δ_{lin}	1.062	1.686

Table 4.1: Overdensities in the linearized and the exact model at different times.

cosmologies, such as OCDM and Λ CDM, it turns out that multiplying the sCDM result with an additional factor yields a good approximation:

$$\delta_c \approx \begin{cases} 1.686 & \text{for sCDM} \\ 1.686[\Omega_m(t_{\text{coll}})]^{0.0185} & \text{for OCDM} \\ 1.686[\Omega_m(t_{\text{coll}})]^{0.0055} & \text{for } \Lambda\text{CDM} \end{cases}$$

4.3 Collisionless dynamics

4.3.1 Time scales

In the previous chapter, we assumed a collisionless medium. To find out when this approximation is valid, we look at the collision time-scales in thermodynamic equilibrium.

We consider a system with N self-gravitating particles and diameter r . Each particle i has a velocity \mathbf{v}_i and a mass m . The average velocity is $\mathbf{v} = \langle \mathbf{v}_i \rangle$. There are several relevant time scales.

- The *crossing time scale* t_{cross} is the time required for a typical particle to cross the system:

$$t_{\text{cross}} = \frac{r}{v}$$

- The *time scale for close encounters* t_{close} is the average time between close encounters. To calculate it, we look at a gravitational two-body interaction of two close particles. We define an encounter as a close encounter if the change of the speed of the particle due to the interaction is similar to the initial speed,

$$\delta_v := |\mathbf{v}_{\text{before}} - \mathbf{v}_{\text{after}}| \approx \mathbf{v}_{\text{before}}.$$

Then we define t_{close} as the average time between close encounters,²⁰

$$t_{\text{close}} = N t_{\text{cross}}.$$

- The *relaxation time scale* t_{relax} is the time required for the particle velocity distribution to relax to a new distribution. Consider a cumulative change in velocity Δv^2 over the motion of a particle. We define t_{relax} as the time scale for which Δv^2 is comparable to v^2 . One can show that

$$t_{\text{relax}} := \frac{N}{10 \ln N} t_{\text{cross}}.$$

In typical astrophysical systems with $N \gg 100$, such as galaxies and galaxy clusters, we usually find

$$t_{\text{close}} \gg t_{\text{relax}} \gg t_H \gg t_{\text{cross}},$$

where $t_H = H^{-1} \approx 10^{10}$ yr is the Hubble time. Collisions are thus so rare that we can treat many systems as effectively collisionless.

For example, in a galaxy we have $r \approx 10$ kpc, $v \approx 200$ km s⁻¹, and $N \approx 10^{10}$, which yields

$$\begin{aligned} t_{\text{close}} &= 5 \cdot 10^{17} \text{ yr}, \\ t_{\text{relax}} &= 2 \cdot 10^{15} \text{ yr}, \\ t_H &= 10^{10} \text{ yr}, \\ t_{\text{cross}} &= 5 \cdot 10^7 \text{ yr}. \end{aligned}$$

Gravitational collisions are apparently very inefficient and can be neglected for these systems, so we can use the collisionless Boltzmann equation. Instead of analysing two-body forces, we can use a smooth mean gravitational potential.

²⁰ It seems counter-intuitive that the time between encounters increases with particle size. This has to do with our assumption that the system is in thermodynamic equilibrium, where the velocity goes down for constant m and r .

4.3.2 Basic Dynamics

We now look at the collisionless Boltzmann equation (CBE), with a phase-space distribution function $f(\mathbf{x}, \mathbf{v}, t)$. As we've seen before, the moments of f are

- $n(\mathbf{x}, t) = \int f(\mathbf{x}, \mathbf{v}, t) d^3 v = \rho(\mathbf{x}, t)/m$
- $\langle Q \rangle(\mathbf{x}, t) = \frac{1}{n} \int d^3 v Q f(\mathbf{x}, \mathbf{v}, t)$

The evolution of the distribution function is given by the CBE,

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \sum_i \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} + \sum_i \frac{\partial f}{\partial v_i} \frac{dv_i}{dt} = 0,$$

which can be rewritten with $v_i = dx_i/dt$ and $dv_i/dt = -\partial\Phi/\partial x_i$ as

$$\frac{\partial f}{\partial t} + \sum_i v_i \frac{\partial f}{\partial x_i} - \sum_i \frac{\partial \Phi}{\partial x_i} \frac{\partial f}{\partial v_i} = 0,$$

where the gravitational potential can be calculated from the Poisson equation, $\nabla^2 \Phi = 4\pi G \rho$.

4.3.3 Equations of motion

We multiply the CBE by Q and integrate over \mathbf{v} to get

$$\frac{\partial}{\partial t} [n \langle Q \rangle] + \sum_i \frac{\partial}{\partial x_i} [n \langle Q v_i \rangle] + n \sum_i \frac{\partial \Phi}{\partial x_i} \left\langle \frac{\partial Q}{\partial v_i} \right\rangle = 0,$$

where the last term follows from partial integration and from $f = 0$ for $v^2 \rightarrow \infty$. It turns out that we can completely recover something that looks almost identical to the ideal fluid equations, just by substituting cleverly for Q .²¹

Continuity equation If we set $Q = 1$, we get

$$\frac{\partial n}{\partial t} + \sum_i \frac{\partial}{\partial x_i} [n \langle v_i \rangle] = 0,$$

which we recognize as the continuity equation of an ideal fluid, with n instead of ρ .

²¹ We have already seen this in section 3.2, but we didn't prove the result.

Euler equation For $Q = v_j$, we obtain

$$\frac{\partial}{\partial t} [n \langle v_j \rangle] + \sum_i \frac{\partial}{\partial x_i} [n \langle v_i v_j \rangle] + n \frac{\partial \Phi}{\partial x_j} = 0. \quad (4.1)$$

We remember the definition²²

$$\sigma_{ij}^2 = \langle v_i v_j \rangle - \langle v_i \rangle \langle v_j \rangle$$

to get

$$\begin{aligned} 0 &= \frac{\partial n}{\partial t} \langle v_j \rangle + n \frac{\partial \langle v_j \rangle}{\partial t} + \sum_i \frac{\partial}{\partial x_i} [n \sigma_{ij}^2] + \sum_i \frac{\partial}{\partial x_i} [n \langle v_i \rangle \langle v_j \rangle] + n \frac{\partial \Phi}{\partial x_j} \\ &= \frac{\partial n}{\partial t} \langle v_j \rangle + n \frac{\partial \langle v_j \rangle}{\partial t} + n \sum_i \frac{\partial}{\partial x_i} [n \sigma_{ij}^2] + \sum_i \frac{\partial}{\partial x_i} [n \langle v_i \rangle] \langle v_j \rangle + \sum_i n \langle v_i \rangle \frac{\partial}{\partial x_i} \langle v_j \rangle + n \frac{\partial \Phi}{\partial x_j} \\ &= n \frac{\partial \langle v_j \rangle}{\partial t} + n \sum_i \frac{\partial}{\partial x_i} [n \sigma_{ij}^2] + \sum_i n \langle v_i \rangle \frac{\partial}{\partial x_i} \langle v_j \rangle + n \frac{\partial \Phi}{\partial x_j} \quad (\text{continuity equation}) \\ &= \frac{\partial \langle v_j \rangle}{\partial t} + \sum_i \langle v_i \rangle \frac{\partial \langle v_j \rangle}{\partial x_i} + \frac{1}{n} \sum_i \frac{\partial (n \sigma_{ij}^2)}{\partial x_i} + \frac{\partial \Phi}{\partial x_j} \end{aligned}$$

²² We have previously defined the stress tensor as $\rho \sigma_{ij}^2$, but we will see the quantity $n \sigma_{ij}^2$ appear now. It does not matter much, because we can convert with $n = \rho/m$.

This is the Euler equation, but with $n\sigma_{ij}^2$ instead of p , and n instead of ρ . We already stated this result in the previous chapter, and now we have proven it. The equation in this form is called the *Jeans equation*.

Discussion If we take the Jeans equation, the continuity equation, and the Poisson equation, we have a total of five equations in the 11 unknowns ρ , v_i , σ_{ij} (6, because it is symmetric) and ρ . Evidently, we have too many unknowns. We could try to take higher moments of the CBE, but this only adds more unknowns in the end. In order to close the system, we need more assumptions.²³

²³ We didn't have this problem for an ideal fluid, because the equation of state was used.

4.4 Virial theorem

4.4.1 Derivation

²⁴ The Virial theorem is also valid for a self-gravitating ideal fluid, since its derivation relies on the equations from section 4.3.3, which are basically the same as the ideal fluid equations.

$$\frac{1}{2} \frac{d^2 I_{jk}}{dt^2} = 2K_{jk} + W_{jk} + \Sigma_{jk},$$

where we have defined

$I_{jk} = \int \rho x_j x_k d^3x$	moment of inertia tensor
$K_{jk} = \frac{1}{2} \int \rho \langle v_j v_k \rangle d^3x$	kinetic energy tensor
$W_{jk} = - \int \rho x_k \frac{\partial \Phi}{\partial x_j} d^3x$	Chandrasekhar potential energy tensor
$\Sigma_{jk} = - \sum_i \int x_k \rho \langle v_j v_i \rangle dS_i$	surface pressure term

We start the derivation with the Euler equation (4.1),

$$0 = \frac{\partial}{\partial t} [n \langle v_j \rangle] + \sum_i \frac{\partial}{\partial x_i} [n \langle v_i v_j \rangle] + n \frac{\partial \Phi}{\partial x_j},$$

which we multiply by mx_k and integrate over space to get

$$\int x_k \frac{\partial [\rho \langle v_j \rangle]}{\partial t} d^3x = - \sum_i \int x_k \frac{\partial [\rho \langle v_i v_j \rangle]}{\partial x_i} d^3x - \int \rho x_k \frac{\partial \Phi}{\partial x_j} d^3x.$$

The first term on the right-hand side can be rewritten with the divergence theorem as

$$- \sum_i \int x_k \frac{\partial [\rho \langle v_i v_j \rangle]}{\partial x_i} d^3x = 2K_{kj} + \Sigma_{kj}.$$

The last term on the right-hand side is the Chandrasekhar potential energy tensor. With the continuity equation, one can show that the left-hand side is equal to the moment of inertia tensor. Putting everything together, we get

$$\frac{1}{2} \frac{d^2 I_{jk}}{dt^2} = 2K_{jk} + W_{jk} + \Sigma_{jk},$$

Scalar Virial theorem. From the tensor Virial theorem, we can derive the scalar Virial theorem,

$$\frac{1}{2} \frac{d^2 I}{dt^2} = 2K + W + \Sigma,$$

where $X = \text{tr}(X_{ij})$. One can show that K and W are the kinetic and potential energy of the system.

Static, localized system. For a system in a steady state, meaning that the distribution function does not depend on time, but only on \mathbf{x} and \mathbf{v} , the derivative of the moment of inertia tensor is zero. If we also assume that the system is isolated, the surface pressure term is zero. This yields

$$2K + W = 0, \quad \text{or equivalently,} \quad E = -K = \frac{W}{2},$$

where E is the total energy of the system.

4.4.2 Applications

Mass estimation. We consider a system composed of many collisionless particles²⁵ with total mass M , average particle velocity v , and diameter R . The kinetic and potential energy can be estimated as

$$K \approx \frac{Mv^2}{2}, \quad W \approx -\frac{GM^2}{R},$$

which can be plugged into the Virial theorem to give

$$v^2 \approx \frac{GM}{R}.$$

Since the radius can be estimated from images, and the average velocity can be estimated from measurements of the Doppler shift, this method is useful for estimating the mass of a system, which is otherwise hard to access experimentally.

²⁵ We have seen in section 4.3.1 that this is an excellent approximation for many systems, such as galaxies and clusters.

Spherical collapse model. The Virial theorem can be applied to the spherical collapse model. Let $E = K + W$ be the conserved total energy of the sphere. At t_{ta} , the shell is stationary, so the kinetic energy is zero, and $E = W_{\text{ta}}$, which is the gravitational potential energy of a solid sphere with radius r_{ta} :

$$E = -\frac{3}{5} \frac{GM^2}{r_{\text{ta}}}.$$

At $t = t_{\text{coll}}$, which is also called t_{vir} , the energy is

$$\begin{aligned} E &= \frac{W_{\text{vir}}}{2} && \text{Virial theorem} \\ &\approx \frac{1}{2} \left(-\frac{3}{5} \frac{GM^2}{r_{\text{vir}}} \right). \end{aligned}$$

Because the energy is conserved, the two results can be combined to yield

$$r_{\text{vir}} = \frac{r_{\text{ta}}}{2}.$$

We can also calculate the mean overdensity $1 + \delta(t_{\text{vir}})$ within the radius r_{vir} , which is commonly written as $1 + \Delta$:

$$\begin{aligned}
 1 + \Delta &= \frac{\rho(t_{\text{vir}})}{\bar{\rho}(t_{\text{vir}})} \\
 &= \frac{\rho(t_{\text{ta}})}{\bar{\rho}(t_{\text{vir}})} \left(\frac{r_{\text{ta}}}{r_{\text{vir}}} \right)^3 \\
 &= \frac{\rho(t_{\text{ta}})}{\bar{\rho}(t_{\text{ta}})} \frac{\bar{\rho}(t_{\text{ta}})}{\bar{\rho}(t_{\text{vir}})} \left(\frac{r_{\text{ta}}}{r_{\text{vir}}} \right)^3 \\
 &= \frac{9\pi^2}{16} 2^2 2^3 \quad \text{see section 4.2, } a \propto t^{2/3}, \bar{\rho} \propto a^{-3} \\
 &\approx 178
 \end{aligned}$$

4.5 Steady state solutions

We want to find solutions of the CBE which are steady states, meaning we want f to be time-independent. The CBE for a steady state is

$$0 = \sum_i \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} + \sum_i \frac{\partial f}{\partial v_i} \frac{dv_i}{dt},$$

because the term $\frac{\partial f}{\partial t} = 0$ for a steady state.

4.5.1 Integrals of motion

Integrals of motion are functions $I(\mathbf{x}, \mathbf{v})$ that are constant along particle trajectories. Examples of such functions are energy and angular momentum. For integrals of motion,

$$\begin{aligned}
 0 &= \frac{dI}{dt} \\
 &= \sum_i \frac{\partial I}{\partial x_i} \frac{dx_i}{dt} + \sum_i \frac{\partial I}{\partial v_i} \frac{dv_i}{dt}.
 \end{aligned}$$

Thus, integrals of motion are solutions of the steady-state CBE. Conversely, a distribution function f which is a function of integrals of motion, written as

$$f(\mathbf{x}, \mathbf{v}) = f[I_1(\mathbf{x}, \mathbf{v}), \dots, I_N(\mathbf{x}, \mathbf{v})],$$

is necessarily a steady-state solution of the CBE.

It follows that all solutions of the steady state CBE depend on \mathbf{x} and \mathbf{v} through integrals of motion. This result is called *Jean's theorem*.

4.5.2 Spherical models

We now analyse spherically symmetric solutions, where ρ only depends on a radius r , and where f only depends on the specific energy of a particle. First, we derive some general properties, and then we consider particular models.

General spherical model. The specific energy of a particle is

$$E = \frac{1}{2}v^2 + \Phi(\mathbf{x}).$$

We can relate $f(E)$ to the density profile $\rho(r)$ of the system. We define

$$\Psi = -\Phi + \Phi_0, \quad \epsilon = -E + \Phi_0,$$

where we demand $\Phi \rightarrow \Phi_0$ at the boundary of the system. One can show that

$$\rho(\Psi) = 4\pi \int_0^\Psi f(\epsilon) \sqrt{2(\Psi - E)} d\epsilon.$$

To find $\rho(r)$, we solve the Poisson equation for a spherically symmetric system,

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\Psi}{dr} \right) = -4\pi G \rho(r).$$

Conversely,

$$f(\epsilon) = \frac{1}{\pi^2 \sqrt{8}} \frac{d}{d\epsilon} \int_0^\epsilon \frac{d\rho}{d\Psi} \frac{d\Psi}{\sqrt{\epsilon - \Psi}}$$

Isothermal sphere.

$$f(\epsilon) = \frac{\rho_0/m}{(2\pi\sigma^2)^{3/2}} \exp(-\epsilon/\sigma^2),$$

where ρ_0 and σ are constants. Then

$$\begin{aligned} \rho(r) &= \rho_0 \exp\left(\frac{\Psi}{\sigma^2}\right) \\ &= \begin{cases} \rho_0 \left[1 + \left(\frac{r}{r_0}\right)^2\right]^{-3/2} & \text{if } r \leq 2r_0, \\ \frac{\sigma^2}{2\pi Gr^2} & \text{if } r \geq 10r_0, \end{cases} \end{aligned}$$

where

$$r_0 = \frac{3\sigma}{\sqrt{4\pi G\rho_0}},$$

which is called the King radius.

Notes:

- The expected velocity is independent of r , which is the reason why the sphere is called isothermal:

$$\langle v^2 \rangle = \frac{1}{\rho} \int d^3v v^2 f(E) = 3\sigma^2.$$

- In the limit where

$$\rho(r) = \frac{\sigma^2}{2\pi Gr^2},$$

the model is called a singular isothermal sphere, because $\rho \rightarrow \infty$ for $r \rightarrow 0$.

- The total mass of an isothermal sphere is infinite, because

$$\begin{aligned} M(< r) &= 4\pi \int_0^r dr' r'^2 \rho(r') \\ &\propto r^3 \frac{1}{r^2} \propto r, \end{aligned}$$

which goes to ∞ as $r \rightarrow \infty$. In practice, the model has to be cut off at some radius.

King Model.

$$f(\epsilon) = \begin{cases} \frac{\rho_0/m}{(2\pi\sigma^2)^{3/2}} \left[\exp\left(\frac{\epsilon}{\sigma^2}\right) - 1 \right] & \text{if } \epsilon > 0 \\ 0 & \text{if } \epsilon \leq 0 \end{cases}$$

$\Psi(r)$ and $\rho(r)$ can be computed numerically. We can find $\rho(r) = 0$ for $r > r_t$, where r_t is called the tidal radius. The mass in this model is finite. The model is a good fit for bright elliptical galaxies.

(α, β, γ)	Name
(1, 3, 1)	NFW
(1, 4, γ)	Dehnen
(1, 4, 1)	Hernquist
(1, 4, 2)	Jaffe
(2, 2, 0)	Modified isothermal sphere
(2, 3, 0)	Modified Hubble
(2, 4, 0)	Perfect sphere
(2, 5, 0)	Plummer sphere

Table 4.2: Several power law models.

Double power law models.

$$\rho(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \left[1 + \left(\frac{r}{r_0}\right)^\alpha\right]^{(\gamma-\beta)/\alpha},$$

where r_0 , α , β , and γ are constants. One can show that

$$\rho(r) = \begin{cases} r - \gamma & \text{for } r \ll r_0, \\ r - \beta & \text{for } r \gg r_0. \end{cases}$$

Several models are shown in table 4.2. Haloes and elliptical galaxies are often described by such models. The NFW model is used for dark matter haloes.

4.6 Collisionless relaxation

There are several questions:

- How does a self-gravitating collisionless system achieve a steady/relaxed state?
- What is the nature of this relaxed state?

Relaxation is the process by which a system reaches thermodynamic equilibrium. In a gas, collisions are very efficient, and quickly drive the gas to thermodynamic equilibrium. For self-gravitating collisionless systems, the situation is much more complicated, because collisions are very inefficient, and they do not drive the system to a relaxed state. Still, many relaxed systems are observed in astrophysics, such as many galaxies, galaxy clusters, and globular clusters. If not collisions, then what has driven the system to a relaxed state?

4.6.1 Processes

There are several processes which are responsible for relaxation.

Phase mixing. Even for perfectly regular orbits, the trajectories in phase space diverge slowly in time, because the fundamental frequencies of different orbits are not the same. Particles in nearby phase-space regions with slightly different velocities can gradually get out of phase over time. Their orbits get sheared and eventually fill the whole phase space volume, which leads to a new steady state distribution.

Chaotic mixing. For chaotic orbits, the trajectories diverge exponentially in time. Chaotic mixing is more efficient at large scales, so this process mostly relaxes the coarse-grained distribution function.

Violent relaxation. If the (smoothed) potential of the system changes quickly, such as during a merger or mass loss, the velocity distribution of a system can be broadened. The time scale for this process is small, which is why it is called violent.

Landau damping. The interaction between perturbation waves and particles can transfer energy into the random motion of particles. This is similar to free streaming.

4.6.2 End states

Can we describe the end state of relaxation with Statistical Mechanics? Is it a thermal equilibrium?

In a very broad sense, thermal equilibrium is a state which maximizes the entropy, which is also the most probable state. Is this applicable here? No, it turns out there is no satisfactory definition of entropy for our systems, and it is not clear whether the relaxed states we introduced previously are the most probable states. As a result, the end state of relaxation cannot be described as thermal equilibrium. In practice, we have to use numerical simulations to characterize the relaxed state.

Furthermore, many astrophysical systems cannot be treated as being isolated, since interactions between galaxies or even galaxy mergers can occur. This would complicate any attempt at thermodynamic equilibrium analysis even further.

We can, however, describe the properties of these systems statistically using simplified models, and compare them to simulations.

4.7 Halo Mass Function

The goal of this section is to predict the number density of dark matter haloes which have formed through gravitational collapse as a function of time.

We consider the linear overdensity field $\delta(\mathbf{x}, t)$, where we have dropped the subscript “lin” from previous sections for easier notation. In the linear regime,

$$\delta(\mathbf{x}, t) = \delta_0(\mathbf{x})D(t),$$

where $\delta_0(\mathbf{x}) = \delta(\mathbf{x}, t_0)$ is linearly extrapolated to today, and $D(t)$ is the linear growth factor. We adopt a slightly different notation for $D(t)$, normalizing it such that $D(t_0) = 1$. We define a smoothed field by convolving with a kernel,

$$\delta_s(\mathbf{x}, t) = \int d^3y W(\mathbf{x} + \mathbf{y}; R)\delta(\mathbf{x}', t),$$

where $W(\mathbf{x}; R)$ is a window function with a characteristic radius R , such as a top hat. To a top hat window function, we can associate a characteristic mass,

$$M = \frac{4\pi}{3}R^3\bar{\rho}.$$

4.7.1 Probability of collapse

For the spherical collapse model, we found that for regions with a smooth density field satisfying

$$\delta_s(\mathbf{x}, R, t) > \delta_c \approx 1.69, \quad (\text{number for matter dominated universe})$$

the region will have collapsed to form a virialized object. This is what we call a dark matter halo. We rewrite this using the linear growth factor. Because $\delta_s(\mathbf{x}, R, t) = \delta_S(\mathbf{x}, R, t_0)D(t)$, we have equivalently

$$\delta_s(\mathbf{x}, R, t_0) > \delta_c(t),$$

where we defined $\delta_c(t) = \delta_c/D(t)$.

We now calculate the probability of collapse. Since δ is a Gaussian field (δ is linear with Gaussian initial conditions in our standard model of cosmology), the smoothed field is also Gaussian, since the smoothing is a linear process. The probability that $\delta_s > \delta_c(t)$ can be written as

$$\text{Prob}(\delta_s > \delta_c(t)) = \int_{\delta_c(t)}^{\infty} d\delta_s \text{Prob}(\delta_s),$$

where $\delta_s = \delta_s(t_0)$. Since the field is Gaussian,

$$\text{Prob}(\delta_s) \propto \frac{\exp\left[\frac{-\delta_s^2}{2\sigma^2(M)}\right]}{\sqrt{2\pi}\sigma(M)},$$

where $\sigma(M)$ is the variance of the smoothed field,

$$\begin{aligned} \sigma^2(M) &= \langle \delta_s^2(\mathbf{x}, R, t_0) \rangle \\ &= \frac{1}{2\pi^2} \int_0^{\infty} dk \tilde{W}(k, R) P(k), \end{aligned}$$

²⁶ As a reminder, \tilde{W} is the Fourier transform of the window function, and $P(k) = P(k, t_0)$ is the linear matter-power spectrum today

$$\text{Prob}(\delta_s > \delta_c(t)) = \frac{1}{2} \text{erf}\left[\frac{\delta_c(t)}{\sqrt{2}\sigma(M)}\right].$$

4.7.2 Mass fraction

We use the PS-ansatz by PRESS and SCHECHTER (1974). The mass fraction of collapsed objects with mass greater than M is

$$F(> M) = 2 \text{Prob}(\delta_s > \delta_c(t)).$$

The factor 2 was found with the following argument. In the limit where $M \rightarrow 0$, we see that $\sigma(M) \rightarrow \infty$ for CDM models. As a result,

$$\text{Prob}(\delta_s > \delta_c(t)) \rightarrow \frac{1}{2}.$$

With our ansatz, the mass fraction then goes to 1. That means that all the mass in the universe is in a collapsed object of some mass. Without the factor of 2, only half the mass in the universe would be inside this collapsed object.

The factor in the PS-ansatz can be justified in a more solid way by using a formalism called *excursion sets*, which is also called the Extended Press-Schechter (EPS) formalism.

We want to find the number density of collapsed objects with a mass between M and $M + dM$,

$$n(M, t) = \frac{\bar{\rho}}{M} \frac{\partial F(> M)}{\partial M} dM$$

The factor $\bar{\rho}/M$ is the number of *all* objects per unit volume, if all objects had the mass M . The next factor is the mass fraction of collapsed objects with mass between M and $M + dM$. We get

$$n(M, t) dM = 2 \frac{\bar{\rho}}{M} \frac{\partial \text{Prob}(\delta_s > \delta_c(t))}{\partial \sigma} \left| \frac{d\sigma}{dM} \right| dM.$$

We plug in

$$\begin{aligned} \frac{\partial \text{Prob}(\delta_s > \delta_c(t))}{\partial \sigma} &= \frac{\partial}{\partial \sigma} \left[\frac{1}{2} \text{erfc} \left(\frac{\delta_c}{2\sigma} \right) \right] \\ &= \frac{1}{\sqrt{2\pi}} \frac{\delta_c}{\sigma^2} \exp \left(-\frac{\delta_c^2}{2\sigma^2} \right) \end{aligned}$$

and get²⁷

$$\begin{aligned} n(M, t) dM &= 2 \frac{\bar{\rho}}{M} \frac{1}{\sqrt{2\pi}} \frac{\delta_c}{\sigma^2} \exp \left(-\frac{\delta_c}{2\sigma^2} \right) \left| \frac{d\sigma}{dM} \right| dM \\ &= \frac{1}{\sqrt{2\pi}} \frac{\bar{\rho}}{M^2} \frac{\delta_c}{\sigma} \exp \left(-\frac{\delta_c}{2\sigma^2} \right) \left| \frac{d \ln \sigma}{d \ln M} \right| dM, \end{aligned}$$

²⁷ We use

$$\left| \frac{d\sigma}{dM} \right| = \frac{\sigma}{M} \left| \frac{d \ln \sigma}{d \ln M} \right|$$

which is called the PS Halo mass function. We define

$$\nu = \frac{\delta_c(t)}{\sigma(M)}$$

and rewrite:

$$n(M, t) dM = \frac{\bar{\rho}}{M^2} f_{\text{PS}}(\nu) \left| \frac{d \ln \nu}{d \ln M} \right| dM,$$

where

$$f_{\text{PS}}(\nu) = \sqrt{\frac{2}{\pi}} \nu \exp \left(-\frac{\nu^2}{2} \right)$$

is called the Multiplicity function. The formula is convenient to express in this way, because the quantities are dimensionless.

In fig. 4.3, we see that $f(\nu)$ from our model (solid line) corresponds acceptably to simulated data. The agreement becomes much better when we allow elliptical collapse, which results in

$$f_{\text{EC}} = A \left(1 + \frac{1}{\tilde{\nu}^{2q}} \right) f_{\text{PS}}(\tilde{\nu}),$$

where $\tilde{\nu} \approx 0.84\nu$, $q = 0.3$, and $A \approx 0.322$. There are also other fitting functions that are commonly used.

Let's look at the shape and evolution of the Mass function, shown in fig. 4.4. The mass function is shown for different redshifts, where time increases from left to right.

The form of the mass function is $n \propto \nu \exp(-\nu^2/2)$, so there is an exponential cut-off for $\nu > 1$, which we can see in the plot. This corresponds to $M > M_*$, with $\sigma(M_*) = \delta_c(t)$. Since in CDM models, $D(t)$ increases with t , and $\sigma(M)$ decreases with M , M_* increases with t . This means that the cut-off occurs at higher masses as time goes on. Thus, small haloes can form first, and massive haloes form later. We have already seen this hierarchical structure formation before.

Today, the largest haloes which have formed have masses $M \approx M_* \approx 10^{14} M_\odot$. At earlier epochs, the largest haloes were found in dwarf galaxies, with $M \approx 10^9$, later galaxies with $M \approx 10^{12} M_\odot$, and finally galaxy clusters with $M \approx 10^{13} M_\odot$.

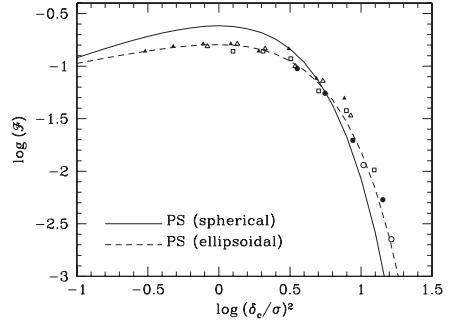


Figure 4.3: The multiplicity factor.

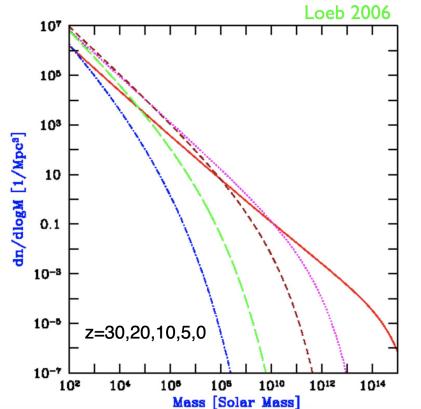


Figure 4.4: The halo mass function.

4.8 Halo mergers

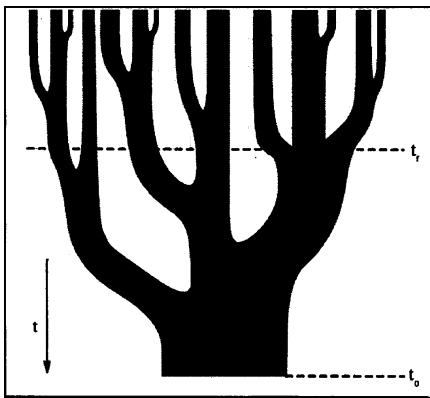


Figure 4.5: A merger tree represents the merging history of haloes. The thickness of a branch corresponds to its mass at that time. We see that several smaller progenitor haloes merge into more massive ones as time goes on.

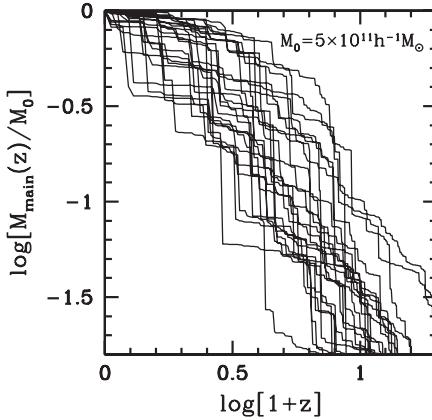


Figure 4.6: A sample of progenitor histories with mass $M_0 = 5 \cdot 10^{11} h^{-1} M_\odot$. Each line shows the evolution of the mass of the main progenitor for a different halo. Note that there is a large scatter in the history of the main progenitors.

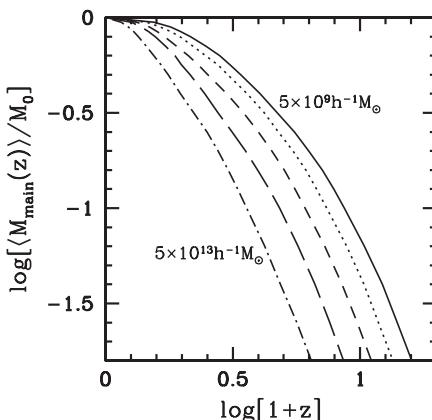


Figure 4.7: The average main progenitor histories of mergers with various masses. Each line is the average of many main progenitors with the same mass. We can see that massive haloes assemble quicker, but later

We consider a halo with mass M_0 at time t_0 . A convenient description of the history of haloes are merger trees, such as the one in fig. 4.5.

Merger trees can be generated from numerical simulations, but the method is slow and not easy to understand intuitively. Fortunately, an analytical approach, known as the *extended Press-Schechter* (EPS) approach, exists.

Progenitor mass function. Consider the average number of progenitors at time t_1 with masses $M_1 \in [M, M + dM]$, which by time t_2 have merged to form a halo of mass M_2 .

$$n(M_1, t_1 | M_2, t_2) dM_1 = \frac{M_2}{M_1} f_{\text{PS}}(\nu_{12}) \left| \frac{d \ln \nu_{12}}{d \ln M_1} \right| dM_1,$$

with

$$\nu_{12} = \frac{\delta_c(t_1) - \delta_c(t_2)}{\sqrt{\sigma^2(M_1) - \sigma^2(M_2)}}.$$

This can be derived using the EPS formalism.

Main progenitor. The most massive branch of a merger tree is called the main progenitor. In fig. 4.6, a sample of main progenitor histories for a given mass is shown, and in fig. 4.7, averages of such histories for various masses are shown.

Characteristic scales The *assembly time* t_a is the characteristic time scale for the formation of the halo. It is defined such that the mass of the main progenitor at t_a is half the final mass.

The *halo merger rate* is the rate at which a halo of mass M changes its mass by ΔM , which can be expressed as a probability $\text{Prob}(\Delta M | M, t) d \ln \Delta M d t$.

The *halo survival time* t_s is the time it takes a halo of mass M to merge to a halo of mass $2M$.

4.9 Halo clustering

We want to study how haloes are clustered in space. Consider the number overdensity of halos,

$$\delta_h(x, t) = \frac{N_h - \bar{N}_h}{\bar{N}_h},$$

where N_h is the number of haloes with mass in $[M, M + dM]$ in a small volume dV centred on position x , and \bar{N}_h is the mean number of such haloes.

4.9.1 Linear Halo bias

On linear (large) scales, one can show with the EPS-approach that

$$\delta_h(\mathbf{x}, t) = b_h(M, t)\delta(\mathbf{x}, t),$$

where b_h is a constant, called the *halo bias*, and δ is the matter overdensity. The halo bias can be derived to be

$$b_h(M, t) = 1 + \frac{1}{D(t)} \left[\frac{v^2 - 1}{\delta_c(t)} \right],$$

where we remember $v = \delta_c(t)/\sigma(M)$.

The model is compared with simulations in fig. 4.8. The dashed line is the spherical-collapse EPS model, and the solid line is a model that also includes non-spherical collapse. Once again, we see that spherical collapse gives us the right trend, but introducing elliptical collapse yields a much better agreement with simulations.

Note that for $M > 1 \cdot 10^{12} M_\odot$, the halo bias is larger than 1, so haloes are more clustered than δ . For $M < 1 \cdot 10^{12} M_\odot$ (not shown in the figure), $b_h < 1$, which is called an *antibias*. In summary, the more massive the haloes, the more strongly clustered they are.

4.9.2 More general bias models

Small scales. On small scales, we have to include non-linear effects in the bias. Also, there could be a stochastic component to the bias. Consider a more general bias model:

$$\delta_h(\mathbf{x}, t) = b_h(M, t, \delta)\delta(\mathbf{x}) + \epsilon(\mathbf{x})$$

where $b_h(\dots, \delta)$ leads to non-linear dependence on the bias, and ϵ describes the scatter between the halo contrast and the density contrast.

Assembly bias. The bias may also depend on other halo properties, such as the concentration of the halo or its spin. An important dependence is on the halo formation history. It is found that haloes with early assembly times are more strongly clustered at a given mass and time. This is called *assembly bias*.

4.10 Internal structure of Haloes

4.10.1 Density profile

From simulations, we know that density profiles of haloes are well fit by an NFW model (see section 4.5.2),

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

where ρ_s and r_s are constants. Note that the NFW profile is one of the double power law models. We see that

$$\rho(r) \propto \begin{cases} r^{-1} & \text{if } r \ll r_s \\ r^{-2} & \text{if } r \approx r_s \\ r^{-3} & \text{if } r \gg r_s, \end{cases}$$

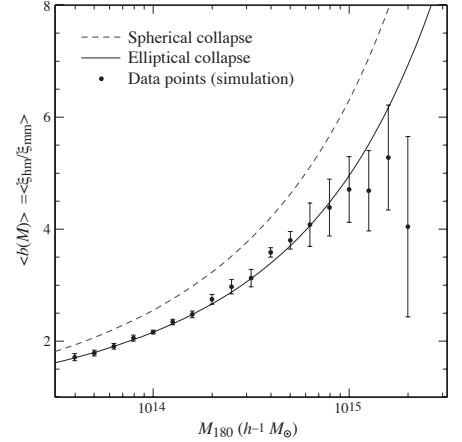


Figure 4.8: The average bias for halos of mass $> M_{180}$ predicted by analytical models of spherical and elliptical collapse.

which is similar to the singular isothermal sphere model. Since the model becomes invalid for $r \rightarrow 0$ because of the influence of baryons, the singularity at $r = 0$ is not a huge problem. The total mass is

$$\begin{aligned} M &= \int_0^{r_v} d^3x \rho(x) \\ &= 4\pi \int_0^{r_v} dr r^2 \rho(r) \\ &= 4\pi \rho_s r_s^3 \left[\ln(1 + x) - \frac{c}{1 + c} \right], \end{aligned}$$

where r_v is the virial radius, defined such that $\rho(< r_v)/\bar{\rho} = 1 + \Delta_v \approx 178$ in a matter dominated universe. The parameter $c = r_v/r_s$ is the concentration parameter.

Notes:

- Sometimes, $\Delta_v = 200$ is approximated as a convention.
- The NFW profile is a good fit to haloes for all masses and redshifts in simulations, which is why it is called a *universal profile*. It is not well-understood why there is such little variety in the large-scale structure of haloes.
- Generally, the concentration parameter $c(M, t)$ depends on mass and time. The dependence can be derived from simulations, but has so far not been predicted from analytical models.
- To a lesser degree, c also depends on the formation history of the haloes, as haloes with early assembly times are more concentrated.
- There is an even better popular fit function, called the Einasto profile. It is flatter as $r \rightarrow 0$, but it also has more parameters.

4.10.2 Halo shapes

Generally, haloes are not spherical, but elliptical. Let $a_1 \geq a_2 \geq a_3$ be the main axis of the ellipsoid, and let $s = a_3/a_1$, which is a measure of flattening of the halo. If $s = 1$, we are looking at a sphere, and $s = 0$ is a perfectly flat two-dimensional pancake. In simulations, the following trends have been found:

- Less massive haloes are more spherical.
- Haloes that assemble early are more spherical.
- The minor axis a_1 tends to be perpendicular to large-scale filaments of the cosmic web.

4.10.3 Halo substructure

In fig. 4.9, we see the result of a large scale simulation of the universe. In the dark matter distribution, there are subhaloes visible inside some haloes. The subhaloes are remnants of progenitor haloes that have merged to form larger haloes.

An interesting quantity is the *subhalo mass function*, which describes the number of subhaloes n with mass m in a main halo with total mass M . To make the result independent of the mass of the main halo, we instead use the mass fraction m/M . A good fit to simulations of the subhalo mass fraction is

$$\frac{dn}{d \ln(m/M)} \approx \frac{f_0}{\beta \Gamma(1 - \gamma)} \left(\frac{m}{\beta M} \right)^{-\gamma} \exp \left(-\frac{m}{\beta M} \right),$$

where $\gamma \approx 0.9$, $\beta \in [0.1, 0.5]$, and

$$f_0 = \frac{1}{M} \int m \frac{dM}{dm} dm.$$

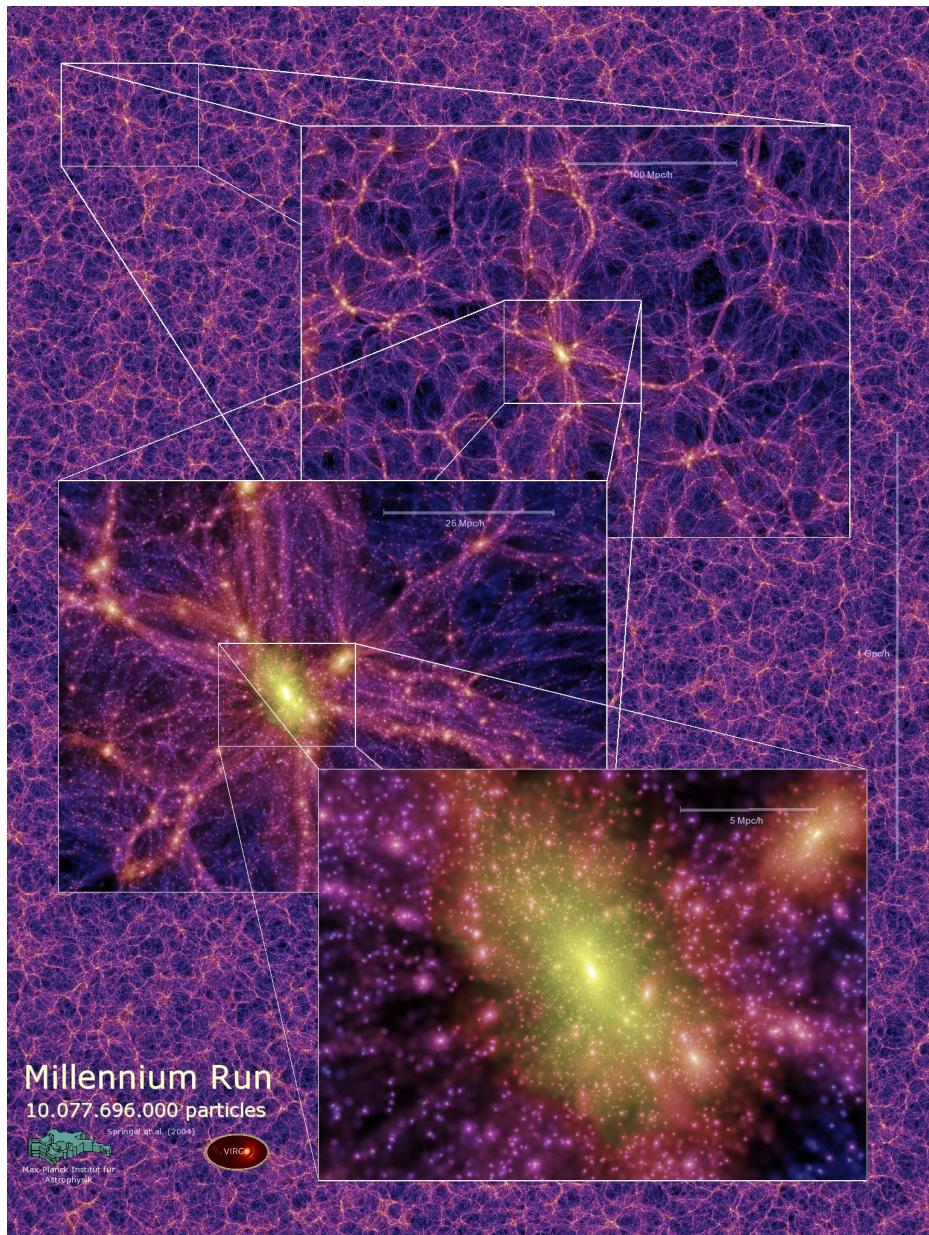


Figure 4.9: The distribution of dark matter in the universe, based on the “Millenium Run” simulation. The halo in the enlarged view consists both of a smooth dark matter distribution and several smaller subhaloes. These subhaloes are remnants of progenitor haloes that have merged to form the main halo.

The quantity f_0 is called the *mean subhalo mass fraction*, and it is the fraction of the main halo mass that is made up of subhaloes.

Subhaloes at a given time can be related to the progenitors and the survival time of progenitors following the merger.

4.10.4 Angular momentum

Haloes generally have a non-zero angular momentum, which can be characterized by the *dimensionless spin parameter*,

$$\lambda = \frac{J|E|^{1/2}}{GM^{5/2}},$$

where J is the angular momentum of the halo, E is its energy, and M is its mass. From simulations, we know that $\lambda \approx 0.03$ is quite small, which is related to haloes being supported against gravitational collapse mostly by random motion rather than rotation.

5 Baryonic Structures

5.1 State and distribution of baryons

We want to include baryons in the study of non-linear structure formation.

Evolution Before recombination and decoupling, the baryons were an ionized plasma.

After recombination, they are mostly neutral and in atomic form. At that point, there is little emission, which is why this period is called the Dark Ages.

At later times ($z \approx 6 - 20$), the first stars start to form, which emit UV photons that re-ionize the baryons and end the Dark Ages. This is called the Epoch of Reionization (EoR).

Following this, the baryons continue to participate in structure formation, driven by dark matter.

Phases

- Gas
 - Hot gas (ionized)
 - Cold gas (atomic or molecular form)
- Stars

These different phases interact and convert to each other.

Distribution

- within galaxies: Interstellar medium (ISM)
- between galaxies: Intergalactic medium (IGM)
- within galaxies clusters: Intra-cluster medium (ICM)

Dynamics We know that there is five times as much dark matter mass as there is baryonic mass. As a result, dark matter mostly drives the dynamics of structure formation and provides the backbone for the dynamics of the baryons. Unlike dark matter, baryons are collisional, and their dynamics include various effects:

- Thermodynamical effects: Heating and cooling
- Radiative transfer: Interaction between baryons and photons
- Hydrodynamical effects: ex. Shocks
- Transitions between different phases: ex. star formation
- Astrophysical effect: ex. supernova explosions, active galactic nuclei

The dynamics of baryons is more complicated than that of dark matter.

Generally, some baryons tend to sink to the bottom of potential wells, driven by dark matter on small scales, because baryons have dissipation.

Our approach here is to first review astrophysical facts about baryonic structures and discuss approaches for modelling the baryons.

5.2 Stars

Stars are formed by the collapse of cold gas in molecular clouds in the ISM. They are supported against gravity by nuclear fusion reactions at their centre. The energy that is produced is transported outwards by photons to the star's outer shells.

5.2.1 Sun

We have already seen that $M_{\odot} \approx 2 \cdot 10^{30}$ kg.

5.2.2 Composition

The solar composition is shown in TODO. Apparently, the abundance of helium and heavier elements in the sun is different from the primordial abundance. This is because heavier elements are produced by nuclear reactions in stars. To characterize these abundances, we define the *metallicity* z , which is the mass fraction of elements heavier than helium. For the sun, $z_{\odot} \approx 2\%$.

Stellar types are classified based on their surface temperature, which is shown in TODO. The spectra of different types are shown in TODO.

5.2.3 Hertzsprung-Russel diagram

TODO figure.

Stars are born on the main sequence (MS). The phenomenological *initial mass function* (IMF) is the distribution of newly born stars of the MS. Eventually, they have burned all their hydrogen and start moving away from the MS towards the red giant branch (RGB). Stars with higher masses exhaust their hydrogen first. As a red giant runs out of nuclear fuel, they collapse into compact objects. Depending on their mass, this compact object can be a white dwarf, a neutron star, or a black hole. In the process, supernova explosions can be generated. They inject kinetic energy into the ISM and disperse the heavy elements that the star generated. This process is called enrichment of the ISM.

5.2.4 Galaxies

Galaxies contain many stars, so they can be used to track the stellar evolution. Because the massive blue stars burn their fuel first, new galaxies are blue, and old galaxies are red.