ENS Lyon, M2 “Champs, Particules et Matière Condensée”
Module de
Cosmologie

Julien Lesgourgues
Chargé de Recherches au CNRS *(LAPTH, Annecy-Le-Vieux)*
Détaché au *CERN (Genève)* et a l'*EPFL (Lausanne)*

February 12, 2015

Contents

Chapter 1

Introduction to the universe expansion

1.1 Historical overview

1.1.1 The Doppler effect

At the beginning of the XX-th century, the understanding of the global structure of the universe beyond the scale of the solar system was still relying on pure speculation. In 1750, with a remarkable intuition, Thomas Wright noticed that the luminous stripe observed in the night sky and called the Milky Way could be a consequence of the spatial distribution of stars: they could form a thin plate, what we call now a galaxy. At that time, with the help of telescopes, many faint and diffuse objects had been already observed and listed, under the generic name of nebulae - in addition to the Andromeda nebula which is visible by eye, and has been known many centuries before the invention of telescopes. Soon after the proposal of Wright, the philosopher Emmanuel Kant suggested that some of these nebulae could be some other clusters of stars, far outside the Milky Way. So, the idea of a galactic structure appeared in the mind of astronomers during the XVIII-th century, but even in the following century there was no way to check it on an experimental basis.

At the beginning of the nineteenth century, some physicists observed the first spectral lines. In 1842, Johann Christian Doppler argued that if an observer receives a wave emitted by a body in motion, the wavelength that he will measure will be shifted proportionally to the speed of the emitting body with respect to the observer (projected along the line of sight):

$$\Delta\lambda/\lambda = \vec{v}.\vec{n}/c \tag{1.1}$$

where c is the celerity of the wave (See figure 1.1). He suggested that this effect could be observable for both light and sound waves. The former assumption was checked experimentally in 1868 by Sir William Huggins, who found that the spectral lines of some neighboring stars were slightly shifted toward the red or blue ends of the spectrum. So, it was possible to know the projection along the line of sight of star velocities, v_r, using

$$z \equiv \Delta\lambda/\lambda = v_r/c \tag{1.2}$$

where z is called the redshift (it is negative in case of blue-shift) and c is the speed of light. Note that the redshift gives no indication concerning

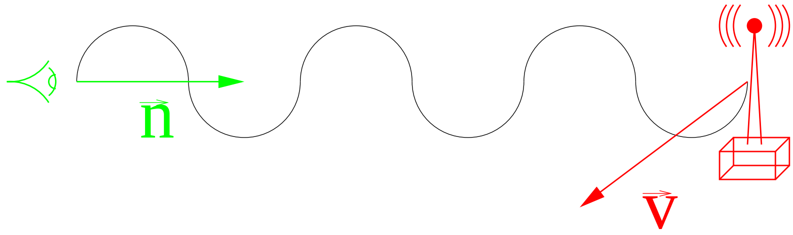

Figure 1.1: The Doppler effect.

the distance of the star. At the beginning of the XX-th century, with increasingly good instruments, people could also measure the redshift of some nebulae. The first measurements, performed on the brightest objects, indicated some arbitrary distribution of red and blue-shifts, like for stars. Then, with more observations, it appeared that the statistics was biased in favor of red-shifts, suggesting that a majority of nebulae were going away from us, unlike stars. This was raising new questions concerning the distance and the nature of nebulae.

1.1.2 The discovery of the galactic structure

In the 1920's, Leavitt and Shapley studied some particular stars, called the cepheids, known to have a periodic time-varying luminosity. They could show that the period of cepheids is proportional to their absolute luminosity L (the absolute luminosity is the total amount of light emitted by unit of time, i.e., the flux integrated on a closed surface around the star). This relation is well understood from current knowledge on stellar physics (it is due to the cycle of ionization of helium in the cepheids's athmosphere). Leavitt and Shapley were already able to measure the coefficient of proportionality (calibrated with the nearest cepheids, for which the parallax method can be employed; the parallax is half the angle under which a star appears to move when the earth makes one rotation around the sun). So, by measuring the apparent luminosity, i.e. the flux l per unit of surface through an instrument pointing to the star, it was easy to get the distance of the star r from

$$L = l \times \left(4\pi r^2\right) \ . \tag{1.3}$$

Using this technique, it became possible to measure the distance of various cepheids inside our galaxies, and to obtain the first estimate of the characteristic size of the stellar disk of the Milky Way (known today to be around 80.000 light-years).

But what about nebulae? In 1923, the 2.50m telescope of Mount Wilson (Los Angeles) allowed Edwin Hubble to make the first observation of individual stars inside the brightest nebula, Andromeda. Some of these were found to behave like cepheids, leading Hubble to give an estimate of the distance of Andromeda. He found approximately 900.000 light-years (but later, when cepheids were known better, this distance was established to be around 2 million light-years). That was the first confirmation of the galactic structure of the universe: some nebulae were likely to be some distant replicas of the Milky Way, and the galaxies were separated by large voids.

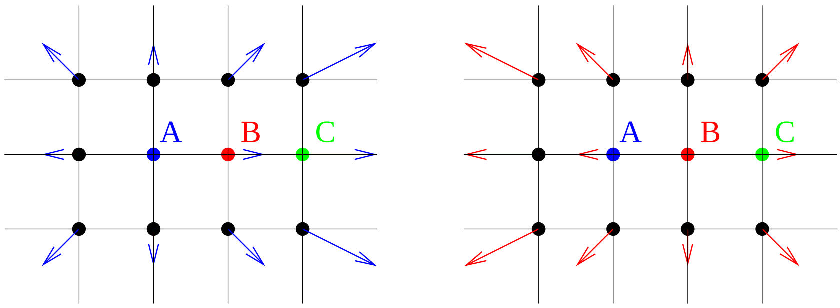

Figure 1.2: Homogeneous expansion on a two-dimensional grid. Some equally-spaced observers are located at each intersection. The grid is plotted twice. On the left, the arrays show the expansion flow measured by A; on the right, the expansion flow measured by B. If we assume that the expansion is homogeneous, we get that A sees B going away at the same velocity as B sees C going away. So, using the additivity of speeds, the velocity of C with respect to A must be twice the velocity of B with respect to A. This shows that there is a linear relation between speed and distance, valid for any observer.

1.1.3 The Cosmological Principle

This observation, together with the fact that most nebulae are redshifted (excepted for some of the nearest ones like Andromeda), was an indication that on the largest observable scales, the universe was expanding. At the beginning, this idea was not widely accepted. Indeed, in the most general case, a given dynamics of expansion takes place around a center. Seeing the universe in expansion around us seemed to be an evidence for the existence of a center in the universe, very close to our own galaxy.

Until the middle age, the Cosmos was thought to be organized around mankind, but the common wisdom of modern science suggests that there should be nothing special about the region or the galaxy in which we leave. This intuitive idea was formulated by the astrophysicist Edward Arthur Milne as the "Cosmological Principle": the universe as a whole should be homogeneous, with no privileged point playing a particular role.

Was the apparently observed expansion of the universe a proof against the Cosmological Principle? Not necessarily. The homogeneity of the universe is compatible either with a static distribution of galaxies, or with a very special velocity field, obeying to a linear distribution:

$$\vec{v} = H \, \vec{r} \tag{1.4}$$

where $\vec{v}$ denotes the velocity of an arbitrary body with position $\vec{r}$, and H is a constant of proportionality. An expansion described by this law is still homogeneous because it is left unchanged by a change of origin. To see this, one can make an analogy with an infinitely large rubber grid, that would be stretched equally in all directions: it would expand, but with no center (see figure 1.2). This result is not true for any other velocity field. For instance, the expansion law

$$\vec{v} = H \, |\vec{r}| \, \vec{r} \tag{1.5}$$

is not invariant under a change of origin: so, it has a center.

1.2 The Hubble Law

1.2.1 Hubble's discovery

So, a condition for the universe to respect the Cosmological Principle is that the speed of galaxies along the line of sight, or equivalently, their redshift, should be proportional to their distance. Hubble tried to check this idea, still using the cepheid technique. He published in 1929 a study based on 18 galaxies (in which cepheids could be seen), for which he had measured both the redshift and the distance. His results were showing roughly a linear relation between redshift and distance (see figure 1.3). He concluded that the universe was in homogeneous expansion, and gave the first estimate of the coefficient of proportionality H, called the Hubble parameter.

Hubble's measurements were rather unprecise. It is now understood that his measurement were not based on regular cepheids. Moreover, at distances of the order of 1 Mpc probed by Hubble's experiment (Mpc denotes a Mega-parsec, the unity of distance usually employed for cosmology; 1 Mpc $\simeq 3\times10^{22}$m $\simeq 3\times10^{6}$ light-years; the proper definition of a parsec is "the distance to an object with a parallax of one arcsecond"), peculiar velocities tend to dominate over the expansion flow. So, Hubble's conclusion was obviously quite biased. However, this experiment is generally considered as the starting point of experimental cosmology. Since then, many similar experiments have been performed with better and better techniques and instruments, using not only cepheids but also supernovae and other "standard candles" (i.e., objects which absolute magnitude can be inferred in some way, without knowing their distance) at larger and lager distances. Recent data (like that shown in figure 1.4) leave no doubt about the proportionality, but there is still an uncertainty concerning the exact value of H. The Hubble constant is generally parametrized as

$$H = 100\ h\ \text{km s}^{-1}\text{Mpc}^{-1} \tag{1.6}$$

where h is the dimensionless "reduced Hubble parameter", currently known to be in the range $h = 0.742 \pm 0.036$ (at the 68% confidence level) from astrophysical observations *(Astrophys.J. 699 (2009) 539)*. As we shall see later most cosmological observations confirm this range. So, for instance, a typical galaxy located at 10 Mpc goes away at a speed close to 740 km s^{-1}.

1.2.2 Homogeneity and inhomogeneities

Before leaving this section, we should clarify one point about the "Cosmological Principle", i.e., the assumption that the universe is homogeneous. Of course, nobody has ever claimed that the universe was homogeneous on small scales, since compact objects like planets or stars, or clusters of stars like galaxies are inhomogeneities in themselves. The Cosmological Principle only assumes homogeneity after smoothing over some characteristic scale. By analogy, take a grid of step l (see figure 1.5), and put one object in each intersection, with a randomly distributed mass (with all masses obeying to the same distribution of probability). Then, make a random displacement of each object (again with all displacements obeying to the same distribution of probability). At small scales, the mass density is obviously inhomogeneous for three reasons: the objects are compact, they have different masses, and they are separated by different distances. However, since the distribution has been obtained by performing a random shift in mass and position, starting from an homogeneous structure, it is clear even

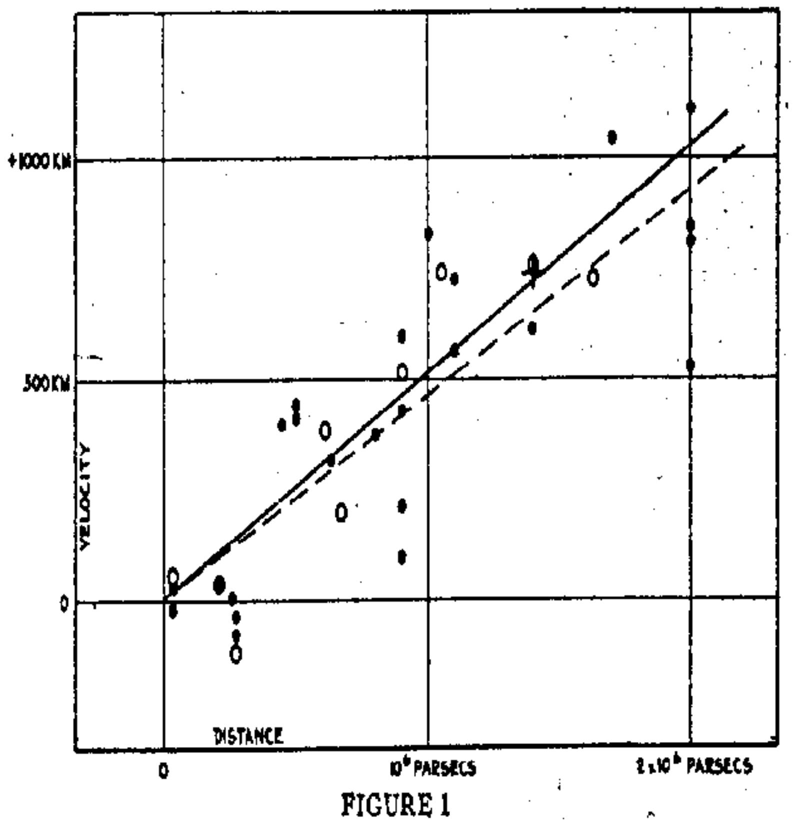

Figure 1.3: The diagram published by Hubble in 1929. The labels of the horizontal (resp. vertical) axis are 0, 1, 2 Mpc (resp. 0, 500, 1000 km.s^{-1}). Hubble estimated the expansion rate to be 500 km.s^{-1}Mpc^{-1}. Today, it is known to be around 70 km.s^{-1}Mpc^{-1}.

intuitively that the mass density smoothed over some large scale will remain homogeneous again.

The Cosmological Principle should be understood in this sense. Let us suppose that the universe is almost homogeneous at a scale corresponding, say, to the typical intergalactic distance, multiplied by thirty or so. Then, the Hubble law doesn't have to be verified exactly for an individual galaxy, because of peculiar motions resulting from the fact that galaxies have slightly different masses, and are not in a perfectly ordered phase like a grid. But the Hubble law should be verified in average, provided that the maximum scale of the data is not smaller than the scale of homogeneity. The scattering of the data at a given scale reflects the level of inhomogeneity, and when using data on larger and larger scales, the scattering must be less and less significant. This is exactly what is observed in practice. An even better proof of the homogeneity of the universe on large scales comes from the Cosmic Microwave Background, as we shall see in section 6.4.

We will come back to these issues in section 6.4, and show how the formation of inhomogeneities on small scales are currently understood and

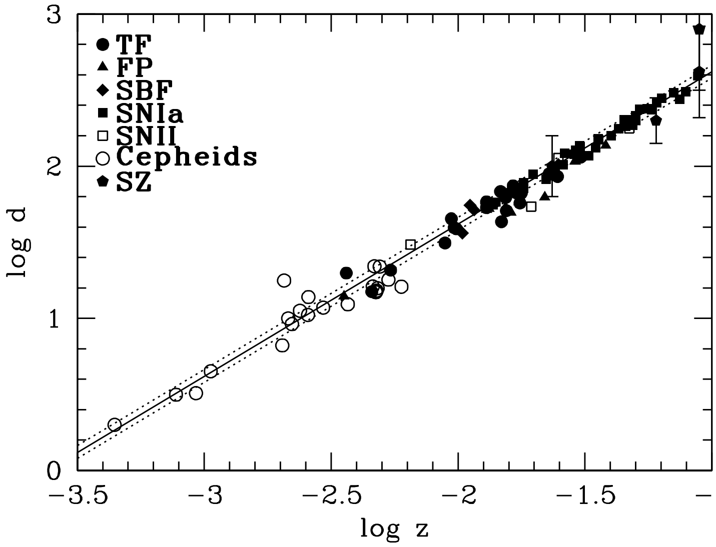

Figure 1.4: An example of Hubble diagram published by the Hubble Space Telescope Key Project in 2000 *(Astrophys.J. 553 (2001) 47-72)*, based on cepheids, supernovae and other standard candles till a distance of 400 Mpc. The horizontal axis gives the radial velocity, expressed as $\log_{10}[v/c] = \log_{10} z$ where z is redshift; the vertical axis shows the distance $\log_{10}[d/(1\text{Mpc})]$.

quantified within some precise physical models.

1.3 The universe Expansion from Newtonian Gravity

It is not enough to observe the galactic motions, one should also try to explain it with the laws of physics.

1.3.1 Newtonian Gravity versus General Relativity

On cosmic scales, the only force expected to be relevant is gravity. The first theory of gravitation, derived by Newton, was embedded later by Einstein into a more general theory: General Relativity (thereafter denoted GR). However, in simple words, GR is relevant only for describing gravitational forces between bodies which have relative motions comparable to the speed of light[1]. In most other cases, Newton's gravity gives a sufficiently accurate description.

The speed of neighboring galaxies is always much smaller than the speed of light. So, *a priori*, Newtonian gravity should be able to explain the Hubble flow. One could even think that historically, Newton's law led to the prediction of the universe expansion, or at least, to its first interpretation. Amazingly, and for reasons which are more mathematical than physical,

[1]Going a little bit more into details, it is also relevant when an object is so heavy and so close that the speed of liberation from this object is comparable to the speed of light.

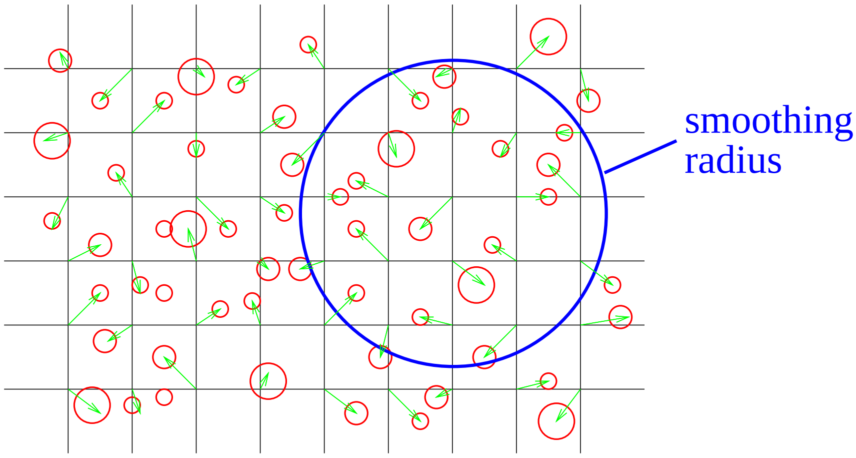

Figure 1.5: We build an inhomogeneous distribution of objects in the following way: starting from each intersection of the grid, we draw a random vector and put an object of random mass at the extremity of the vector. Provided that all random vectors and masses obey to the same distributions of probability, the mass density is still homogeneous when it is smoothed over a large enough smoothing radius (in our example, the typical length of the vectors is smaller than the step of the grid; but our conclusion would still apply if the vectors were larger than the grid step, provided that the smoothing radius is even larger). This illustrates the concept of homogeneity above a given scale, like in the universe.

it happened not to be the case: the first attempts to describe the global dynamics of the universe came with GR, in the 1910's. In this course, for pedagogical purposes, we will not follow the historical order, and start with the Newtonian approach.

Newton himself did the first step in the argumentation. He noticed that if the universe was of finite size, and governed by the law of gravity, then all massive bodies would unavoidably concentrate into a single point, just because of gravitational attraction. If instead it was infinite, and with an approximately homogeneous distribution at initial time, it could concentrate into several points, like planets and stars, because there would be no center to fall in. In that case, the motion of each massive body would be driven by the sum of an infinite number of gravitational forces. Since the mathematics of that time didn't allow to deal with this situation, Newton didn't proceed with his argument.

1.3.2 The rate of expansion from Gauss theorem

In fact, using Gauss theorem, this problem turns out to be quite simple. Suppose that the universe consists in many massive bodies distributed in an isotropic and homogeneous way (i.e., for any observer, the distribution looks the same in all directions). This should be a good modelling of the universe on sufficiently large scales. We wish to compute the motion of a particle located at a distance $r(t)$ away from us. Because the universe is assumed to be isotropic, the problem is spherically symmetric, and we can employ Gauss theorem on the sphere centered on us and attached to the particle (see figure 1.6). The acceleration of any particle on the surface of

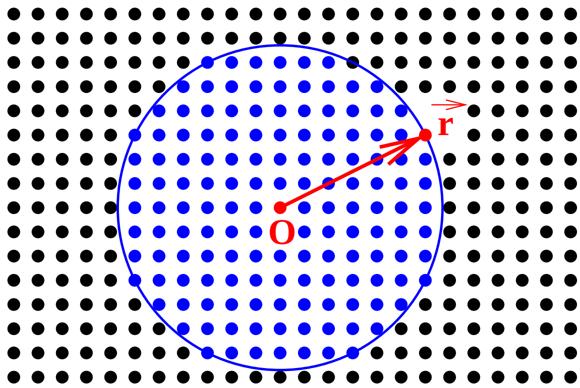

Figure 1.6: Gauss theorem applied to the local universe.

this sphere reads

$$\ddot{r}(t) = -\frac{\mathcal{G}M(r(t))}{r^2(t)} \tag{1.7}$$

where $\mathcal{G}$ is Newton's constant and $M(r(t))$ is the mass contained inside the sphere of radius $r(t)$. In other words, the particle feels the same force as if it had a two-body interaction with the mass of the sphere concentrated at the center. Note that $r(t)$ varies with time, but $M(r(t))$ remains constant: because of spherical symmetry, no particle can enter or leave the sphere, which contains always the same mass.

Since Gauss theorem allows us to make completely abstraction of the mass outside the sphere[2], we can make an analogy with the motion e.g. of a satellite ejected vertically from the Earth. We know that this motion depends on the initial velocity, compared with the speed of liberation from the Earth: if the initial speed is large enough, the satellites goes away indefinitely, otherwise it stops and falls down. We can see this mathematically by multiplying equation (1.7) by $\dot{r}$, and integrating it over time:

$$\frac{\dot{r}^2(t)}{2} = \frac{\mathcal{G}M(r(t))}{r(t)} - \frac{k}{2} \tag{1.8}$$

where k is a constant of integration. We can replace the mass $M(r(t))$ by the volume of the sphere multiplied by the homogeneous mass density

[2]The argumentation that we present here is useful for guiding our intuition, but we should say that it is not fully self-consistent. Usually, when we have to deal with a spherically symmetric mass distribution, we apply Gauss theorem inside a sphere, and forget completely about the external mass. This is actually not correct when the mass distribution spreads out to infinity. Indeed, in our example, Newtonian gravity implies that a point inside the sphere would feel all the forces from all bodies inside and outside the sphere, which would exactly cancel out. Nevertheless, the present calculation based on Gauss theorem does lead to a correct prediction for the expansion of the universe. In fact, this can be rigorously justified only *a posteriori*, after a full general relativistic study. In GR, the Gauss theorem can be generalized thanks to the consequences of Birkhoff's theorem, which is valid also when the mass distribution spreads to infinity. In particular, for an infinite spherically symmetric matter distribution, Birkhoff's theorem implies that we can isolate a sphere as if there was nothing outside of it. Once this formal step has been performed, nothing prevents us from using Newtonian gravity and Gauss theorem inside a smaller sphere, as if the external matter distribution was finite. This argument justifies rigorously the calculation of this section.

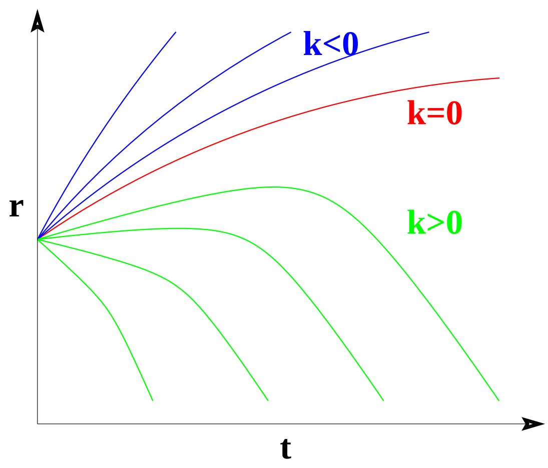

Figure 1.7: The motion of expansion in a Newtonian universe is equivalent to that of a body ejected from Earth. It depends on the initial rate of expansion compared with a critical density. When the parameter k is zero or negative, the expansion lasts forever, otherwise the universe re-collapses ($r \rightarrow 0$).

$\rho_{\text{mass}}(t)$, and rearrange the equation as

$$\left(\frac{\dot{r}(t)}{r(t)}\right)^2 = \frac{8\pi\mathcal{G}}{3}\rho_{\text{mass}}(t) - \frac{k}{r^2(t)} \ . \tag{1.9}$$

The quantity $\dot{r}/r$ is called the rate of expansion. Since $M(r(t))$ is time-independent, the mass density evolves as $\rho_{\text{mass}}(t) \propto r^{-3}(t)$ (i.e., matter is simply diluted when the universe expands). The behavior of $r(t)$ depends on the sign of k. If k is positive, $r(t)$ can grow at early times but it always decreases at late times, like the altitude of the satellite falling back on Earth: this would correspond to a universe expanding first, and then collapsing. If k is zero or negative, the expansion lasts forever.

In the case of the satellite, the critical value, which is the speed of liberation (at a given altitude), depends on the mass of the Earth. By analogy, in the case of the universe, the important quantity that should be compared with some critical value is the homogeneous mass density. If at all times $\rho_{\text{mass}}(t)$ is bigger than the critical value

$$\rho_{\text{mass}}(t) = \frac{3(\dot{r}(t)/r(t))^2}{8\pi\mathcal{G}} \tag{1.10}$$

then k is positive and the universe will re-collapse. Physically, it means that the gravitational force wins against inertial effects. In the other case, the universe expands forever, because the density is too small with respect to the expansion velocity, and gravitation never takes over inertia. The case $k = 0$ corresponds to a kind of equilibrium between gravitation and inertia in which the universe expands forever, following a power–law: $r(t) \propto t^{2/3}$.

1.3.3 The limitations of Newtonian predictions

In the previous calculation, we cheated a little bit: we assumed that the universe was isotropic around us, but we didn't check that it was isotropic everywhere (and therefore homogeneous). Following what we said before, homogeneous expansion requires proportionality between speed and distance at a given time. Looking at equation (1.9), we see immediately that this is true only when $k = 0$. So, it seems that the other solutions are not compatible with the Cosmological Principle. We can also say that if the universe was fully understandable in terms of Newtonian mechanics, then the observation of linear expansion would imply that k equals zero and that there is a precise relation between the density and the expansion rate at any time.

This argument shouldn't be taken seriously, because the link that we made between homogeneity and linear expansion was based on the additivity of speed (look for instance at the caption of figure 1.2), and therefore, on Newtonian mechanics. But Newtonian mechanics cannot be applied at large distances, where v becomes large and comparable to the speed of light. This occurs around a characteristic scale called the Hubble radius R_H :

$$R_H = cH^{-1}, \tag{1.11}$$

at which the Newtonian expansion law gives $v = HR_H = c$.

So, the full problem has to be formulated in relativistic terms. In the GR results, we will see again some solutions with $k \neq 0$, but they will remain compatible with the homogeneity of the universe.

Chapter 2

Homogeneous Cosmology

From now on, we will adopt units in which $c = \hbar = k_b = 1$ in most equations.

2.1 The Lemaître, Friedmann, Robertson & Walker metric

2.1.1 Cosmological background and perturbations

As already suggested in section 1.2.2, most calculations and predictions in cosmology are done under the assumption that the exact description of the universe can be decomposed in two problems: the background problem (which should be an independent, self-consistent problem) and the inhomogeneity problem (within a given imposed background). This is the usual approach in any theory of perturbations.

In the background problem, one assumes that in first approximation we can see the universe as a smooth distribution of matter, i.e. that one can average over small inhomogeneities like stars, galaxies and clusters, which are replaced by an idealized "cosmological fluid". The cosmological fluid can be thought to be a truly continuous distribution of matter, or equivalently, a regular distribution of compact objects, smoothed over a bigger scale than the smallest distance between these objects. The background problem consists in computing the evolution of the cosmological fluid (i.e., the distortions due to its own gravitational field, its possible transformations under phase transitions, etc.). The goal is to understand e.g. the average expansion rate as a function of time, the age of the universe, etc.

The perturbation problem consists in writing first-order (linear) perturbations in a given background and solve for their evolution. The goal is to understand, for instance, the large-scale structure of the universe or the Cosmic Microwave Background (CMB) anisotropies. The approach can even be pushed to second-order (quadratic) perturbations, but then equations become extremely complicated.

Of course, this approach cannot work for describing the formation of small scale structures. For instance, the merging of two galaxies is a fully non-linear gravitational problem which cannot be addressed by a perturbed expansion. On the other hand, it is not necessarily sensitive to General Relativity and to the expansion of the universe. The interesting question is to understand whether the cosmological perturbation theory is self-consistent on the largest scales today, and possibly on all scales in the remote past.

Today, all physicists agree that the cosmological perturbation theory provides an excellent description of the universe at early times on all scales (we will quantify the statement "early time" later in the course), which can accurately explain e.g. observations of the CMB or of light element abundances. In addition, a large majority of cosmologists believes that cosmological perturbation theory is able to explain the structure and evolution of the universe on the largest observables scales until today. On small scales, the relativistic cosmological perturbation theory should be substituted by a Newtonian non-linear approach (involving N-body gravitational clustering simulations)[1].

2.1.2 General Relativity in two words

Since in this Master, General Relativity is being taught in parallel to cosmology, we will assume in this subsection that the reader is a complete newcomer in the field, and provide some very basic intuition on General Relativity. Then, in the following sections, we will derive step by step the general relativistic laws governing the evolution universe, and stress the differences with their Newtonian counterparts.

When Einstein tried to build a theory of gravitation compatible with the invariance of the speed of light, the equivalence principle and Newton's law in some particular limit, he found that the minimal price to pay was :

- to abandon the idea of a gravitational potential, related to the distribution of matter, and whose gradient gives the gravitational field in any point.
- to assume that our four-dimensional space-time is curved, and that free-falling objects describe geodesics in this space-time.
- to relate the properties of curvature in a given point to the properties of matter in the same point.

What does that mean in simple words?

First, let's recall briefly what a curved space is, first with only two-dimensional surfaces. Consider a plane, a sphere and an hyperboloid. For us, it's obvious that the sphere and the hyperboloid are curved, because we can visualize them in our three-dimensional space: so, we have an intuitive notion of what is flat and what is curved. But if there were some two-dimensional people living on these surfaces, not being aware of the existence of a third dimension, how could they know whether they leave in a flat or a in curved space-time? There are several ways in which they could measure it. One would be to obey the following prescription: walk in straight line on a distance d; turn 90 degrees left; repeat this sequence three times again; see whether you are back at your initial position. The people on the three surfaces would find that they are back there as long as they walk along a small square, smaller than the radius of curvature. But a good test is to repeat the operation on larger and larger distances. When the size of the

[1] However, it is worth pointing out that a minority of researchers (e.g. Thomas Buchert) would say that at least one problem cannot be understood within the framework of cosmological perturbations in the recent universe and on the largest scales: namely, the current acceleration of the universe expansion. In few words, the validity of a perturbation theory is difficult to prove when the equation of motions are non-linear (which is the case of the Einstein equation involved in cosmology, see the next sections). Starting from this point, it is possible to argue that perturbation theory could require some kind of non-trivial renormalization of the background in the recent universe, when small-scale structures are very non-linear.

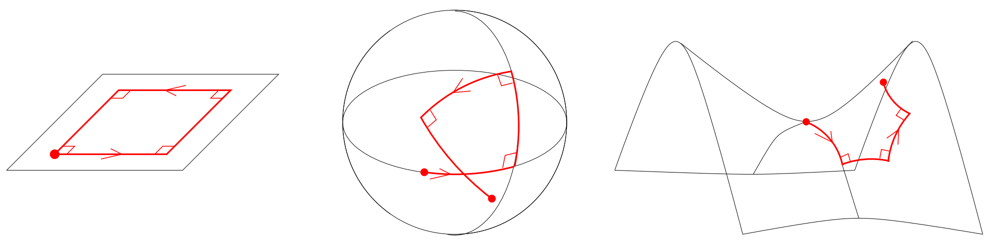

Figure 2.1: Measuring the curvature of some two-dimensional spaces. By walking four times in straight line along a distance d, and turning 90 degrees left between each walk, a small man on the plane would find that he is back at his initial point. Doing the same thing, a man on the sphere would walk across his own trajectory and stop away from his departure point. Instead, a man on the hyperboloid would not close his trajectory.

square will be of the same order of magnitude as the radius of curvature, the habitant of the sphere will notice that before stopping, he crosses the first branch of his trajectory (see figure 2.1). The one on the hyperboloid will stop without closing his trajectory. Another way to specify the curvature of a two-dimensional surface is to map it with an arbitrary coordinate system (x, y), and to use a scaling law or *line element*, i.e. a function $dl(x, y, dx, dy)$ providing a measure of infinitesimal distances as a function of position and of infinitesimal coordinate differences. For example, on projected maps of the earth's surface, one should know the scaling law in order to correctly estimate distances between two points of given latitude and longitude. At the next level of precision, the surface of the earth is curved by mountains and valleys. In a given region, having under disposal a precise topological map with contour lines of constant elevation, one can use a scaling law for estimating the physical distance between two neighboring points as a function of their latitude and longitude difference, and of the number of contour lines between the two points.

Getting an intuitive representation of a three-dimensional curved space is much more difficult. A 3-sphere and a 3-hyperboloid could be defined analytically as some 3-dimensional spaces obeying to the equation $a^2 + b^2 + c^2 \pm d^2 = R^2$ inside a 4-dimensional Euclidean or Minkowski space with coordinates (a, b, c, d). If we wanted to define them by making use of only three dimensions, the problem would be exactly like for drawing a projected map of the Earth. We would need to specify the line element $dl(x, y, z, dx, dy, dz)$ everywhere, within a given (arbitrary) coordinate system. Of course, the coordinates can be defined arbitrarily, but the physical distances computed from dl are related to intrinsic properties of the curved space, invariant under a change of coordinate. The scaling law leads to the definition of a spatial metric tensor defined through $dl^2 = g_{ij}(x^i)\, dx^i dy^j$, and to the whole formalism of Riemannian geometry (curvature tensor, intrinsic curvature scalar, geodesics, etc.).

That was still for three dimensions. The curvature of a four-dimensional space-time is very difficult to visualize intuitively, first because it has even more dimensions, and second because in special and general relativity, there is a difference between time and space. For a given space-time manifold, one can choose an arbitrary system of coordinates (time x^0 and space x^1, x^2, x^3) and describe the space-time curvature by the line element ds (which represents the infinitesimal distance betwen two closeby *events*

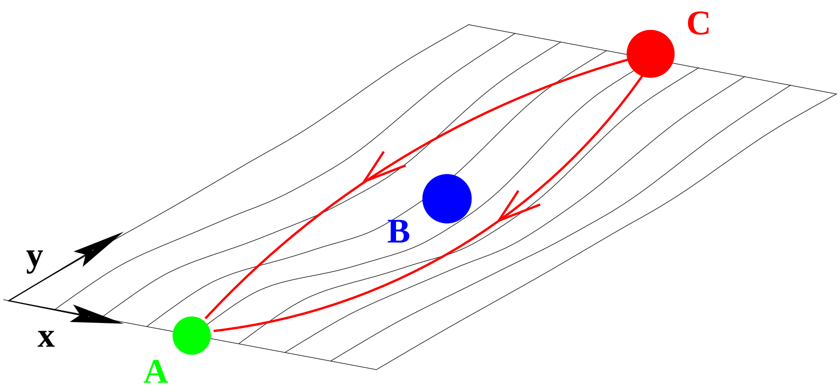

Figure 2.2: Gravitational lensing. Somewhere between an object C and an observer A, a massive object B - for instance, a galaxy - curves its surrounding space-time. Here, for simplicity, we only draw two spatial dimensions. In absence of gravity and curvature, the only possible trajectory of light between C and A would be a straight line. But because of curvature, the straight line is not anymore the shortest trajectory. Photons prefer to follow two geodesics, symmetrical around B. So, the observer will not see one image of C, but two distinct images. In fact, if we restore the third spatial dimension, and if the three points are perfectly aligned, the image of C will appear as a ring around B. This phenomenon is observed in practice.

rather than two closeby spatial points). The 4×4 metric defined through $ds^2 = g_{\mu\nu}(x^\mu)\, dx^\mu dy^\nu$ must have a negative signature (i.e. negative determinant) in order to recover locally Lorentz invariance and special relativity.

Now, the definition of geodesics is the following. Take an initial point and an initial direction. They define a unique line, called a geodesic, such that any portion of the line gives the shortest trajectory between the two points (so, for instance, on a sphere of radius R, the geodesics are all the great circles of radius R, and nothing else). In general relativity (as in any theory of gravity respecting the equivalence principle and hence based on geometry and a metric tensor), the trajectories $x^\mu(\lambda)$ of free-falling bodies are geodesics of the space-time specified by the metric $g_{\mu\nu}$. The geodesics obey to

$$\frac{d^2x^\alpha}{d\lambda^2} + \Gamma^\alpha_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0 \tag{2.1}$$

and depend on curvature through the Christoffel symbols $\Gamma^\alpha_{\mu\nu}$, which in turn can be expressed as a function of the metric as

$$\Gamma^\alpha_{\mu\nu} = \frac{1}{2} g^{\alpha\beta}(g_{\mu\beta,\nu} + g_{\beta\nu,\mu} - g_{\mu\nu,\beta}) \; . \tag{2.2}$$

Note that the expression of $g_{\mu\nu}$ and $\Gamma^\alpha_{\mu\nu}$ are not invariant under a change of coordinate, while the curvature of the underlying manifold and the ensemble of geodesics on this manifold are intrinsic, coordinate-independent properties of the manifold.

All free-falling bodies follow geodesics, including light rays. This leads for instance to the phenomenon of gravitational lensing (see figure 2.2).

The Einstein theory of gravitation says that four-dimensional space-time is curved, and that the properties of curvature in each point (related to the metric and its derivatives) depends entirely on the matter distribution in that point. In simple words, this means that the metric tensor plays more or less the same role as the gravitational potential in Newtonian gravity.

So, in General Relativity, gravitation is not formulated as a force or a field, but as the curvature of space-time, sourced by matter. All isolated systems follow geodesics which are bent by the curvature. In this way, their trajectories are affected by the distribution of matter around them: this is precisely what gravitation means.

2.1.3 Frame comoving with an observer

Let us consider a free-falling observer M in an arbitrary curved space-time. The observer's trajectory (which is a geodesic) can be described parametrically by a set of functions $\{x^1(t), x^2(t), x^3(t)\}$.

We can always perform a change of coordinates in such way that this particular observer has fixed spatial coordinates x^i_M: along this geodesic and in the new coordinates, $dx^i/dx^0 = 0$ and $x^i = x^i_M$. This frame is said to be comoving with the observer, and locally all terms g_{0i} vanish: in each point $x^\mu = (t, x^1_M, x^2_M, x^3_M)$ one has $g_{0i}(x^\mu) = 0$.

In addition, it is possible to define the time coordinate in such way that the coefficient $g_{00}^{1/2}$ (called the lapse function) is constant all over the geodesic of our particular observer, and equal to the speed of light c. If this is the case, the line element between two closeby events on the observer's trajectory is given by $ds^2 = c^2(dx^0)^2 + g_{ij}dx^i dx^j = c^2(dx^0)^2$ (since $dx^i = 0$). Hence, $dx^0 = ds/c$: the time coordinate $x^0 \equiv t$ obeys to the definition of proper time. It represents the physical time measured by our free-falling observer.

2.1.4 Building the first cosmological models

After obtaining the mathematical formulation of General Relativity around 1916, Einstein considered various testable consequences of his theory in the solar system (e.g., corrections to the trajectory of Mercury, or to the apparent diameter of the sun during an eclipse). But remarkably, he immediately understood that GR could also be applied to the universe as a whole, and published some first attempts in 1917. However, Hubble's results concerning the expansion were not known at that time, and most physicists had the prejudice that the background universe should be not only isotropic and homogeneous, but also static – or stationary. As a consequence, Einstein (and other people like De Sitter) found some interesting static cosmological solutions, but not the ones that really describe our universe.

A few years later, some other physicists tried to relax the assumption of stationarity. The first was the Russian physicist Alexander Friedmann (in 1922), followed independently by the Belgian physicist and priest Lemaître (in 1927), and then by some Americans, Robertson and Walker. When the Hubble flow was discovered in 1929, it became clear for a fraction of the scientific community that the universe could be described by the equations of Friedmann, Lemaître, Robertson and Walker. However, many people – including Hubble and Einstein themselves – remained reluctant to this idea for many years. Today, the Friedmann – Lemaître model is considered as one of the major achievements of modern physics.

2.1.5 Coordinate choice in the FLRW universe

The LFRW model is the most general solution of the GR equations under the assumption that the background universe is homogeneous and isotropic.

The fact that the universe is postulated to be homogeneous and isotropic (but not necessarily static) means that there exist a definition of time such that at each instant, all points and all directions are equivalent. For instance, the energy density should only be a function of this time, not of space: $\rho(x^\mu) = \rho(x^0)$.

We immediately notice that the fact of being "homogeneous, isotropic and non-stationary" cannot be a coordinate-independent property of a given universe, by construction: it privileges a particular definition of time, or more precisely, a particular time-slicing. A redefinition of time $t \longrightarrow t'(t)$ does not change the time-slicing. In the new time coordinate, at a given time t', all spatial points are still equivalent. A more general redefinition of time mixing time and space, $t \longrightarrow t'(t, x^1, x^2, x^3)$, changes the time-slicing: 3D hypersurfaces of constant t' are no longer homogeneous.

The easiest way to build a system of coordinates in a homogeneous universe is to start from an initial homogeneous hypersurface, to assign it a time coordinate t_1 and some arbitrary spatial coordinates. In each point, we can place an observer a rest with respect to the coordinate system: for any of these observers, $dx^i/dx^0(t_1) = 0$. This is possible by assumption: since the hypersurface is assumed to be homogeneous, there is no "force" imposing some "bulk motion" to all observers. We then give a clock to each of our observers. These clocks indicate the proper time measured by each of them. We define a new hypersurface as the ensemble of all points in space-time such that the clocks indicate a common value t_2. We assign to this hypersurface the time coordinate t_2, and some spatial coordinates such that each of our observers keeps fixed spatial coordinates. This can be repeated in order to map the entire space-time with a set of coordinates such that: all our observers keep fixed spatial coordinates; and the time coordinate corresponds to the proper time measured by all observers. In other words, we have built a frame which is comoving not just with one observer (as in a previous subsection), but with an infinity of observers mapping the entire space. These particular observers are called "comoving observers", and any set of coordinates built in that way is called a comoving coordinate system.

In comoving coordinates and using proper time, the metric describing the whole space-time reads

$$ds^2 = c^2 dt^2 + g_{ij} dx^i dx^j \tag{2.3}$$

in which t is the proper time, (x^i) are some spatial comoving coordinates, and g_{ij} must have a special form preserving the homogeneity and isotropy of three-dimensional space at any given time t. We will write down this form of g_{ij} in the next subsection. One is still free to perform some change of coordinates, and it is worth noticing that:

- a simple redefinition of time $t \longrightarrow t'(t)$ preserves the above form of the metric, excepted that $g_{00} \neq c^2$. The new time coordinate does not represent the proper time of comoving observers anymore, but it still defines a time-slicing of space-time in homogeneous hypersurfaces. In the following, we will sometimes use different definitions of the time coordinate. Physical problems can be solved with any of these time coordinates, although observables involving physical periods of time or rates should always be computed with the proper time of the observer making the experiment.

- an internal redefinition of spatial coordinates $x^i \longrightarrow x^{i'}(x^i)$ preserves the above form, and the universe will still appear as homogeneous in the new system. Hence, there is an infinite number of possible comoving spatial coordinate systems. In the following we will use cartesian coordinates, spherical coordinates, etc.

- a general change of coordinates mixing space and time would not preserve the above form of the metric. In the new coordinate system, the universe would not appear as homogeneous, since quantities like e.g. the spatial curvature or the total energy density would depend on both time and space. The new frame could represent locally the comoving frame of an observer leaving in a homogeneous universe, but not being at rest with the ensemble of comoving observers (who see homogeneous and isotropic observables). Such an observer with a peculiar velocity should not perceive the universe as isotropic: for instance, if the universe is filled with a homogeneous background of light, a non-comoving observer should see a Doppler effect affecting the color of this light (bluer in front of him, redder behind). It is important to understand that the FLRW assumption does not say that all possible observers see a homogeneous universe (this would only be possible within a homogeneous static universe), but simply that there exists an ensemble of observers seeing a homogeneous universe, and hence, a global "comoving frame".

2.1.6 The curvature of the FLRW universe

So far, we have not specified the part g_{ij}. We only assumed that it preserves homogeneity and isotropy. So, the curvature should be the same everywhere at a given time. The list of possible three-dimensional spaces with constant curvature is very short: flat Euclidean space, 3-sphere and 3-hyperboloid.

In flat space, one can use e.g. Cartesian or polar coordinates and write the spatial line element as

$$dl^2 = dx^2 + dy^2 + dz^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta \; d\phi^2) \; . \tag{2.4}$$

All possible changes of coordinate preserve this flatness. Let us rewrite the line element after the simplest possible change, namely an homothetic transformation with respect to the origin of coordinates:

$$dl^2 = a^2(dx^2 + dy^2 + dz^2) = a^2[dr^2 + r^2(d\theta^2 + \sin^2\theta \; d\phi^2)] \; . \tag{2.5}$$

Let's go back now to the full FLRW space-time. It is obvious that

$$ds^2 = c^2dt^2 - a^2(dx^2 + dy^2 + dz^2) = c^2dt^2 - a^2[dr^2 + r^2(d\theta^2 + \sin^2\theta \; d\phi^2)] \tag{2.6}$$

describes a flat, isotropic universe, but this universe is static. In fact we only want the universe to be homogeneous/isotropic *at any given time*, so

$$ds^2 = c^2dt^2 - a(t)^2(dx^2 + dy^2 + dz^2) = c^2dt^2 - a(t)^2[dr^2 + r^2(d\theta^2 + \sin^2\theta \; d\phi^2)] \tag{2.7}$$

(where we made the a factor time-dependent) is another obvious solution to the FLRW problem leading to a homogeneous, non-stationary and spatially flat universe. This is even the most general FLRW solution with zero spatial curvature (as usual, modulo trivial time redefinitions and spatial changes of coordinates)[2].

[2]Above, we performed a homothetic transformation of coordinates and allowed the factor appearing in the transformation to become a time-dependent function. We could have made a different transformation, generating other factors, and tried to make these other factors time-dependent. But in general this would break homogeneity and isotropy, unless the time-dependent factor can be factored out like in the above solution!

Again, three-dimensional spaces with constant non-zero curvature fall in two categories: 3-spheres and 3-hyperboloids. A convenient choice of polar coordinate leads to the following expression for the line elements in such spaces:

$$dl^2 = \left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta \ d\phi^2)\right] \tag{2.8}$$

where k is a constant number, related to the spatial curvature: if $k = 0$, the universe is Euclidean (and called a "flat universe"), if $k > 0$, it is positively curved (and called a "closed universe"), and if $k < 0$, it is negatively curved (and called an "open universe"). In the last two cases, the radius of curvature is given by

$$r_c(t) = \frac{1}{\sqrt{|k|}}. \tag{2.9}$$

When $k > 0$, the universe has a finite volume, and the coordinate r is defined only in the range $0 \leq r < r_c$. This is the reason for which positively curved universes are usually called "closed". The terms "open universe" just refer to the opposite case.

The most general solution for an homogeneous, isotropic, non-stationary universe is obtained again by multiplying the above spatial line element by the square of a time-dependent factor $a(t)$ called the scale factor:

$$ds^2 = c^2dt^2 - a(t)^2\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta \ d\phi^2)\right] \ . \tag{2.10}$$

The corresponding metric is called the FLRW metric (in comoving spherical coordinates). So, in three-dimensional space, infinitesimal physical distances dl are always given by the scale factor $a(t)$ times the comoving line element computed from Eq. (2.8). This is still true for a macroscopic length obtained by integrating dl over a given path in three-dimensional space: in the FRLW universe, the physical size of an object at a given time is always equal to its comoving size multiplied by the scale factor at that time[3]. In particular we can immediately notice that the physical size of the radius of curvature in the FLRW universe is

$$R_c^{\text{physical}}(t) = \frac{a(t)}{\sqrt{|k|}}. \tag{2.11}$$

The previous expression in Eq. (2.9) provides only the comoving radius of curvature. Note that r and $a(t)$ can always be rescaled by $r \longrightarrow r\sqrt{|k|}$, $a(t) \longrightarrow a(t)/\sqrt{|k|}$. After the rescaling, the metric reads like in Eq. (2.10), but with k restricted to the three possible values +1 (positive curvature), 0 (flat) or -1 (negative curvature) without loss of generality.

We know that observers at rest with the cosmological fluid have fixed comoving coordinates (it is trivial to check that all trajectories parametrized by ($x^i = x^i_M$=constant) are a solutions of the geodesics equations in the FLRW metric). This doesn't mean that the universe is static, because all distances grow proportionally to $a(t)$: so, the scale factor accounts for the homogeneous expansion. An analogy helps in understanding this concept. Let us take a rubber balloon and draw some points on the surface. Then, we inflate the balloon. The distances between all the points grow proportionally to the radius of the balloon. This is not because the points have

[3]We will see however in the next sections that due to the finite speed of light, speaking of macroscopic distance in cosmology can be somewhat subtle and require more work and definitions.

a proper motion on the surface, but because all the lengths on the surface of the balloon increase with time. In other words, in general relativity, the universe expansion is not described anymore through the velocity of objects like in Newtonian cosmology, but through the expansion of the background spacetime.

Intuitively, the FLRW metric describes a curved space-time with two types of curvature:

- the spatial curvature, described by $\pm a(t)/\sqrt{|k|}$ at each time.
- the space-time curvature described by the time evolution of $a(t)$.

The second is maybe more difficult to visualize as a curvature term, but we will see later that both terms contribute e.g. to the curvature of light ray trajectories in space-time. In a few sections, we will also see that the scale factor defines an actual radius of curvature, the Hubble radius $R_H(t) = ca(t)/\dot{a}(t)$.

If k was equal to zero and a was constant in time, we could redefine the coordinate system with $(r', \theta', \phi') = (ar, \theta, \phi)$, obtain the Minkowski metric and go back to Newtonian gravity. So, we stress again that the curvature really manifests itself as $k \neq 0$ (for spatial curvature) and $\dot{a} \neq 0$ (for the remaining space-time curvature).

Note finally that in the rest of the course, some equations may take a simpler form after the time redefinition $dt = a(t)d\tau$. In this case, the time dependence factors out from the full FLRW line element:

$$ds^2 = a^2(\tau)\left(c^2 d\tau^2 - \left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\ d\phi^2)\right]\right) \qquad (2.12)$$

where the scalar factor $a(t)$ as been re-expressed as a function of the new time variable τ. This metric exhibits conformal symmetry; hence, τ is called *conformal time*, by opposition to the proper time t, also called *cosmological time*.

2.2 Curvature of light-rays in the FLRW universe

Our goal in this section is to understand the concrete consequences of the universe expansion for observers looking at the sky. Hence, we need to understand how light rays propagate in the universe.

2.2.1 Photon geodesics

Photon propagate in the vacuum at the speed of light along geodesics. Hence, over an infinitesimal time interval time dt, they run over a distance $dl^2 = c^2 dt^2$. On macroscopic scales, the relation between distance and time is given by integrating $dl = \pm c dt$ over the geodesic.

By definition, we are only interested in photons reaching us at some point, and allowing us to observe an object. Lets us consider that we are a comoving observer and choose the origin of spherical comoving coordinates to coincide with us (this choice is only made for getting simple calculations; it doesn't imply at all that we occupy some privileged point in space or anything like that). In the FLRW universe, a photon reaching us with a momentum aligned with a given direction (θ_e, ϕ_e) must have travelled along a straight line in space, starting from an unknown emission point

(r_e, θ_e, ϕ_e). If its spatial trajectory was not a straight line, there would be a contradiction with the assumption of an isotropic universe. However the photon trajectory in space-time is curved, as can be checked by integrating over the infinitesimal distance between the emission point $(t_e, r_e, \theta_e, \phi_e)$ and a later point (t, r, θ_e, ϕ_e) with $t > t_e$, $r < r_e$:

$$\int_{r_e}^{r} -\frac{dr}{\sqrt{1-kr^2}} = \int_{t_e}^{t} \frac{c\,dt}{a(t)} \tag{2.13}$$

On can check that this trajectory is indeed a solution of the geodesic equation, and that it corresponds to a curved trajectory in space-time: if we draw this trajectory in two-dimensional (t, r) space, we see that the slope $dr/dt = -c\sqrt{1-kr^2}/a(t)$ changes along the trajectory. The photon is seen by the observer (at the origin of coordinates) at a reception time t_0 which can be deduced from r_e and t_e through the implicit relation:

$$\int_{r_e}^{0} -\frac{dr}{\sqrt{1-kr^2}} = \int_{t_e}^{t_0} \frac{c\,dt}{a(t)} \,. \tag{2.14}$$

The ensemble of all points (t_e, r_e, θ, ϕ) for which eq. (2.14) holds define our past light-cone at time t_0, as illustrated in figure 2.3. Note that the right-hand side corresponds exactly to the conformal time interval $(\tau_r - \tau_e)$ times the speed of light.

The equation (2.13) describing the propagation of light (more precisely, of radial incoming photons) is extremely useful - probably, one of the two most useful equations of cosmology, together with the Friedmann equation, that we will present soon. It is on the basis of this equation that we are able today to measure the curvature of the universe, its age, its acceleration, and other fundamental quantities.

2.2.2 A new definition of redshift

First, a simple calculation based on equation (2.13) gives the redshift associated with a given source of light. Let's still play the role of a comoving observer sitting at the origin of coordinates. We observe a galaxy located at (r_e, θ_e, ϕ_e), emitting light at a given frequency λ_e. The corresponding wave crests are emitted by the galaxy at a frequency $\nu_e = c/\lambda_e$ with a period $dt_e \equiv 1/\nu_e$. Each wave crests follows the trajectory described by Eq. (2.13). We receive the light signal with a frequency $\nu_r = c/\lambda_r = 1/dt_r$ such that

$$\int_{r_e}^{0} -\frac{dr}{\sqrt{1-kr^2}} = \int_{t_e}^{t_r} \frac{dt}{a(t)} = \int_{t_e+dt_e}^{t_r+dt_r} \frac{dt}{a(t)} \,. \tag{2.15}$$

The second equality gives:

$$\int_{t_e}^{t_e+dt_e} \frac{dt}{a(t)} = \int_{t_r}^{t_r+dt_r} \frac{dt}{a(t)} \,. \tag{2.16}$$

Hence in very good approximation:

$$\frac{dt_e}{a(t_e)} = \frac{dt_r}{a(t_r)}. \tag{2.17}$$

We infer a simple relation between the emission and reception wavelengths:

$$\frac{\lambda_r}{\lambda_e} = \frac{dt_r}{dt_e} = \frac{a(t_r)}{a(t_e)}. \tag{2.18}$$

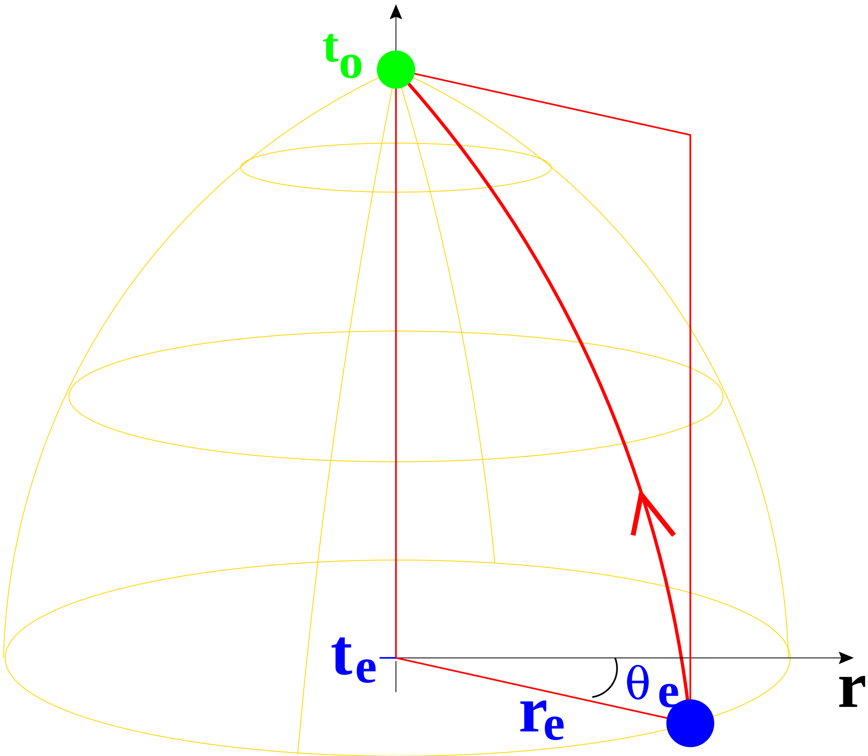

Figure 2.3: An illustration of the propagation of photons in our universe, skipping one spatial dimension. We are sitting at the origin, and at a time t_0 we can see the light of a galaxy emitted at (t_e, r_e, θ_e). Before reaching us, this light has travelled over a trajectory which is straight in three-dimensional space (constant angles), but curved in space-time. In any point, the slope dr/dt is given by equation (2.13). So, the relation between r_e, t_0 and t_e depends on the spatial curvature and on the scale factor evolution. The trajectory would be a straight line in space-time only if $k = 0$ and $a = constant$, i.e., in the limit of Newtonian mechanics in Euclidean space. The ensemble of all possible photon trajectories crossing $r = 0$ at $t = t_0$ is called our "past light cone", visible here in orange. Asymptotically, near the origin, it can be approximated by a linear cone with $dl = cdt$, showing that at small distance, the physics is approximately Newtonian. **Important remark: here, the past line cone has be drawn as a convex cone. Instead, for realistic cosmological scenarios, the cone is concave.**

This result could have been easily guessed: a wavelength is a distance, subject to the same stretching as all physical distances when the scale factor increases. Hence, in the FLRW universe, the redshift imposed by the expansion is given by

$$z = \frac{\Delta\lambda}{\lambda} = \frac{\lambda_r - \lambda_e}{\lambda_e} = \frac{a(t_r)}{a(t_e)} - 1 \ . \tag{2.19}$$

In other words, if we observe an object now, at time t_0, its absorption lines are redshifted by a factor

$$z = \frac{a(t_0)}{a(t_e)} - 1 \ . \tag{2.20}$$

This is a crucial difference with respect to Newtonian mechanics, in which the redshift was defined as $z = v/c$, and seemed to be limited to $|z| < 1$. The true GR expression doesn't have such limitations, since the ratio of the scale factors can be arbitrarily large without violating any fundamental principle. And indeed, observations do show many objects - like quasars - at redshifts of $z \sim 4$ or even bigger. We'll see later that we also observe the Cosmic Microwave Background at a redshift of approximately $z = 1100$!

Note finally that in the real perturbed universe, objects are never exactly comoving, they have small peculiar velocities $\vec{v}_c$ with respect to the comoving frame. Hence, the observed redshift is given by the sum of a General Relativity contribution given by eq. (2.20), and a Doppler contribution given by $(\vec{v}_c.\hat{n})/c$. The second term rarely exceeds $\mathcal{O}(10^{-3})$, while the first term grows from zero for nearby objects to infinity for remote objects. Hence, we expect that at very short distances, the Doppler contribution can dominate, while at larger distances the GR contribution takes over.

2.2.3 A new definition of the Hubble parameter

In the limit of small redshift, we expect to recover the Newtonian results, and to find a relation similar to $z = v/c = HL/c$ (where L is the physical distance to the object). To show this, let's assume again that t_0 is the present time, and that we are a comoving observer at $r = 0$. We want to compute the redshift of a nearby galaxy, which emitted the light that we receive today at a time $t_0 - dt$. In the limit of small dt, the equation of propagation of light shows that the physical distance L between the galaxy and us is simply

$$L \simeq dl = c\,dt \tag{2.21}$$

while the redshift of the galaxy is

$$z = \frac{a(t_0)}{a(t_0 - dt)} - 1 \simeq \frac{a(t_0)}{a(t_0) - \dot{a}(t_0)dt} - 1 = \frac{1}{1 - \frac{\dot{a}(t_0)}{a(t_0)}dt} - 1 \simeq \frac{\dot{a}(t_0)}{a(t_0)}dt \ . \tag{2.22}$$

By combining these two relations we obtain

$$z \simeq \frac{\dot{a}(t_0)}{a(t_0)} L/c \ . \tag{2.23}$$

So, at small redshift, we recover the Hubble law, and the role of the Hubble parameter is played by $\dot{a}(t_0)/a(t_0)$. In the Friedmann universe, we will define the Hubble parameter at any time as the expansion rate of the scale factor:

$$H(t) = \frac{\dot{a}(t)}{a(t)} \ . \tag{2.24}$$

The current value of the Hubble parameter (the one measured by Hubble himself) will be noted as H_0.

We have proved that in the FLRW universe, the proportionality between distance and velocity (or redshift) is recovered for small distances and redshifts. What happens at larger distance? This question actually raises a non-trivial problem: the definition of distances for objects which are so far from us that the (Euclidean) approximation $L = dl = dt$ becomes inaccurate.

2.2.4 The notion of distance to an object

Let us assume again that sitting at the origin of spherical coordinates at time t_0, we observe a remote comoving object emitting light from $(t_e, r_e, \theta_e, \phi_e)$. What is the physical distance to the object? This question is ambiguous in an expanding universe. Are we asking about the distance in units of today, i.e. the distance between us and the position of this object today? If it is a comoving object, it should be located now at coordinates $(t_0, r_e, \theta_e, \phi_e)$. Then, the distance computed on the constant-time hypersurface with $t = t_0$ is given by

$$d = \int_0^{r_e} dl = a(t_0) \int_0^{r_e} \frac{dr}{\sqrt{1 - kr^2}} \ . \tag{2.25}$$

Very often, the scale factor is defined in such way that $a(t_0) = 1$, and the above distance d coincides with the *comoving distance* $\chi(r_e)$:

$$\chi(r_e) \equiv \int_0^{r_e} \frac{dr}{\sqrt{1 - kr^2}} \ , \tag{2.26}$$

which can be integrated to

$$\chi(r) = \begin{cases} \sin^{-1}(r) & \text{if} \quad k = 1, \\ r & \text{if} \quad k = 0, \\ \sinh^{-1}(r) & \text{if} \quad k = -1. \end{cases} \tag{2.27}$$

Hence, it is useful to define the function

$$f_k(x) \equiv \begin{cases} \sin(x) & \text{if} \quad k = 1, \\ x & \text{if} \quad k = 0, \\ \sinh(x) & \text{if} \quad k = -1, \end{cases} \tag{2.28}$$

so that $r = f_k(\chi)$.

It follows from Eq. (2.14) that $\chi(r)$ is equal to the conformal age of the object, $(\tau_0 - \tau_e)$, times the speed of light:

$$\chi(r) = \int_{t_e}^{t_0} \frac{c\, dt}{a(t)} = c(\tau_0 - \tau_e) \ . \tag{2.29}$$

At this point, conformal time takes a particular signification: it is a particular measure of time, which is equal to the comoving distance traveled by a light signal divided by c. In units in which $c = 1$ and assuming $a(t_0) = 1$, both χ and τ can be expressed in units of physical distances today, e.g. in Mega-parsecs. These are indeed the most comon units for comoving distance and conformal time.

Comoving distances are well-defined quantities, up to a choice of normalization for $a(t)$. They are used by observers in many circumstances. By construction, the comoving distance between two comoving objects does not depend on time, unlike the physical distance between them. However, this is a purely conventional and rather artificial definition of distances, since we can't see remote objects today - they might even have disappeared. Anyway, we should not argue about the definition of distances, because distances are not directly measurable quantities in cosmology. We should concentrate on experimental, indirect ways to probe them. Each experimental technique will lead to a particular definition of distance.

In astrophysics, distances are usually measured in three ways:

- *From the redshift.* In principle the observed redshift of objects measures the ratio $a(t_0)/a(t_e)$ plus corrections due to the local effects of small-scale inhomogeneities (peculiar velocity of the object, ...). On very large distances, one can neglect the impact of inhomogeneities and assume in first approximation that the observed redshift is really equal to $a(t_0)/a(t_e) - 1$. Then, if we know in advance the function $a(t)$, we can identify the time t_e and compute the comoving distance $\chi(t_e)$ by integrating $(c\,dt/a(t))$ from t_e to t_0. This method is (in first approximation) the one used by observers trying to infer the spatial distribution of galaxies from *galaxy redshift surveys.* The distance reported in pictures showing the distribution of galaxies in slices of our universe is obtained in that way. However, it assumes an *a priori* knowledge of the function $a(t)$. In many cases, this function is precisely what one wants to measure.

- *From the angular diameter of standard rulers.* Surprisingly, there exist a few objects in astrophysics and cosmology which physical size can be known in advance, given some physical properties of these objects. They are called *standard rulers.* In the next chapters we will introduce one example of standard ruler: the sound horizon at decoupling, "observed" in CMB anisotropies. In Euclidean space, the distance d to a spherical object can be inferred from its physical diameter dl and its angular diameter $d\theta$ through $dl = d \times d\theta$. In FLRW cosmology, although the geometry is not Euclidean, we will adopt exactly this relation as one of the possible definitions of distance. The corresponding quantity is called the angular diameter distance d_A,
$$d_A \equiv \frac{dl}{d\theta} \ . \tag{2.30}$$
In Euclidean space, d_A would be proportional to the usual Euclidean distance to the object and therefore to its redshift. In the FLRW universe, the relation between the angular diameter distance and the redshift is non-trivial and depends on the spacetime curvature, as we shall see in the next subsection.

- *From the luminosity of standard candles.* As we have seen already with Cepheids, there exists also objects called *standard candles* for which the absolute luminosity (i.e. the total luminous flux emitted per unit of time) can be estimated independently of its distance and apparent luminosity. In Euclidean space, the distance could be inferred from the absolute luminosity L and apparent one l through $l = L/(4\pi d^2)$. In cosmology, although the geometry is not Euclidean, we will adopt exactly this relation as one of the possible definitions of distance. The corresponding quantity is called the luminosity distance d_L,
$$d_L \equiv \sqrt{\frac{L}{4\pi l}} \ . \tag{2.31}$$
In Euclidean space, d_L would be again proportional to the usual Euclidean distance to the object and therefore to its redshift, while in the FLRW universe the relation between the luminosity distance and the redshift is as subtle as for the angular diameter distance.

2.2.5 Angular diameter distance – redshift relation

Recalling that in Euclidean space with Newtonian gravity and homogeneous (linear) expansion, one has $z = v/c$ and $v = H_0 d$, we easily find a trivial relation between the angular diameter distance and the redshift:

$$d_A = d = (c/H_0)\, z. \tag{2.32}$$

In General Relativity, because of the bending of light-rays by gravity, the steps of the calculation are different. Using the definition of infinitesimal distances (2.10), we see that the physical size dl (evaluated at time t_e) of an object orthogonal to the line of sight is related to its angular diameter $d\theta$ through

$$dl = a(t_e)\; r_e\; d\theta \tag{2.33}$$

where t_e is the time at which the galaxy emitted the light ray that we observe today on Earth, and r_e is the comoving coordinate of the object. Hence

$$d_A = a(t_e)\; r_e = a(t_0)\frac{r_e}{1+z_e}\;. \tag{2.34}$$

We can replace r_e using Eqs. (2.26) - (2.29):

$$\begin{aligned} d_A &= \frac{a(t_0)}{1+z_e}\; f_k(\chi) && (2.35)\\ &= \frac{a(t_0)}{1+z_e}\; f_k\left(\int_{t_e}^{t_0} \frac{c\,dt}{a(t)}\right) && (2.36)\\ &= \frac{a(t_0)}{1+z_e}\; f_k\left(\int_{a_e}^{a_0} \frac{c\,da}{a^2 H(a)}\right) && (2.37)\\ &= \frac{a(t_0)}{1+z_e}\; f_k\left(\int_{0}^{z_e} \frac{c\,dz}{a(t_0)H(z)}\right) && (2.38) \end{aligned}$$

If we know the the curvature sign k and the function $H(z)$ up to z_e, we can compute d_A as a function of z_e. The function $d_A(z_e)$ is called the "angular diameter distance – redshift relation".

A generic consequence is that in the Friedmann universe, for an object of fixed size and redshift, the angular diameter depends on the spatial curvature - as illustrated graphically in figure 2.4. Therefore, if we know in advance the physical size of an object, we can measure on the one hand its angular diameter, on the other hand its redshift z_e, and then look for cosmological models predicting the correct value for $d_A(z_e)$.

2.2.6 Luminosity distance – redshift relation

In absence of expansion and curvature, d_L would simply correspond to the Euclidean distance to the source. On the other hand, in general relativity, it is easy to understand that the apparent luminosity is given by

$$l = \frac{L}{4\pi\, a^2(t_0)\, r_e^2 (1+z_e)^2} \tag{2.39}$$

leading to

$$d_L = a(t_0)\, r_e (1+z_e)\;. \tag{2.40}$$

Let us explain this result. First, the reason for the presence of the factor $[4\pi\, a^2(t_0)\, r_e^2]$ in equation (2.39) is obvious. The photons emitted at a comoving coordinate r_e are distributed today on a sphere of comoving radius r_e surrounding the source. Following the expression for infinitesimal distances (2.10), the physical surface of this sphere is obtained by integrating over the infinitesimal surface element $dS^2 = a^2(t_0)\, r_e^2 \sin\theta\, d\theta\, d\phi$, which gives precisely $4\pi\, a^2(t_0)\, r_e^2$. In addition, we should keep in mind that L is

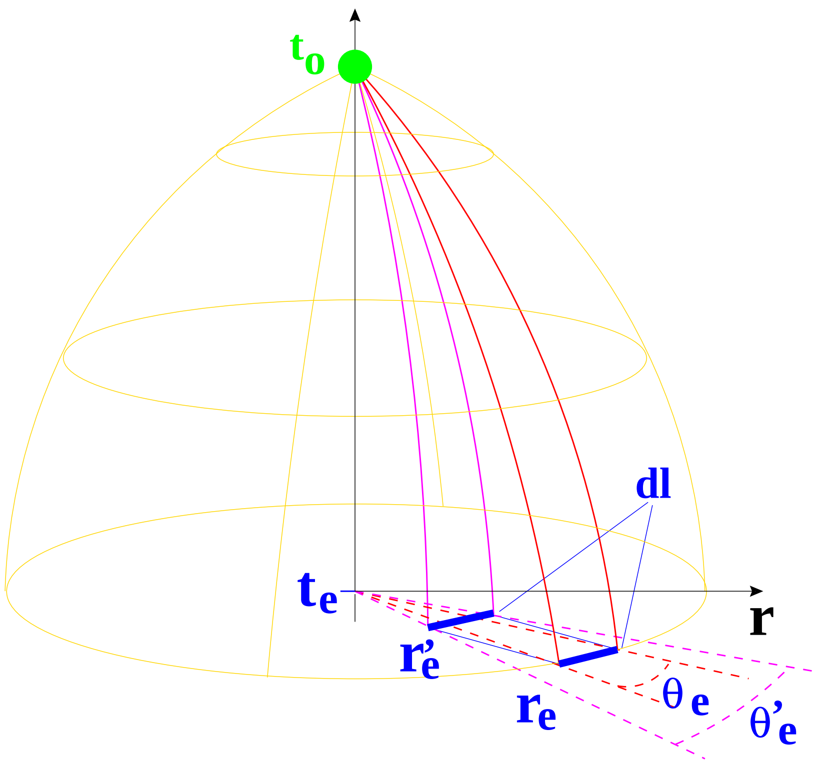

Figure 2.4: Angular diameter – redshift relation. We consider an object of fixed size dl and fixed redshift, sending a light signal at time t_e that we receive at present time t_0. All photons travel by definition with θ =constant. However, the bending of their trajectories in the (t, r) plane depends on the spatial curvature and on the scale factor evolution. So, for fixed t_e, the comoving coordinate of the object, r_e, depends on curvature. The red lines are supposed to illustrate the trajectory of light in a flat universe with $k = 0$. If we keep dl, $a(t)$ and t_e fixed, but choose a positive value $k > 0$, we infer from equation (2.13) that the new coordinate ${r_e}'$ has to be smaller. But dl is fixed, so the new angle $d\theta'$ has to be bigger, as easily seen on the figure for the purple lines. So, in a positively curved universe, objects are seen under a larger angle. Conversely, in a negatively curved universe, they are seen under a smaller angle. **Important remark: here, the past line cone has be drawn as a convex cone. Instead, for realistic cosmological scenarios, the cone is concave.**

a flux (i.e., an energy by unit of time) and l a flux density (energy per unit of time and surface). But the energy carried by each photon is inversely proportional to its physical wavelength, and therefore to $a(t)$. This implies that the energy of each photon has been divided by $(1 + z_e)$ between the time of emission and now, and explains one of the two factors $(1 + z_e)$ in (2.39). The other factor comes from the change in the rate at which photons are emitted and received (we have already seen in section 2.2.2 that since λ scales like $(1 + z_e)$, both the energy and the frequence scale like $(1 + z_e)^{-1}$).

We see that the luminosity distance is not indepent from the angular distance:

$$d_L = a(t_0)\ r_e\ (1+z_e) = a(t_e)\ r_e\ (1+z_e)^2 = (1+z_e)^2 d_A\ . \tag{2.41}$$

Like d_A, the luminosity distance can be written formally as a function of z_e:

$$d_A = a(t_0)\ (1+z_e)\ f_k \left(\int_0^{z_e} \frac{c\, dz}{a(t_0) H(z)} \right)\ . \tag{2.42}$$

Again, we would need to know the function $H(z)$ and the value of k in order to calculate explicitly the luminosity distance – redshift relation $d_L(z_e)$. In the limit $z \longrightarrow 0$, the three definition of distances given in the past sections (namely: $a(t_0)\chi$, d_A and d_L) are all equivalent and reduce to the usual definition of distance d in Euclidean space, related to the redshift through $d = z(c/H_0)$. Hence, the measurement of $d_A(z)$ and $d_L(z)$ at small redshift does not bring new information with respect to a Hubble diagram (i.e., it only allows to measure one number H_0), while measurements at high redshift depend on the spatial curvature and the dynamics of expansion. We will see in the next chapter that $d_L(z)$ has been measured for many supernovae of type Ia till roughly $z \sim 2$, leading to one of the most intriguing discovery of the past years.

In summary of this section, according to General Relativity, the homogeneous universe is curved by its own matter content, and the space–time curvature can be described by one number plus one function: the comoving spatial curvature k, and the scale factor $a(t)$. We should now be able to relate these two quantities with the source of curvature: the matter density.

2.3 The Friedmann law

In the rest of this course, we will use units such that $c = \hbar = k_B = 1$ for simplicity.

2.3.1 Einstein's equation

The relationship between the properties of matter in one point and those of curvature in the same point is given by the Einstein equation

$$G_{\mu\nu} = 8\pi G\ T_{\mu\nu}\ . \tag{2.43}$$

The Einstein tensor $G_{\mu\nu}$ can be computed for the FLRW metric using Christoffel's symbols. It is found to be diagonal ($G_{0i} = G_{i\neq j} = 0$) and isotropic ($G_{11} = G_{22} = G_{33}$). In fact, only diagonal and isotropic Einstein and energy-momentum tensors are compatible with the assumption of a homogeneous, isotropic universe with a comoving coordinate system. The most general energy-momentum tensor in such an idealized universe must be in the form

$$T^\mu_\nu = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -p & 0 & 0 \\ 0 & 0 & -p & 0 \\ 0 & 0 & 0 & -p \end{pmatrix} \tag{2.44}$$

where ρ and p stand for the energy density and pressure of the cosmological fluid. The first component of the Einstein equation reads

$$G_{00} = 3 \left[\frac{k}{a^2} + \left(\frac{\dot{a}}{a} \right)^2 \right]\ . \tag{2.45}$$

This expression is interesting to discuss. In units with $c = 1$, G_{00} appears with the dimension of an inverse squared distance, representing intuitively the curvature of the space-time manifold. Here, indeed, G_{00} is the sum of the inverse squared spatial curvature radius, $R_c(t) = \pm a/\sqrt{|k|}$, and of the inverse squared Hubble radius, $R_H(t) = a/\dot{a}$, with a multiplicative factor 3 (coming from the number of spatial dimensions). We see that the Hubble radius really plays the role of a curvature radius for space-time. We can write now the first Einstein equation $G_{00} = 8\pi G\, T_{00}$ in the FLRW universe,

$$3\left[\frac{k}{a^2} + \left(\frac{\dot{a}}{a}\right)^2\right] = 8\pi G\, \rho \ , \tag{2.46}$$

or equivalently,

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi\mathcal{G}}{3}\rho - \frac{k}{a^2} \ . \tag{2.47}$$

The above relation between the scale factor $a(t)$, the comoving spatial curvature k and the homogeneous energy density of the universe $\rho(t)$ is called the Friedmann law. Together with the propagation of light equation, this law is the key ingredient of the Friedmann-Lemaître model.

In special/general relativity, the total energy of a particle is the sum of its rest energy $E_0 = mc^2$ (i.e. $E_0 = m$ in units $c = 1$), plus its momentum energy. So, if we consider only non-relativistic particles like those forming galaxies, we can neglect the momentum energy and write $\rho = \rho_{\text{mass}}$. Then, the Friedmann equation looks exactly like the Newtonian expansion law (1.9), excepted that the function $r(t)$ (representing previously the position of objects) is replaced by the scale factor $a(t)$. Of course, the two equations look the same, but they are far from being equivalent. First, we have already seen in section 2.2.2 that although the distinction between the scale factor $a(t)$ and the classical position $r(t)$ is irrelevant at short distance, the difference of interpretation between the two is crucial at large distances – of order of the Hubble radius (in particular, in one case the existence of objects with $d > R_H$ and $z > 1$ is violating the speed-of-light limit, in the other case it is not). Second, we have seen in section 1.3.3 that the term proportional to k seems to break the homogeneity of the universe in the Newtonian formalism, while in the Friedmann model, when it is correctly interpreted as the spatial curvature term, it is perfectly consistent with the Cosmological Principle.

The next crucial difference between the Friedmann law and the Newtonian expansion law is the possibility to account for a homogeneous, isotropic fluid of relativistic particles, as we shall see in the next subsection.

2.3.2 Energy conservation

The Einstein equation implies Bianchi identities of the form $G^{\nu}_{\mu;\nu} = T^{\nu}_{\mu;\nu} = 0$. The first Bianchi identity $T^{\nu}_{0;\nu} = 0$ is nothing but the energy conservation equation. In the FLRW universe it reduces to:

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p) \ . \tag{2.48}$$

Hence, the relation between ρ and a (i.e. the way in which the energy gets diluted with the universe expansion) depends crucially on the pressure – or more precisely, on the equation of state $p(\rho)$. The most important limiting case in cosmology are:

- *non-relativistic matter.* In the limit of strongly non-relativistic matter, such as comobile objects, the negligible kinetic energy implies $p = 0$ (in absence of kinetic energy, a box enclosing the fluid would not feel any kind of pressure). If the comobile fluid represents a large-scale approximation for a homogeneous distribution of galaxies, then this approximation is fine. Hence:

$$\dot{\rho} = -3\frac{\dot{a}}{a}\rho \qquad \Rightarrow \qquad \rho \propto a^{-3}. \tag{2.49}$$

  This result is obvious. For objects with negligible velocities, the energy density is equal to the mass density, which is conserved inside any given comoving volume, since the number of comobile objects in a comoving volume is by definition constant. Since a comoving volume V increases like $V \propto a^3$ in physical units, ρ decreases like a^{-3}.

- *ultra-relativistic matter.* In the limit of ultra-relativistic matter, such as photons or massless neutrinos, the particle velocity $v = c$ generates pressure. We know from statistical thermodynamics that an ultra-relativistic gas has an equation of state $p = \rho/3$. Hence:

$$\dot{\rho} = -3\frac{\dot{a}}{a}(1 + \frac{1}{3})\rho = -4\frac{\dot{a}}{a}\rho \qquad \Rightarrow \qquad \rho \propto a^{-4}. \tag{2.50}$$

  We conclude that an ultra-relativistic fluid dilutes *faster* than a non-relativistic medium with the universe expansion. This can be understood in the following way. A homogeneous, ultra-relativistic fluid can be thought to be a gas of fast moving particles, each with $v = c$, either free-streaming or interacting with Brownian motions, such that at any time the density of particles is the same everywhere in the universe. The cosmological fluid invoked in the FLRW model could include such a component. In this case, at a given time, a comoving volume V contains N ultra-relativistic particles of individual energy $E = \nu = 1/\lambda$ (still in units with $c = \hbar = 1$). As time passes by, V increases like a^3, N is fixed (the particles move in and out of the volume, but the number of particles remains constant, otherwise the assumption of homogeneity would be violated, since V would become an over dense or underdense region). Finally, E scales like a^{-1}. Hence the energy density in the volume scales like $\rho \propto E/V \propto a^{-4}$.

In the jargon of cosmology, the ultra-relativistic component of the cosmological fluid is usually called "radiation", while the word "matter" is reserved to the non-relativistic one. The Friedmann equation is true for any types of matter, relativistic or non-relativistic; if there are different species, the total energy density ρ is the sum over the density of all species.

2.3.3 Cosmological constant

When Einstein introduced its theory, he noticed that a simple geometrical term can be added to the left-hand side without violating any principle:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G\, T_{\mu\nu}\ . \tag{2.51}$$

The number Λ (which has the dimension of an inverse squared time, as can be seen when c is restored) should depend neither on space, neither on time. It is called the cosmological constant. At some point Einstein proposed that Λ could be non-zero and negative in order to allow for a static solution to the universe equations. Then he stepped back. Anyway, we see that the cosmological constant above is rigorously equivalent to a homogeneous fluid with energy-momentum tensor

$$T^{\mu}_{\nu} = \frac{\Lambda}{8\pi G} g^{\mu}_{\nu} = \begin{pmatrix} \frac{\Lambda}{8\pi G} & 0 & 0 & 0 \\ 0 & \frac{\Lambda}{8\pi G} & 0 & 0 \\ 0 & 0 & \frac{\Lambda}{8\pi G} & 0 \\ 0 & 0 & 0 & \frac{\Lambda}{8\pi G} \end{pmatrix} . \tag{2.52}$$

By comparison with Eq. (2.44), we find that this fluid has $\rho = -p = \Lambda/8\pi G$. Looking at Eq. (2.48), we see that the equation of state $p = -\rho$ implies $\dot{\rho} = 0$, consistently with the fact that Λ should not vary with time.

A priori, a cosmological constant could be present in the universe, wither as a purely geometrical term (as in the Einstein proposal) or as some form of energy never being diluted. The vacuum energy which appears in quantum field theory (in particular, during a phase transition such as a spontaneous symmetry breaking) is of this kind: it does not dilute, and as long as the fundamental state of the theory is invariant, it remains indistinguishable from a cosmological constant. We will see that this term is probably playing an important role in our universe.

2.3.4 Various possible scenarios for the history of the universe

Let us write the Friedmann law including all possible contributions to the homogeneous cosmological fluid mentioned so far:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi\mathcal{G}}{3}\rho_{\rm R} + \frac{8\pi\mathcal{G}}{3}\rho_{\rm M} - \frac{kc^2}{a^2} + \frac{\Lambda}{3} \tag{2.53}$$

where $\rho_{\rm R}$ is the radiation density and $\rho_{\rm M}$ the matter density. The order in which we wrote the four terms on the right–hand side – radiation, matter, spatial curvature, cosmological constant – is not arbitrary. Indeed, they evolve with respect to the scale factor as a^{-4}, a^{-3}, a^{-2} and a^{0}. So, if the scale factors keeps growing, and if these four terms are present in the universe, there is a chance that they all dominate the expansion of the universe one after each other (see figure 2.5). Of course, it is also possible that some of these terms do not exist at all, or are simply negligible. For instance, some possible scenarios would be:

- only matter domination, from the initial singularity until today (we'll come back to the notion of Big Bang later).
- radiation domination $\rightarrow$ matter domination today.
- radiation dom. $\rightarrow$ matter dom. $\rightarrow$ curvature dom. today
- radiation dom. $\rightarrow$ matter dom. $\rightarrow$ cosmological constant dom. today

But all the cases that do not respect the order (like for instance: curvature domination $\rightarrow$ matter domination) are impossible.

During each stage, if we assume that one component strongly dominates the others, the behavior of the scale factor, Hubble parameter and Hubble radius are given by:

1. Radiation domination:

$$\frac{\dot{a}^2}{a^2} \propto a^{-4}, \qquad a(t) \propto t^{1/2}, \qquad H(t) = \frac{1}{2t}, \qquad R_H(t) = 2t. \tag{2.54}$$

   So, the universe is in decelerated power–law expansion.

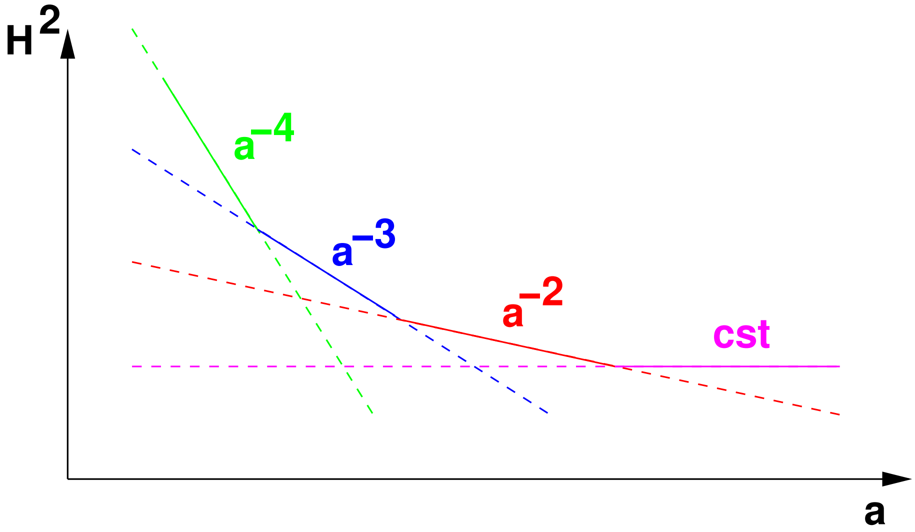

Figure 2.5: Evolution of the square of the Hubble parameter, in a scenario in which all typical contributions to the universe expansion (radiation, matter, curvature, cosmological constant) dominate one after each other.

2. Matter domination:

$$\frac{\dot{a}^2}{a^2} \propto a^{-3}, \qquad a(t) \propto t^{2/3}, \qquad H(t) = \frac{2}{3t}, \qquad R_H(t) = \frac{3}{2}t. \quad (2.55)$$

Again, the universe is in power–law expansion, but it decelerates more slowly than during radiation domination.

3. Negative curvature domination ($k < 0$):

$$\frac{\dot{a}^2}{a^2} \propto a^{-2}, \qquad a(t) \propto t, \qquad H(t) = \frac{1}{t}, \qquad R_H(t) = t. \qquad (2.56)$$

A negatively curved universe dominated by its curvature is in linear expansion.

4. Positive curvature domination: if $k > 0$, and if there is no cosmological constant, the right–hand side finally goes to zero: expansion stops. After, the scale factor starts to decrease. H is negative, but the right–hand side of the Friedmann equation remains positive. The universe recollapses. We know that we are not in such a phase, because we observe the universe expansion. But *a priori*, we might be living in a positively curved universe, slightly before the expansion stops.

5. Cosmological constant domination:

$$\frac{\dot{a}^2}{a^2} \rightarrow \text{constant}, \qquad a(t) \propto \exp(\Lambda t/3), \qquad H = 1/R_H = \sqrt{\Lambda/3}. \qquad (2.57)$$

The universe ends up in exponentially accelerated expansion.

So, in all cases, there seems to be a time in the past at which the scale factor goes to zero, called the initial singularity or the "Big Bang". The

Friedmann description of the universe is not supposed to hold until $a(t) = 0$. At some time, when the density reaches a critical value called the Planck density, we believe that gravity has to be described by a quantum theory, where the classical notion of time and space disappears. Some proposals for such theories exist, mainly in the framework of "string theories". Sometimes, string theorists try to address the initial singularity problem, and to build various scenarios for the origin of the universe. Anyway, this field is still very speculative, and of course, our understanding of the origin of the universe will always break down at some point. A reasonable goal is just to go back as far as possible, on the basis of testable theories.

The future evolution of the universe heavily depends on the existence of a cosmological constant. If the latter is exactly zero, then the future evolution is dictated by the curvature (if $k > 0$, the universe will end up with a "Big Crunch", where quantum gravity will show up again, and if $k \leq 0$ there will be eternal decelerated expansion). If instead there is a positive cosmological term which never decays into matter or radiation, then the universe necessarily ends up in eternal accelerated expansion.

2.3.5 Cosmological parameters

In order to know the past and future evolution of the universe, it would be enough to measure the present density of radiation, matter and Λ, and also to measure H_0. Then, thanks to the Friedmann equation, it would be possible to extrapolate $a(t)$ at any time[4]. Let us express this idea mathematically. We take the Friedmann equation, evaluated today, and divide it by H_0^2:

$$1 = \frac{8\pi\mathcal{G}}{3H_0^2}(\rho_{\mathrm{R}0} + \rho_{\mathrm{M}0}) - \frac{k}{a_0^2 H_0^2} + \frac{\Lambda}{3H_0^2}. \tag{2.58}$$

where the subscript 0 means "evaluated today". Since by construction, the sum of these four terms is one, they represent the relative contributions to the present universe expansion. These terms are usually written

$$\Omega_{\mathrm{R}} = \frac{8\pi\mathcal{G}}{3H_0^2}\rho_{\mathrm{R}0}, \tag{2.59}$$

$$\Omega_{\mathrm{M}} = \frac{8\pi\mathcal{G}}{3H_0^2}\rho_{\mathrm{M}0}, \tag{2.60}$$

$$\Omega_k = \frac{k}{a_0^2 H_0^2}, \tag{2.61}$$

$$\Omega_\Lambda = \frac{\Lambda}{3H_0^2}, \tag{2.62}$$

$$\tag{2.63}$$

and the "matter budget" equation is

$$\Omega_{\mathrm{R}} + \Omega_{\mathrm{M}} - \Omega_k + \Omega_\Lambda = 1. \tag{2.64}$$

The universe is flat provided that

$$\Omega_0 \equiv \Omega_{\mathrm{R}} + \Omega_{\mathrm{M}} + \Omega_\Lambda \tag{2.65}$$

[4] At least, this is true under the simplifying assumption that one component of one type does not decay into a component of another type: such decay processes actually take place in the early universe, and could possibly take place in the future.

is equal to one. In that case, as we already know, the total density of matter, radiation and Λ is equal at any time to the critical density

$$\rho_c(t) = \frac{3H^2(t)}{8\pi\mathcal{G}}. \tag{2.66}$$

Note that the parameters Ω_x, where $x \in \{R, M, \Lambda\}$, could have been defined as the present density of each species divided by the present critical density:

$$\Omega_x = \frac{\rho_{x0}}{\rho_{c0}}. \tag{2.67}$$

The physical density today ρ_{x0} of a component can be expressed in standard units, e.g. g.cm^{-3}. Another alternative is to decompose it as:

$$\begin{aligned} \rho_{x0} &= \Omega_x \frac{3H_0^2}{8\pi\mathcal{G}} = \Omega_x h^2 \, \frac{3(100\,\mathrm{km.s^{-1}.Mpc^{-1}})^2}{8\pi\mathcal{G}} && (2.68) \\ &= \Omega_x h^2 \times 1.8788 \times 10^{-29}\mathrm{g.cm^{-3}} \; . && (2.69) \end{aligned}$$

Hence, the physical density can be parametrized with the dimensionless number $\Omega_x h^2$. Later we will adopt the notation $\omega_x \equiv \Omega_x h^2$.

So far, we conclude that the evolution of the Friedmann universe can be described entirely in terms of four parameters, called the “cosmological parameters”:

$$\Omega_{\mathrm{R}}, \Omega_{\mathrm{M}}, \Omega_\Lambda, H_0. \tag{2.70}$$

One of the main purposes of observational cosmology is to measure the value of these cosmological parameters.

Chapter 3

The Hot Big Bang cosmological model

3.1 Historical overview

Curiously, after the discovery of the Hubble expansion and of the Friedmann law, there were no significant progresses in cosmology for a few decades. The most likely explanation is that most physicists were not considering seriously the possibility of studying the universe in the far past, near the initial singularity, because they thought that it would always be impossible to test any cosmological model experimentally.

Nevertheless, a few pioneers tried to think about the origin of the universe. At the beginning, for simplicity, they assumed that the expansion of the universe was always dominated by a single component, the one forming galaxies, i.e., pressureless matter. Since going back in time, the density of matter increases as a^{-3}, matter had to be very dense at early times. This was formulated as the "Cold Big Bang" scenario.

According to Cold Big Bang, in the early universe, the density was so high that matter had to consist in a gas of nucleons and electrons. Then, when the density fell below a critical value, some nuclear reactions formed the first nuclei - this era was called nucleosynthesis. But later, due to the expansion, the dilution of matter was such that nuclear reactions were suppressed (in general, the expansion freezes out all processes whose characteristic time–scale becomes smaller than the so–called Hubble time–scale H^{-1}). So, only a given number of nuclei had time to form, in some proportions which remained frozen afterward. After nucleosynthesis, matter consisted in a gas of nuclei and electrons, with electromagnetic interactions. When the density became even smaller, they finally combined into atoms – this second transition is called recombination. At late time, any small density inhomogeneity in the gas of atoms was enhanced by gravitational interactions. The atoms started to accumulate into clumps like stars and planets - but this is a different story.

In the middle of the XX-th century, a few particle physicists tried to build the first models of nucleosynthesis – the era of nuclei formation. In particular, four groups – each of them not being aware of the work of the others – reached approximately the same negative conclusion: in the Cold Big Bang scenario, nucleosynthesis does not work properly, because the formation of hydrogen is strongly suppressed with respect to that of heavier elements. But this conclusion is at odds with observations: using spectrometry, astronomers know that there is a lot of hydrogen in stars and clouds of gas. The groups of the Russo-American Gamow in the 1940's, of the Russian Zel'dovitch (1964), of the British Hoyle and Tayler (1964), and of Peebles in Princeton (1965) all reached this conclusion. They also proposed a possible way to reconcile nucleosynthesis with observations. If one assumes that during nucleosynthesis, the dominant energy density is that of photons, the expansion is driven by $\rho_{\mathrm{R}} \propto a^{-4}$, and the rate of expansion is different. This affects the kinematics of the nuclear reactions in such way that enough hydrogen can remain.

In that case, the universe would be described by a Hot Big Bang scenario, in which the radiation density dominated at early time. Before nucleosynthesis and recombination, the mean free path of the photons was very small, because they were continuously interacting – first, with electrons and nucleons, and then, with electrons and nuclei. So, their motion could be compared with the Brownian motion in a gas of particles: they formed what is called a "black–body". In any black–body, the many interactions maintain the photons in thermal equilibrium, and their spectrum (i.e., the number density of photons as a function of wavelength) obeys to a law found by Planck in the 1890's. Any "Planck spectrum" is associated with a given temperature.

Following the Hot Big Bang scenario, after recombination, the photons did not see any more charged electrons and nuclei, but only neutral atoms. So, they stopped interacting significantly with matter. Their mean free path became infinite, and they simply traveled along geodesics – excepted a very small fraction of them which interacted accidentally with atoms, but since matter got diluted, this phenomenon remained subdominant. So, essentially, the photons traveled freely from recombination until now, keeping the same energy spectrum as they had before, i.e., a Planck spectrum, but with a temperature that decreased with the expansion. This is an effect of General Relativity: the wavelength of an individual photon is proportional to the scale factor; so the shape of the Planck spectrum is conserved, but the whole spectrum is shifted in wavelength. The temperature of a black–body is related to the energy of an average photon with average wavelength: $T \sim < E > \sim \hbar c / < \lambda >$. So, the temperature decreases like $1/ < \lambda >$, i.e., like $a^{-1}(t)$.

The physicists that we mentioned above noticed that these photons could still be observable today, in the form of a homogeneous background radiation with a Planck spectrum. Following their calculations – based on nucleosynthesis – the present temperature of this cosmological black–body had to be around a few Kelvin degrees. This would correspond to typical wavelengths of the order of one millimeter, like microwaves.

These ideas concerning the Hot Big Bang scenario remained completely unknown, excepted from a small number of theorists.

In 1964, two American radio–astronomers, A. Penzias and R. Wilson, decided to use a radio antenna of unprecedented sensitivity – built initially for telecommunications – in order to make some radio observations of the Milky Way. They discovered a background signal, of equal intensity in all directions, that they attributed to instrumental noise. However, all their attempts to eliminate this noise failed.

By chance, it happened that Penzias phoned to a friend at MIT, Bernard Burke, for some unrelated reason. Luckily, Burke asked about the progresses of the experiment. But Burke had recently spoken with one of his colleagues, Ken Turner, who was just back from a visit Princeton, during which he had followed a seminar by Peebles about nucleosynthesis and possible relic radiation. Through this series of coincidences, Burke could

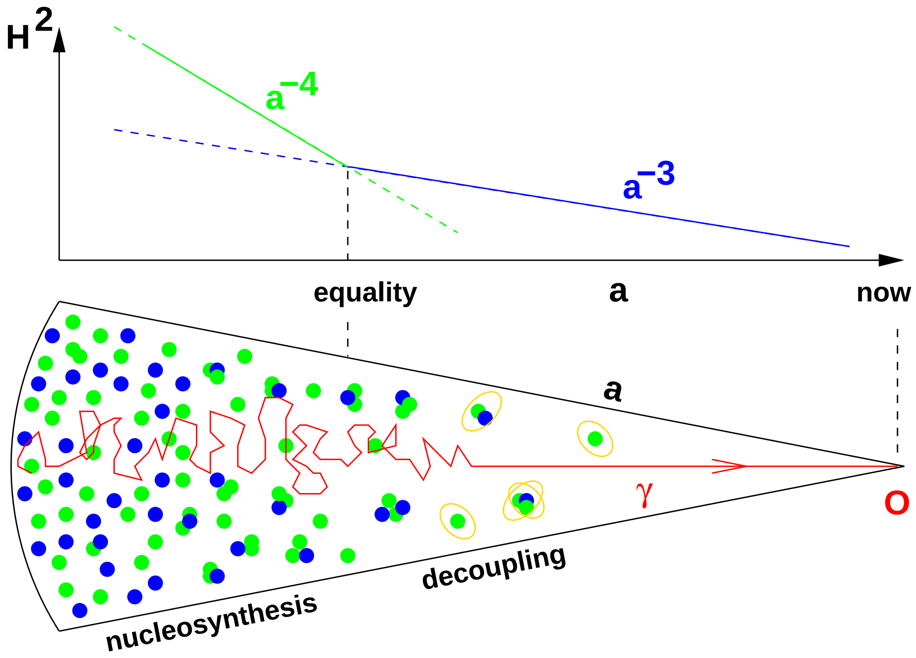

Figure 3.1: On the top, evolution of the square of the Hubble parameter as a function of the scale factor in the Hot Big Bang scenario. We see the two stages of radiation and matter domination. On the bottom, an idealization of a typical photon trajectory. Before decoupling, the mean free path is very small due to the many interactions with baryons and electrons. After decoupling, the universe becomes transparent, and the photon travels in straight line, indifferent to the surrounding distribution of electrically neutral matter.

put Penzias in contact with the Princeton group. After various checks, it became clear that Penzias and Wilson had made the first measurement of a homogeneous radiation with a Planck spectrum and a temperature close to 3 Kelvins: the Cosmic Microwave Background (CMB). Today, the CMB temperature has been measured with great precision: $T_0 = 2.726$ K.

This fantastic observation was a very strong evidence in favor of the Hot Big Bang scenario. It was also the first time that a cosmological model was checked experimentally. So, after this discovery, more and more physicists realized that reconstructing the detailed history of the universe was not purely science fiction, and started to work in the field.

The CMB can be seen in our everyday life: fortunately, it is not as powerful as a microwave oven, but when we look at the background noise on the screen of a TV set, one fourth of the power comes from the CMB!

3.2 Quantum thermodynamics in the FLRW universe

We recall that we are using units such that $c = \hbar = k_B = 1$.

Let us assume that the cosmological fluids is formed of many different species X_i (which can be either interacting with other species or free-streaming), each described by a phase-space distribution function $f_i(x^\mu, p_\nu)$. The homogeneity assumption implies that f_i should be the same everywhere; isotropy implies that it should not depend on the direction of the three-momentum p^i, but only on its modulus p; finally, the energy of each particle is given by $E = p_0 = \sqrt{m_i^2 + p^2}$. Hence the phase-space distribution can only be a function of time and of the modulus p: $f_i = f_i(p, t)$. The number density, energy density and pressure of each species read:

$$n_i(t) = \frac{g_i}{(2\pi)^3} \int d^3p \; f_i(p,t) \; , \tag{3.1}$$

$$\rho_i(t) = \frac{g_i}{(2\pi)^3} \int d^3p \; E_i \; f_i(p,t) \; , \tag{3.2}$$

$$p_i(t) = \frac{g_i}{(2\pi)^3} \int d^3p \; \frac{p^2}{3E_i} \; f_i(p,t) \; , \tag{3.3}$$

where $E_i = \sqrt{m_i^2 + p^2}$, and g_i is the number of quantum degrees of freedom (spin or helicity states) of the considered species (e.g. $g_i = 2$ for photons γ, electrons e^-, positrons e^+, protons p, anti-protons $\bar{p}$, neutrons n, anti-neutrons $\bar{n}$, or $g_i = 1$ for neutrinos ν_i and anti-neutrons $\bar{\nu}_i$ where i is one of e, μ or τ).

Interactions can be represented by a set of reactions $1 + 2 \leftrightarrow 3 + 4$ (for elastic scattering, $1 = 3$ and $2 = 4$). In general the evolution of each species due to the above reaction is represented by a Boltzmann equation of the type:

$$\frac{df_i}{dt} = F[f_1, f_2, f_3, f_4] \tag{3.4}$$

where the right-hand side, which is quite complicated to write in the general case, is a function of the distribution of each species involved in the reaction.

3.2.1 Kinetic (or thermal) equilibrium

If two species i and j have frequent interactions (like elastic scattering $i + j \longrightarrow i + j$), they exchange momentum in a random way and reach a kinetic equilibrium called "thermal equilibrium". Many species can be in thermal equilibrium, forming a so-called "thermal bath" or "thermal plasma". In thermal equilibrium, the distributions of each species depend on a common parameter, the temperature T. However the distributions f_i are not all equal to each other. They depend on:

- the mass m_i of each species (the mass appears in the energy of each particle, $E_i = \sqrt{m_i^2 + p^2}$) ;
- an additional parameter μ_i, the "chemical potential" of the species, which encodes the effect of the balance between the many reactions (inelastic scatterings) involved in the plasma;
- at the quantum level, the fact that each species should obey to the Bose-Einstein statistics for bosons (e.g. photons), or to the Fermi-Dirac statistics for fermions (in this chapter, apart from photons, we will only consider fermions).

Hence, a plasma of N species in thermal equilibrium with known masses m_i and known statistics (fermion or boson) can be entirely described in terms of a maximum of $N + 1$ free parameter $(T, \mu_1, ...\mu_N)$, which values can be inferred from considerations e.g. on energy conservation, quantum number conservation, and on the the kinetic of the various reactions involved. Thermal distributions read

$$f_i = \begin{cases} \frac{1}{\exp[\frac{E_i - \mu_i}{T}]+1} & \text{(Fermi-Dirac)} \; , \\ \\ \frac{1}{\exp[\frac{E_i - \mu_i}{T}]-1} & \text{(Bose-Einstein)} \; . \end{cases} \tag{3.5}$$

The probability of interaction between individual particles depends on a cross-section σ and on their relative velocity v. In thermal equilibrium, the interaction between two species i and j is characterized by a "thermally averaged cross-section − velocity product" $\langle \sigma v \rangle$. The interaction rate (or scattering rate) of i is given by $\Gamma_i = n_j \langle \sigma v \rangle$, that of j by $\Gamma_j = n_i \langle \sigma v \rangle$. A detailed study would show that the scattering is efficient enough for maintaining i in thermal equilibrium with j provided that the scattering rate Γ_i is larger than the inverse of the characteristic time set by the universe expansion: namely, $\Gamma_i > H$. Intuitively, when $\Gamma_i < H$, the cross-section is so low or the species j is so diluted that the chance for i to scatter over j within a time comparable to the age of the universe becomes negligible. When all possible scattering reactions which could maintain i in thermal equilibrium have $\Gamma_i < H$, the species i decouples from thermal equilibrium. In this case, assuming that the particles are stable and non-interacting, they can only free-stream with a frozen distribution (i.e., the distribution remains identical to the one at last scattering, apart from the effect of the universe expansion: $p \propto a$).

Let us review a few basic properties of thermal equilibrium which will be useful in the following sections.

- *Density of relativistic particles with negligible chemical potential.* Let us assume for simplicity that $|\mu_i| \ll T$. In this case,

$$f_i = \frac{1}{\exp[\sqrt{m_i^2 + p^2}/T] \pm 1} \; . \tag{3.6}$$

From Eq. (3.1), we see that in general the particles contributing mostly to the number density are those for which $p^2 f_i(p)$ is maximum. If $T \gg m_i$, the function $p^2 f_i(p)$ peaks at a value of p of the same order of magnitude as T, and hence for a huge majority of particles $p \gg m_i$. This corresponds to a gas of relativistic particles. The number density, energy density and pressure can be computed by integrating over the above distribution in the limit $m_i \longrightarrow 0$. The result is found to be:

$$n_i = \frac{\zeta(3)}{\pi^2} g_i T^3 \quad \left(\times \frac{3}{4} \quad \text{for fermions} \right) \; , \tag{3.7}$$

$$\rho_i = \frac{\pi^2}{30} g_i T^4 \quad \left(\times \frac{7}{8} \quad \text{for fermions} \right) \; , \tag{3.8}$$

$$p_i = \frac{1}{3} \rho_i \; , \tag{3.9}$$

where $\zeta(x)$ is the Riemann zeta function ($\zeta(3) \simeq 1.20206...$), and the extra factors for fermions come from the $+1$ term instead of -1 in the denominator of f_i. Note that the usual equation of state of a relativistic gas, $p = \sum_i p_i = \sum_i \rho_i/3 = \rho/3$, is recovered here. We conclude that boson and fermions in thermal equilibrium with each other and such that $m_i \ll T$ and $|\mu_i| \ll T$ share roughly the same number/energy density, apart from possible factors of order one.

- *Density of non-relativistic particles.* In the non-relativistic limit $m_i \gg T$, a detailed integration shows that for both fermions and bosons

$$n_i = g_i \left(\frac{m_i T}{2\pi}\right)^{3/2} \exp[-\frac{(m_i - \mu_i)}{T}] \,, \tag{3.10}$$
$$\rho_i = m_i n_i \,, \tag{3.11}$$
$$p_i = T n_i \ll \rho_i \,. \tag{3.12}$$

Let us compare the number density of these particles with that of relativistic ones still in thermal equilibrium with them:

$$\frac{n_i^{\rm NR}}{n_j^{\rm R}} = e^{\frac{\mu_i}{T}} \left[\frac{g_i}{g_j} \frac{\sqrt{\pi}}{2\sqrt{2}\zeta(3)}\right] \left(\frac{m_i}{T}\right)^{3/2} e^{-\frac{m_i}{T}} \,. \tag{3.13}$$

The factor between brackets is of order one. The part after the brackets is much smaller than one since we assumed $m_i \gg T$. Hence, unless the chemical potential is huge ($\mu_i \gg m_i \gg T$, a case that will never occur in the realistic situations considered later), the number density of non-relativistic species in thermal equilibrium is exponentially suppressed with respect to that of relativistic ones. The total number density in the thermal plasma is dominated by relativistic components.

3.2.2 Chemical equilibrium

Let's consider an inelastic scattering reaction of the type $1 + 2 \longleftrightarrow 3 + 4$. When this reaction is frequent enough, the relative number density of particles cannot be arbitrary, it must obey to the chemical equilibrium relation:

$$\mu_1 + \mu_2 = \mu_3 + \mu_4 \,. \tag{3.14}$$

When the reaction is not frequent, it is unable to maintain chemical equilibrium, and the kinetic of each particle production/annihilation must be followed using the Boltzmann equation. However, these particules can still be in thermal equilibrium (for instance, due to e.g. elastic scattering with photons). If all four species are still in thermal equilibrium, the Boltzmann equation describing e.g. the evolution of n_1 due to the above reaction takes a much simpler form than in the general case:

$$\dot{n}_1 + 3Hn_1 = n_1 n_2 \langle \sigma v \rangle \left[\exp\left(\frac{-\mu_1 - \mu_2 + \mu_3 + \mu_4}{T}\right) - 1\right] \,. \tag{3.15}$$

Here, we did two assumptions (apart for thermal equilibrium). First, we assumed that the cross section $\langle \sigma v \rangle$ is the same for the reactions $1+2 \longrightarrow 3+4$ and $3+4 \longrightarrow 1+2$. Otherwise, the right-hand side would split in two terms for creation and annihilation. However, for the realistic cases considered later, it is sufficient to consider a symmetric cross section. Second, we assumed that $1 + 2 \longleftrightarrow 3 + 4$ is the only reaction leading to the creation or annihilation of type 1 particles. If there are other processes involved, the right-hand side should contain a sum over all possible creation and decay channels.

Note that the factor $n_2 \langle \sigma v \rangle$ in the right-hand side is precisely the scattering rate Γ_1 for the scattering of type 1 particles. Hence, the second term on the left-hand side is of the order of Hn_1, while the right-hand side is of the order of $n_1 \Gamma_1$ times the brackets. We see that if $\Gamma_1 \gg H$, the term involving H can be neglected; in this regime, the differential equation forces

n_1 to reach an equilibrium value for which the brackets vanish, i.e. for which $\mu_1+\mu_2=\mu_3+\mu_4$: chemical equilibirum will be maintained at any time. In the other limit, when $\Gamma_1 \ll H$, the right-hand side is negligible, and there is no reason for the relation $\mu_1+\mu_2=\mu_3+\mu_4$ to be maintained; instead, $\dot{n}_1 = -3Hn_1$, which is equivalent to $n_1 \propto a^{-3}$: this simply corresponds to particle number conservation for a decoupled species. The intermediate regime can only be followed by integrating the above Boltzmann equation.

3.2.3 Conservation of quantum numbers

If the number of particles of a given type i was conserved in any comoving volume, we would have $n_i a^3$ =constant. This is usually *not* the case since in general, the particles i can be destroyed or created by various inelastic scatterings. So, conservation laws do not apply to the number density of individual particles, but to that of quantum numbers.

Let us consider for instance the conservation of electric charge. We can define n_+ as the sum over the number density of all particles with positive charge, weighted by the value of their charge; same for n_-. The total density of electric charge in the universe is then simply $n_Q \equiv n_+ - n_-$. Electric charge is a conserved number, so the charge in any comoving volume must be constant. Hence $n_Q a^3$ is constant. The same holds for other quantities such as baryon number ($n_B a^3$ =constant), lepton number ($n_L a^3$ =constant), etc. (excepted at very early times for which baryon or lepton number conservation can be violated in special circumstances, as we shall see later).

However, in the case of the electric charge, we have an even stronger constraint: since the electric charge is associated with Coulomb forces and the universe expansion is only governed by gravitational forces, the universe must be globally neutral: hence $n_Q = 0$ and $n_+ = n_-$.

Note that each conserved quantum number is *usually* associated with a non-zero chemical potential. When a particle X_i carries no conserved charge, nothing prevents reactions of the type $nX_i \rightarrow mX_i$ with $n \neq m$. This is the case for photons. For instance, as long as the universe contains electrons and positrons, the two reactions

$$3\gamma \longleftrightarrow e^+ + e^- \longleftrightarrow 2\gamma \tag{3.16}$$

are in chemical equilibrium, hence $2\mu_\gamma = 3\mu_\gamma$ and $\mu_\gamma = 0$. In addition, the above reactions tell us that electrons and positrons (which carry electric charges ± 1 and lepton numbers ± 1) have opposite chemical potentials, $\mu_{e^+} = -\mu_{e^-}$. It is not possible to find a reaction that would lead to the conclusion that $\mu_{e^+} = \mu_{e^-} = 0$ without violating charge or lepton number conservation. A species carrying a conserved charge can have a zero chemical potential, but only if we invoke external constraints on top of chemical equilibrium considerations.

3.2.4 Entropy conservation in the thermal bath.

We just said that there is no reason for conserving the total number density of particles in a given comoving volume. However, it is possible to show (although we skip the proof in this course) that the total entropy (i.e. the number of possible states) in any comoving volume is conserved, and that the entropy density of a thermal plasma reads

$$s = \frac{\rho + p}{T} \tag{3.17}$$

where ρ and p are the total density and pressure of species in thermal equilibrium. **Further justifications of this result will be provided in the course of Pierre Salati (next semester).** Let us consider a thermal bath composed of a number of relativistic and non-relativistic species, and let us assume further that the density of non-relativistic particles is negligible with respect to that of relativistic ones (this assumption holds throughout the radiation dominated era in the early universe). The total density and pressure are then equal to

$$\rho_{\rm tot} = \frac{\pi^2}{30} g_* T^4 \;, \qquad p_{\rm tot} = \frac{1}{3} \rho_{\rm tot} \;, \tag{3.18}$$

where we have introduced the *number of relativistic degrees of freedom* g_* defined through

$$g_* = \sum_{\rm rel.bosons} g_i + \frac{7}{8} \sum_{\rm rel.fermions} g_i \;. \tag{3.19}$$

The entropy density is then

$$s = \frac{4}{3} \frac{\pi^2}{30} g_* T^3 \;, \tag{3.20}$$

and its conservation implies $g_* T^3 a^3$ =constant. We see that as long as g_* is constant, $T \propto a^{-1}$. However, when g_* varies (which can happen e.g. if one species becomes non-relativistic at some point), the temperature varies like $T \propto g_*^{-1/3} a^{-1}$.

Note that entropy conservation is really different from number density conservation. For instance, in the above example, the number density reads

$$n_{\rm tot} = \frac{\zeta(3)}{\pi^2} \left[\sum_{\rm rel.bosons} g_i + \frac{3}{4} \sum_{\rm rel.fermions} g_i \right] T^3 \;. \tag{3.21}$$

The term between brackets differs from g_* due to the factor 7/8. Hence, when g_* varies, the quantity $n_{\rm tot} a^3$ is *not* constant, since the entropy is *not* equivalent to the number density!

3.3 The Thermal history of the universe

3.3.1 Early stages

The most early stages in the evolution of the universe are still partially unknown and source of active research, while late times are very well constrained and obey to models validated by observations. In summary, the epoch during which the energy scale $\rho_{\rm tot}^{1/4}$ of the universe was smaller than 100 MeV is rather well understood, while early stages are still quite uncertain. In this subsection, we will provide a very brief overview of what could have happened above 100 MeV. In the next subsections, we will describe in more details the main events taking place below 100 MeV.

Following the most conventional picture, gravity became a classical theory (with well–defined time and space dimensions) at a time called the Planck time [1]: $t \sim 10^{-36}$s, $\rho \sim M_P^4 \sim (10^{18}\text{GeV})^4$ (where the Planck mass is defined by $M_P = G^{-1/2}$: the Friedmann equation can also be written as $3 M_P^2 H^2 = 8\pi\rho$, and the Planck time corresponds to $H = M_P$, i.e. to a Hubble radius equal to the Planck length $R_H = 1/M_P = \lambda_P$; all these relations

[1]By convention, the origin of time is chosen by extrapolating the scale-factor to $a(0) = 0$. Of course, this is only a convention, it has no physical meaning.

are written as usual for $c = \hbar = k_B = 1$ units). Later, there was most probably a stage of accelerated expansion called *inflation*. Current observations provide some indirect, but precise information on inflation, which is quite extraordinary since this stage took place at extremely high energy. Inflation might be related to the spontaneous symmetry breaking of the GUT (Grand Unified Theory) symmetry around $t \sim 10^{-32}$s, $\rho \sim (10^{16}\text{GeV})^4$. We will describe the motivations and predictions of inflation in the last chapter.

After inflation, during a stage called reheating, the scalar field responsible for inflation decayed into the particles of the standard model (three families of quarks, anti-quarks, leptons and anti-leptons; Higgs boson(s); gauge bosons; and possibly also some particles belonging to extensions of the standard model, like maybe supersymmetric particles), which reached thermal equilibrium after some time. At such high energy, most (if not all) particles where ultra-relativistic ($T > m_i$), and the total energy and pressure were given by Eq. (3.18). The end of reheating marks the beginning of the radiation dominated era suggested by Gamow, Peebles and others. Note that during this era, $T \propto a^{-1}$ and $\rho \propto a^{-4}$ in good approximation, although these scaling are slightly violated each time that g_* varies (this occurs from time to time e.g. when some particles become non-relativistic). Around $t \sim 10^{-6}$s, $\rho \sim (100\ \text{GeV})^4$, the EW (Electro Weak) symmetry is spontaneously broken and the quarks acquire a mass through the Higgs mechanism. Later, at $t \sim 10^{-4}$s, $\rho \sim (100\ \text{MeV})^4$, the QCD (Quantum Chromo Dynamics) transition forces quarks to get confined into hadrons: baryons and mesons.

All these stages are quite complicated and extremely interesting to investigate in details (here we will not touch this at all). Let us mention that a particularly fascinating and important issue is the evolution of the baryon and lepton number.

Let us focus first on the baryon number. Before reheating, there is no baryon number. Hence, if the baryon number is always conserved, each time that a particle is created during reheating with a given baryon number, its anti-particle with opposite baryon number will also be created. The pairs of particle-antiparticles will not annihilate in the relativistic regime. For simplicity, let us do as is there was only one type of particle with a baryon number, say b with baryon number $B = 1$ and its antiparticle $\bar{b}$ with $B = -1$. These particles could in principle annihilate through e.g.

$$b + \bar{b} \leftrightarrow n\gamma \tag{3.22}$$

(n being the number of produced photons). Note that a particle and its anti-particle should share the same mass m_b. Intuitively, as long as $T \gg m_b$, the photons carry enough energy for creating pairs of b and $\bar{b}$, so they will coexist in the thermal plasma: annihilation and creation compensate each other. However, when $T < m_b$, the photons do not carry enough energy for creating pairs, and only annihilation can occur: so, b and $\bar{b}$ annihilate. If we assume that the baryon number is always conserved, then the annihilation will be total and we will be left with no baryons at all today. This is not the case since the nuclei of atoms are made of protons and neutrons. Hence, the baryon number conservation has to be violated at some point between reheating and $T \sim m_b$. When the violation occurs, an excess of particles with positive B can be created. This is called baryogenesis. When $T \sim m_b$, all baryons annihilate with antibaryons, excepted the few ones in excess, which remain till today.

Let us give a very simplified mathematical description of this phenomenon: after baryogenesis, the universe contains relativistic baryons and anti-baryons in thermal and kinetic equilibrium. The reaction

$$b + \bar{b} \leftrightarrow n\gamma \tag{3.23}$$

with different possible values of n guaranties that $\mu_\gamma = 0$ and $\mu_b = -\mu_{\bar{b}}$. If $\mu_b = 0$, then n_b is exactly equal to $n_{\bar{b}}$. The outcome of baryogenesis should be a small excess of baryons, hence $\mu_b > 0$. The conserved baryon number $n_B a^3$ is non-zero and obtains from

$$n_B = n_b - n_{\bar{b}} = \frac{g_b}{(2\pi)^3} \int d^3p \left[\frac{1}{\exp(\frac{E-\mu_b}{T}) + 1} - \frac{1}{\exp(\frac{E+\mu_b}{T}) + 1} \right] . \tag{3.24}$$

In the relativistic limit $E = p$ this gives

$$n_B = \frac{g_b T^3}{6\pi^2} \left[\pi^2 \left(\frac{\mu_b}{T} \right) + \left(\frac{\mu_b}{T} \right)^3 \right] . \tag{3.25}$$

which is positive for $\mu_b > 0$. As long as Ta =constant (i.e. as long as g_* is constant in the thermal bath), the conservation of $n_B a^3$ implies that μ_b/T is also constant. The baryon asymmetry can be parametrized by

$$\frac{n_B}{n_b + n_{\bar{b}}} = n_B / \left[2 \times \frac{3}{4} \frac{\zeta(3)}{\pi^2} g_b T^3 \right] \sim \left[\pi^2 \left(\frac{\mu_b}{T} \right) + \left(\frac{\mu_b}{T} \right)^3 \right] \tag{3.26}$$

but this is not a conserved number. Usually, the asymmetry is parameterized by n_B/s, which is really a conserved number since both the baron number $n_B a^3$ and entropy sa^3 are conserved. We will see later that in order to obtain the correct baryon density today, we must assume that n_B/s is of the order of 10^{-10}.

Note that when the universe is filled with a thermal plasma, s is of the order of $g_* T^3$, while n_γ is of the order of $g_\gamma T^3$ with $g_\gamma = 2$. So, instead of n_B/s, we will often use the ratio n_B/n_γ, although strictly speaking the second number if not conserved and differs from the first one by a factor of the order of g_* (which can vary between ~ 3 and ~ 10 during the period that we will study in the next sections). In the recent universe we will see that

$$\eta_b \equiv \frac{n_B}{n_\gamma} \sim 5 \times 10^{-10} . \tag{3.27}$$

When $T \sim m_B$, the number density of both n_b and $n_{\bar{b}}$ drops down very quickly due to the $\exp(-m_b/T)$ factor. Intuitively, this means that a smaller and smaller fraction of photons have enough energy for producing $b+\bar{b}$ pairs. The assumption of thermal and kinetic equilibrium and the conservation of entropy and baryon number provide enough equations for following $\mu_b(T)$ and $T(a)$ until $n_{\bar{b}}$ becomes really negligible. We don't even need to do that: it is enough to know that when $n_{\bar{b}} = 0$, baryon number conservation simply implies that $n_b a^3 = n_B a^3$ is constant. Note that at that time

$$n_b = g_b \left(\frac{m_b T}{2\pi} \right)^{3/2} e^{-\frac{(m_b - \mu_b)}{T}} , \tag{3.28}$$

so the quantity μ_b/T now varies with time, in order to maintain $n_b a^3$ =constant.

This description of the matter-antimatter asymmetry in the early universe was quite simplistic with respect to reality. Actually, baryogenesis and baryon-antibaryon annihilation are two active topics of research. Baryogenesis could be associated with B-violating processes during GUT symmetry breaking or EW symmetry breaking, or could also be induced by leptogenesis, for which a similar discussion can hold. The baryon-antibaryon annihilation is expected to take place roughly around $T \sim 1000$ MeV, which is the order of magnitude of the proton mass; it is intimately related to the quark-hadron transition.

3.3.2 Content of the universe around $T \sim$10 MeV

In the next sections, we will describe a list of phenomena induced by the fact that the weak interactions become inefficient around 1 MeV, and also that the MeV is the order of magnitude of binding energies in light nuclei. Before these sections, we should look at initial conditions before $T \sim$MeV.

Let us list the species present after the quark-hadron transition. A species can be present at a given time if it satisfies one of two conditions:

- either it is relativistic: $m \ll T$. In this case the particle can be easily produced by other species in the thermal bath (annihilation and creation compensate each other).

- or it is stable thanks to the conservation of a quantum number. In this case, the particle may have $m \gg T$, but cannot decay with violating the conservation of this number. Typically, the particles in the category are the lightest ones carrying a given quantum number. For instance, the proton is the lightest baryon.

Generally speaking, hadrons consist of baryons, mesons and their antiparticles. Mesons carry zero baryon number and quickly annihilate. Antibaryons annihilate well before $T \sim 10$MeV, as described above. Baryons made of heavy quarks are unstable at the temperature considered here since they can decay into lighter baryons (protons and neutrons). Protons are perfectly stable in the limit of no B violation since they are the lightest baryons. Neutrons can decay into protons through beta decay ($n \longrightarrow p + e^- + \bar{\nu}_e$) but it is possible to show that at the temperature considered here, the inverse process is still efficient (electrons and neutrons carry enough energy for converting a proton into a neutron: this only requires $m_n - m_p = 1.203$ MeV). So, around $\sim$10 MeV, both protons and neutrons are present. They are still maintained in thermal and kinetic equilibrium by weak and electromagnetic interactions. They are of course both non-relativistic since $m_n \sim m_p \sim$ GeV. They have approximately the same density $n_n = n_p$, as will be shown explicitly in the section on nucleosynthesis.

In the lepton sector, μ, $\bar{\mu}$, τ and τ are so heavy that they decay into electrons and positrons. The mass of electrons and positrons is close to 0.5 MeV, so they are still relativistic at that time. Electric neutrality implies $n_{e^-} - n_{e^+} = n_p$. Does this imply a large asymmetry for electrons versus positrons? Remember that n_B/s is conserved and of the order of 10^{-10}. At the temperature considered here, we can consider that $n_B = n_p + n_n \simeq 2n_p$ and that $s \sim n_\gamma \sim n_{e^-}$ modulo factors of order at most ten. Hence, speaking only of orders of magnitude,

$$\frac{n_{e^-} - n_{e^+}}{n_{e^-} + n_{e^+}} \sim \frac{n_{e^-} - n_{e^+}}{s} \sim \frac{n_B}{s} \sim 10^{-10} \; . \tag{3.29}$$

We see that electric neutrality implies that the electron-positron asymmetry is as tiny as the initial baryon asymmetry.

Besides, the universe contains all six neutrinos: ν_e, ν_μ, ν_τ and their antiparticles, maintain in thermal and kinetic equilibrium by weak interactions. Their mass is at most of the order of eV, so they have no reason to annihilate and contribute to the thermal plasma as ultra-relativistic components. They could in principle carry some asymmetry associated to chemical potentials μ_e, μ_μ and μ_τ (each antineutrino would then have an opposite chemical potential due to the chemical equilibrium of the reactions $\nu_e + \bar{\nu}_e \longleftrightarrow e^- + e^+ \longleftrightarrow \gamma$). Due to the large mixing angles in the neutrino mass matrix, the three potentials should share a unique value at this epoch. This issue is still a topic of research, but since such an asymmetry is difficult to motivate and has not been observed so far, we will assume throughout this course that neutrino chemical potentials are null, and hence that at the time considered here all six neutrino species share exactly the same number density.

Finally, the universe should contain photons. All other particles are expected to have decayed by that time, excepted one or more stable "dark matter particle" that we will not be discussed in this course. In summary, around $T \sim 10$ MeV, the universe should contain: p, n, e^-, e^+, six neutrino species, γ and possibly dark matter particles. The latter, if they exist, are expect to be non-relativistic at that time. So the number of relativistic degrees of freedom is given by photons, electrons, positrons and six neutrinos:

$$g_*(\sim 10\text{MeV}) = 2 + \frac{7}{8}(2+2+6) = 10.75 \ . \tag{3.30}$$

3.3.3 Neutrino decoupling

Weak interactions maintain neutrinos in thermal equilibrium through elastic and inelastic interactions like e.g.

$$\begin{aligned}
\nu_e + e^- &\longleftrightarrow \nu_e + e^- && (3.31)\\
\nu_e + \bar{\nu}_e &\longleftrightarrow \nu_i + \bar{\nu}_i \qquad (i = \mu \ \text{ or } \ \tau)\\
\nu_e + \nu_i &\longleftrightarrow \nu_e + \nu_i\\
\nu_e + \bar{\nu}_i &\longleftrightarrow \nu_e + \bar{\nu}_i\\
&\text{etc.} && (3.32)
\end{aligned}$$

which are all of the weak interaction type (they involve exchanges of weak bosons Z^0, $W^\pm$). The thermally averaged cross sections of these reactions are of the order of $\langle \sigma v \rangle \sim G_F^2 T^2$, where $G_F \sim 10^{-5}\text{GeV}^{-2}$ is the Fermi constant (which characterizes the magnitude of weak interactions). Hence the relevant scattering rates are of the order of $\Gamma = n_{e^-}\langle \sigma v \rangle \sim G_F^2 T^5$. Let us compare the evolution of Γ with that of the Hubble rate $H^2 = (8\pi G/3)\rho \sim M_P^{-2} T^4$. We find that

$$\frac{\Gamma}{H} \sim M_P G_F^2 T^3 \sim \left(\frac{T}{1 \text{ MeV}}\right)^3 \ . \tag{3.33}$$

Hence, when the temperature of the plasma drops below $T \sim$ MeV, the neutrinos leave thermal equilibrium, and their distribution remains frozen, with

$$f_i(p) = \frac{1}{\exp[p/T_\nu] + 1} \ . \tag{3.34}$$

By "frozen", one means that f_i varies only due to the universe expansion, which imposes a very trivial evolution. Each decoupled particle is a free-falling in the FLRW universe. The geodesic equation shows that for such particles $p \propto a^{-1}$ (we already used this result many times for photons). Hence each individual particle has a momentum redshifting like $p(t) = p(t_D)a(t_D)/a(t)$ where t_D is the time of decoupling. For particles which decoupled when they were relativistic (like the neutrinos considered in this section), the distribution $f_i(p)$ depends on p only through the ratio p/T_ν. So, saying that all momenta shift like a^{-1} is strictly equivalent to saying that T_ν shifts like a^{-1}. Hence, after neutrino decoupling and for each of the six species i, the product $(T_\nu a)$ remains *exactly* constant at all times. Besides, as long as they remain relativistic with $T_\nu \gg m_{\nu_i}$, they obey to:

$$n_{\nu_i} = \frac{3}{4}\frac{\zeta(3)}{\pi^2}T_\nu^3 \propto a^{-3} , \tag{3.35}$$

$$\rho_{\nu_i} = \frac{7}{8}\frac{\pi^2}{30}g_i T_\nu^4 \propto a^{-4} , \tag{3.36}$$

$$p_{\nu_i} = \frac{1}{3}\rho_{\nu_i} . \tag{3.37}$$

Neutrino decoupling is a very smooth process because before decoupling (and as long as the number of relativistic degrees of freedom g_* was conserved), we already had $T = T_\nu =\propto a^{-1}$, $n_{\nu_i} \propto a^{-3}$, $\rho_{\nu_i} \propto a^{-4}$ and $p_{\nu_i} = \rho_{\nu_i}/3$. Hence, from the point of view of the universe expansion, one could say that "nothing particular happens" when neutrinos decouple. The temperature of neutrinos and of the thermal bath remain equal, both scaling like a^{-1}. The entropy density before decoupling reads:

$$s = \left.\frac{\rho + p}{T}\right|_{\text{plasma}} = \frac{4}{3}\frac{\pi^2}{30}g_* T^3 \qquad \text{with} \quad g_* = 2 + \frac{7}{8}(2+2+6) = 10.75 . \tag{3.38}$$

After decoupling, the entropy receives contribution from the plasma and from neutrinos. We have not derived the expression of entropy for a decoupled relativistic species, but it is simple: it reads alike the entropy of relativistic species in equilibrium, with the appropriate value of the temperature:

$$s = \left.\frac{\rho + p}{T}\right|_{\text{plasma}} + \left.\frac{\rho_\nu + p_\nu}{T_\nu}\right|_{\text{neutrinos}} \tag{3.39}$$

$$= \frac{4}{3}\frac{\pi^2}{30}\left(2 + \frac{7}{8}(2+2)\right)T^3 + \frac{4}{3}\frac{\pi^2}{30}\left(\frac{7}{8}\times 6\right)T_\nu^3 . \tag{3.40}$$

Since both T and T_ν scale like a^{-1} around the time of neutrino decoupling, they remain equal to each other, and the expression of the entropy is absolutely unchanged.

3.3.4 Positron annihilation

The electron and positron mass is close to 0.5 MeV. Hence, when the temperature of the plasma drops below this value, electrons and positron become gradually non-relativistic. This is the same situation as the one described previously for b and $\bar{b}$: the number density of e^- and e^+ drops down very quickly with respect to that of photons, due to the suppression factor $\exp[-m_e/T]$. Basically, this means that electrons and positrons annihilate each other without being recreated, until positrons disappear completely; a small number of electrons survives, in equal proportion to protons in order to ensure electric neutrality. After this process, $n_{e^-} = n_p \sim n_B \sim 10^{-10}n_\gamma$.

It is particularly interesting to follow the evolution of entropy during electron-positron annihilation. Intuitively, entropy conservation implies that when electrons and positrons annihilate each other, their entropy has to be mediated to other species, namely: photons, which are the only remaining relativistic species in the plasma. In other words, the reaction $e^- + e^+ \longrightarrow$ generates an excess of photons; since photons are in thermal equilibrium, any excess in the number density must be described in terms of an increase in the product (Ta). Let us check this explicitly. Before positron annihilation, the expression of entropy is given by Eq. (3.40). After annihilation, it reads:

$$s = \left.\frac{\rho + p}{T}\right|_{\text{plasma}} + \left.\frac{\rho_\nu + p_\nu}{T_\nu}\right|_{\text{neutrinos}} \tag{3.41}$$

$$= \frac{4}{3}\frac{\pi^2}{30}(2)\,T^3 + \frac{4}{3}\frac{\pi^2}{30}\left(\frac{7}{8}\times 6\right)T_\nu^3 \,. \tag{3.42}$$

Note that the total entropy in a comoving volume sa^3 is conserved, but the separate entropy of neutrinos is also conserved since they are decoupled and $(T_\nu a)$ is exactly constant. This implies that $s_{\text{plasma}}a^3$ is also conserved separately. Hence:

$$\frac{11}{2}(Ta)^3_{\text{before}} = 2(Ta)^3_{\text{after}} \,. \tag{3.43}$$

We conclude that the temperature of the plasma does not scale like a^{-1} during electron positron annihilation: this is a typical example in which it is rescaled according to $g_*^{-1/3}$. In fact, Ta increases in order to compensate the loss of the electron and positron degrees of freedom. But the most interesting outcome of this is that the temperature of photons and neutrinos after annihilation differs by:

$$\frac{(T_\nu a)_{\text{after}}}{(Ta)_{\text{after}}} = \frac{(T_\nu a)_{\text{before}}}{(11/4)^{1/3}(Ta)_{\text{before}}} = \left(\frac{4}{11}\right)^{1/3} \,. \tag{3.44}$$

After positron annihilation, the photons are the only remaining species in thermal equilibrium, hence $g_* = 2$ and (Ta) is exactly constant. Finally, we will see that photons decouple around $T \sim 0.3$ eV. Like for neutrinos, the distribution of photons remains frozen after decoupling, with $T(t) = T(t_D)a(t_D)/a(t)$ until today. We conclude that between $T \sim 0.5$ MeV and today, the relation $T_\nu = (4/11)^{1/3}T$ holds at any time, with the photon temperature given by $T = T_0(a_0/a)$. Here, T_0 is the CMB temperature measured today, $T_0 = 2.726$K. So $T_{\nu 0} = 1.946$K. Knowing the photon and neutrino temperature today, we can infer their number density:

$$n_\gamma^0 = \frac{\zeta(3)}{\pi^2}\times 2\,T_0^3 = 137\ \text{cm}^{-3} \,, \tag{3.45}$$

$$n_\nu^0 = \frac{\zeta(3)}{\pi^2}\times\frac{3}{4}\times 6\times\frac{4}{11}\,T_0^3 = 112\ \text{cm}^{-3} \,. \tag{3.46}$$

(the second number being the total density summed over the six neutrinos).

3.3.5 nucleosynthesis

A nucleus X containing Z protons can have various isotopes ${}^AX$ of mass number A (hence containing $A - Z$ neutrons). The following reactions can increase Z by one unit, starting from a simple proton (i.e. ionized hydrogen nucleus $H^+ = p$; in the following we will omit to write the charge of the various ions):

$$p + n \longrightarrow D + \gamma \tag{3.47}$$

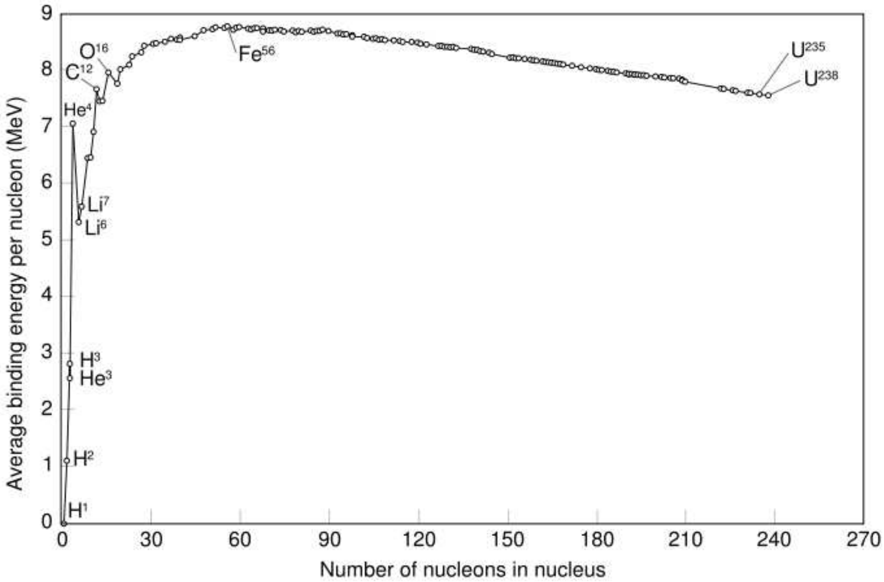

Figure 3.2: Average binding energy per nucleon B/A as a function of A.

$$D + D \longrightarrow {}^3He + n \quad (3.48)$$
$${}^3He + D \longrightarrow {}^4He + p \quad (3.49)$$
$$\ldots \quad (3.50)$$

In order to know whether these reactions are favored or not from the point of view of energetics, we should know the binding energy B of each element. We recall that the binding energy is the minimal amount of energy which must be furnished in order to break a nucleus X in Z protons and $A - Z$ neutrons. Hence the rest energy of X reads:

$$E_0(X) = m_X = Zm_p + (A - Z)m_n - B \ . \quad (3.51)$$

For instance, the binding energy of deuterium is $B_D = 2.22$ MeV, since $m_p = 938.27$ MeV, $m_n = 939.57$ MeV, $m_p + m_n = 1877.84$ MeV and $m_D = 1875.62$ MeV. Hence, from a purely energetic point of view, protons and neutrons should combine and form the isotope with the largest possible binding energy per nucleon B/A: once this isotope exists, any nuclear reaction destroying it would cost energy. Figure 3.2 shows the average binding energy per nucleon B/A as a function of A. Starting from zero for hydrogen ${}^1H$ (p), the curve raises for deuterium ${}^2H$ (pn), helium ${}^3He$ (ppn), tritium ${}^3H$ (pnn), and reaches a local maximum for ${}^4He$ ($ppnn$). The first isotope with a ratio B/A larger than that of ${}^4He$ is ${}^{12}C$. The global maximum is reached at $A = 56$ for iron ${}^{56}Fe$.

Preliminary overview of nucleosynthesis. From a purely energetic point of view, we could expect the following picture. The reaction

$$D + \gamma \longrightarrow p + n \quad (3.52)$$

requires an energy of at least $B_D = 2.22$ MeV. For $T > B_D$, photons carry enough energy for breaking any deuterium nucleus into pairs $p + n$.

Hence, protons and neutrons can be significantly converted into deuterium only when the temperature drops below B_D. Once deuterium forms, it is energetically more favorable to convert it in ^3He, and so on and so on, until the universe contains only heavy elements like iron.

In the above reasoning, we forgot that the kinetic of the various reactions involved does not depend only on initial and final energies, but also on number densities and cross sections. In fact, the previous reasoning is more or less correct in the frame of the Cold Big Bang scenario, which was rejected on this basis: far from stars, the real universe seems dominated by hydrogen rather than heavy elements. In the Hot Big Bang scenario, a key feature is that baryons are considerably suppressed with respect to photons, $n_B \sim 10^{-10} n_\gamma$. So, our argument that when $T < B_D$ the reaction (3.52) cannot occur is wrong. There are so many photons that even if the average photon energy is much less than T_D, but a tiny fraction of them (of order 10^{-10}) have a momentum larger than B_D (which is possible if they are in the high-momentum tail of the Fermi-Dirac distribution), then the reaction is still very efficient. So, in the Hot Big Bang scenario, neutrons and protons start forming deuterium at a significantly *smaller* temperature than B_D. The formation of heavier elements is also suppressed by consideration on number densities. Once deuterium is formed, most of it is efficiently converted into ^4He as could be expected from energetics, but then the gap between ^4He and ^{12}C is very difficult to cross: it requires a three-body reaction $3 \times {}^4He \longrightarrow {}^{12}C$. When ^4He forms, the temperature is far too low for the scattering rate of the above reaction to be comparable with H. Hence the chain will stop at ^4He. Let us now check these qualitative expectations using our knowledge of thermal and chemical equilibrium. The discussion can be carried in two steps.

Formation of Deuterium. We first study the reaction of deuterium formation:

$$n + p \longleftrightarrow D + \gamma \ . \tag{3.53}$$

The cross-section of this reaction is large enough for ensuring chemical equilibrium in the temperature range considered here. Hence $\mu_D = \mu_n + \mu_p$. At $T \ll$ GeV, neutrons, protons and deuterium are all non-relativistic with densities given by Eq. (3.10). Hence

$$\frac{n_D}{n_p n_n} = \exp\left(\frac{\mu_D - \mu_p - \mu_n}{T}\right) \frac{3}{4} \left(\frac{2\pi m_D}{m_p m_n T}\right)^{3/2} \exp\left(\frac{m_p + m_n - m_D}{T}\right) \ , \tag{3.54}$$

where we used the number of spin states: $g = 2$ for p and n, $g = 3$ for deuterium. The argument of the first exponential cancels because of chemical equilibrium. The argument of the second one involves the binding energy B_D of deuterium:

$$\frac{n_D}{n_p n_n} = \frac{3}{4} \left(\frac{2\pi m_D}{m_p m_n T}\right)^{3/2} \exp\left(\frac{B_D}{T}\right) \ . \tag{3.55}$$

We will now use this equation for getting a rough estimate of the order of magnitude of the deuterium density to baryon number density ratio. We know that roughly, $n_p \sim n_n \sim n_B \sim 10^{-10} n_\gamma \sim 10^{-10} T^3$. Hence we obtain

$$\begin{aligned} \frac{n_D}{n_B} &\sim 10^{-10} \left(\frac{T}{m_p}\right)^{3/2} \exp\left(\frac{B_D}{T}\right) \\ &\sim 10^{-10} \left(\frac{T}{0.94 \text{ GeV}}\right)^{3/2} \exp\left(\frac{2.22 \text{ MeV}}{T}\right) \ . \end{aligned} \tag{3.56}$$

As long as $T > B_D$, it is clear that the ratio remains tiny. As expected, there is no significant deuterium abundance above that scale; all baryons are in the form of neutrons and protons. The first terms can be compensated only if the argument of the exponential is large enough. A quick estimate shows that for $T \sim 0.06$ MeV, the above ratio reaches the order of one. A more careful estimate shows that the deuterium abundance becomes sizable around 0.07 MeV. We will retain 0.07 MeV as the temperature of nucleosynthesis.

Once deuterium forms, one can show that it is efficiently converted to ^3He and ^4He, since the scattering rate of the relevant reactions exceeds the Hubble rate, and ^4He is the most stable configuration. However, at $T \sim$ 0.07 MeV, the scattering rate of the three-body reaction $3 \times {}^4He \longrightarrow {}^{12}C$ is considerably suppressed and the chain stops. We conclude that below $T \sim$ 0.07 MeV, nucleons combine into ^4He, which is formed of two protons and two neutrons. However, protons and neutrons are not necessarily in exactly equal proportions before this temperature is reached. Hence, together with ^4He, there might be a relic density of protons or neutron. We see that it is crucial to compute the neutron over proton ratio for $T \geq 0.07$ MeV.

Neutron versus proton density above $T \sim 0.07$ MeV. The balance between neutrons and protons depends essentially on the reaction:

$$p + e^- \longleftrightarrow n + \nu_e \ . \tag{3.57}$$

At high energy ($T >$ MeV), this reaction is in chemical equilibrium, with $\mu_p + \mu_e = \mu_n + \mu_{\nu_e}$. The chemical potential of neutrinos is zero in the simplest cosmological model considered in this course. The one of electrons is non-zero, but before electron-positron annihilation the asymmetry between electrons and positrons is so small ($\mu_e/T \sim 10^{-10}$) that we can work in the approximation $\mu_e \simeq 0$. Hence:

$$\begin{aligned} 1 &= \exp\left(\frac{\mu_p + \mu_e - \mu_n - \mu_{\nu_e}}{T}\right) \\ &\simeq \exp\left(\frac{\mu_p - \mu_n}{T}\right) \\ &= \frac{n_p}{n_n}\left(\frac{m_n}{m_p}\right)^{3/2} \exp\left(\frac{m_p - m_n}{T}\right) \ . \end{aligned} \tag{3.58}$$

(for the last equality, we used Eq. (3.10) for the number density of non-relativistic species). The difference between the neutron and proton mass is $Q \equiv m_n - m_p = 1.203$ MeV. Hence, for $T \gg 1$ MeV, the neutron to proton ratio is given by:

$$\left.\frac{n_n}{n_p}\right|_{T \gg 1 \text{ MeV}} = (m_n/m_p)^{3/2} = 1.002 \ , \tag{3.59}$$

i.e. the density of neutrons and protons is essentially the same. When $T \sim 1$ MeV, chemical equilibrium would force the neutron to proton ratio to drop exponentially like $\exp(-Q/T)$. If this was true, at 0.07 MeV there would be essentially no neutron left, and nucleosynthesis would not happen: the primordial universe would contain only hydrogen.

However, the above reaction is mediated by weak interactions. Hence, it becomes quite weak around $T \sim$MeV, and we are forced to consider its departure from chemical equilibrium. In fact we will see that the reaction freezes out with a significant leftover of neutrons. The neutron density obeys to the Boltzmann equation:

$$\dot{n}_n + 3Hn_n = n_n[n_{\nu_e}\langle\sigma v\rangle]\left\{\exp\left(\frac{\mu_e+\mu_p-\mu_n-\mu_{\nu_e}}{T}\right)-1\right\} . \tag{3.60}$$

The term between brackets is the scattering rate Γ_{np} for neutron to proton conversion, and the exponential can be approximated using Eq. (3.58). Hence

$$\dot{n}_n + 3Hn_n = n_n\Gamma_{np}\left\{\frac{n_p}{n_n}\left(\frac{m_n}{m_p}\right)^{3/2}e^{-Q/T}-1\right\} . \tag{3.61}$$

This equation can be written in terms of a dimensionless variable, the neutron fraction $X_n = n_n/(n_n+n_p)$. We have

$$n_n = X_n(n_n+n_p) = X_n n_B, \qquad n_p = (1-X_n)n_B . \tag{3.62}$$

The conservation of the baryon number implies $n_B \propto a^{-3}$, so

$$\dot{n}_n = \dot{X}_n n_B - 3HX_n n_B . \tag{3.63}$$

Replacing n_n and n_p in Eq. (3.61) and dividing by n_B, we get

$$\dot{X}_n = \Gamma_{np}\left[(1-X_n)\,e^{-Q/T}-X_n\right] . \tag{3.64}$$

The dependence of Γ_{np} with respect to T can be computed using nuclear physics. Still, in order to integrate the equation, we need to know the relation between time t and temperature T. This relation can be inferred from the Friedmann equation. In first approximation, $T \propto a^{-1}$ (neglecting the effect of the electron-positron annihilation on Ta) and $dT/T = -da/a$. So,

$$\frac{dT}{dt} = -T\frac{da}{a\,dt} = -TH \tag{3.65}$$

$$= -\sqrt{\frac{8\pi G}{3}\rho T^2} \tag{3.66}$$

$$= -\sqrt{\frac{8\pi^3 G}{90}g_* T^6} \tag{3.67}$$

with $g_* = 10.75$ before electron-positron annihilation. Hence the reaction reads

$$\frac{dX_n}{dT} = -\sqrt{\frac{90}{8\pi^3 g_*}}\frac{M_P}{T^3}\Gamma_{np}(T)\left[(1-X_n)e^{-Q/T}-X_n\right] . \tag{3.68}$$

Knowing $\Gamma_{np}(T)$, this equation can be integrated. The result is that around $T \sim 0.1$ MeV, X_n gets close to an asymptotic value of 0.15, corresponding to the freeze-out of the neutron to proton ratio.

Equation (3.68) is just a first-order approximation. The precise calculation includes two additional effects: the change in g_* and Ta due to the electron-positron annihilation, and the neutron beta-decay ($n \longrightarrow p + e^- + \bar{\nu}_e$) which should be included in the right-hand side of the Boltzmann equation since it represents another decay channel. Altogether, these effects lead to a slightly different neutron to proton ratio at freeze-out, $X_n(T < 0.1\text{ MeV}) \sim 0.11$. Hence, at $T \sim 0.07$ MeV, $n_n = 0.11\,n_B$ and $n_p = 0.89 n_B$. At $T \sim 0.07$eV, all available neutrons will combine into deuterium, ^3He and finally ^4He nuclei, together with the same number of protons. The final ^4He density should be $n_{^4He} = 0.055\, n_B$, with a leftover of $n_H = 0.78\, n_B$ protons. The helium fraction, usually defined as:

$$Y_P \equiv \frac{4 n_{^4He}}{n_B} , \tag{3.69}$$

is predicted to be 0.22 at any time after nucleosynthesis, in every region of the universe not affected by the ejection of particles from stars (since inside stars, nuclear reactions can form other elements in very different proportions).

Exact results from a full calculation. The above calculation was rather simplistic. A full simulations of nucleosynthesis can be performed using numerical codes (a few nucleosynthesis codes are even publicly available). Instead of studying the kinetics of just two reactions, these codes follow of the order of one hundred possible reactions between neutrons, protons and heavier nuclei (typically, till ^{12}C). The main differences between the outcome of a full simulation and the results of the above section are:

- when reactions freeze-out, the density n_i of other elements than ^4He is nonzero - but still very small: the number density of D and ^3He is smaller than that of ^4He by a factor $\sim 10^5$, the density of ^7Li is smaller by $\sim 10^9$, and all other species are even more suppressed.

- the final helium fraction depends slightly on the free parameter of this problem, namely $\eta_b \sim 10^{-10}$, which controls mainly the temperature at which deuterium starts forming (see Eq. (3.56)). Hence the neutron-to-proton ratio at the beginning of deuterium formation depends on η_b, as well as the final helium abundance. However the dependence is only logarithmic. Precise simulations yield

$$Y_P = 0.2262 + 0.0135 \ln(\eta_b/10^{-10}) . \tag{3.70}$$

  This result is in very good agreement with the approximate calculation presented before.

3.3.6 Recombination

After nucleosynthesis, the universe contains a thermal plasma composed essentially of relativistic photons and non-relativistic electrons, hydrogen nuclei and helium nuclei; plus decoupled relativistic neutrinos. At $T \ll$ MeV, weak interactions are inefficient, but electromagnetic interactions ensure equilibrium between electrons, nuclei and photons. More precisely, photons remain tightly coupled to electrons via Compton scattering ($e^- + \gamma \longrightarrow e^- + \gamma$) and electrons to nuclei via Coulomb scattering ($e^- + p \longrightarrow e^- + p$ or $e^- + {}^4He \longrightarrow e^- + {}^4He$). These interactions are efficient at least as long as hydrogen and helium remain ionized.

The goal of this section is precisely to establish until which time hydrogen remains ionized (a time called recombination). For simplicity, we will not consider helium recombination, and even neglect completely the presence of helium (although in reality $\sim 22\%$ of nucleons are inside ^4He nuclei). The formation of neutral hydrogen depends on the reaction:

$$e^- + p \longleftrightarrow H + \gamma . \tag{3.71}$$

Like for nucleosynthesis, let us start from purely energetic considerations. The binding energy of hydrogen, defined through:

$$m_H = m_p + m_e - \epsilon_0 , \tag{3.72}$$

is equal to $\epsilon_0 = 13.6$ eV. Hence we expect that for $T \gg 13.6$ eV hydrogen is fully ionized: any neutral hydrogen atom would immediately interact with an energetic photon and get ionized. This does not mean that neutral hydrogen forms immediately below $T \sim 13.6$ eV. Just like for the formation of deuterium during nucleosynthesis, the balance of the above reaction depends on relative abundances. We know that the density of electrons and protons is 10^{10} smaller than that of photons. So, much below $T \sim 13.6$ eV, there should still be enough energetic photons for preventing recombination. This can be checked assuming (as a first step) chemical equilibrium for the above reaction:

$$\mu_{e^-} + \mu_p = \mu_H \ . \tag{3.73}$$

Now, let us write the ratio of densities, using Eq. (3.10) with $g = 2$ for e^- and p, and $g = 4$ for H:

$$\frac{n_e n_p}{n_H} = \exp\left(\frac{\mu_e + \mu_p - \mu_H}{T}\right)\left(\frac{m_e m_p T}{m_H 2\pi}\right)^{3/2} \exp\left(\frac{-m_e - m_p + m_H}{T}\right) \ . \tag{3.74}$$

Using the chemical equilibrium relation (3.73) and approximating the ratio m_p/m_H by one, we get

$$\frac{n_e n_p}{n_H} = \left(\frac{m_e T}{2\pi}\right)^{3/2} e^{-\epsilon_0/T} \ . \tag{3.75}$$

Since in this section we decided to neglect the presence of helium, electric neutrality ensures that $n_e = n_p$. Hence the free electron fraction is equal to the hydrogen ionization fraction:

$$X_e \equiv \frac{n_e}{n_e + n_H} = \frac{n_p}{n_p + n_H} \ . \tag{3.76}$$

Replacing n_e, n_p and n_H in Eq. (3.75), we get:

$$\frac{X_e^2}{1 - X_e} = \frac{1}{n_e + n_H}\left(\frac{m_e T}{2\pi}\right)^{3/2} e^{-\epsilon_0/T} \ . \tag{3.77}$$

This equation (based on the assumption of chemical equilibrium) is called the Saha equation. Since we neglect helium, the sum $n_e + n_H = n_p + n_H$ is equal to the baryon number density, $n_B \sim 10^{-10} n_\gamma \sim 10^{-10} T^3$. Hence:

$$\frac{X_e^2}{1 - X_e} \sim 10^{10}\left(\frac{m_e}{T}\right)^{3/2} e^{-\epsilon_0/T} \sim 10^{10}\left(\frac{0.5 \text{ MeV}}{T}\right)^{3/2} e^{-13.6 \text{ eV}/T} \ . \tag{3.78}$$

So, for $T \sim 13.6$ eV, one has $X_e^2/(1 - X_e) \sim 10^{17}$ and hence $X_e \simeq 1$: hydrogen is still fully ionized. It is only for $T \ll 13.6$ eV that the exponential can compensate the first factors in such way that X_e drops below one. An estimate shows that ionization can take place only below $T \sim 0.25$ eV. Let us translate this in terms of redshift. We know that $T_0 = 2.726 \text{ K} = 2.726 \times 8.617 \times 10^{-5}$ eV in our $k_B = c = \hbar = 1$ units. The temperature scales like a^{-1}, i.e. like $(1 + z)$. Hence recombination takes place around:

$$z_{\rm rec} \sim \frac{0.25}{2.726 \times 8.617 \times 10^{-5}} - 1 \sim 1100 \ . \tag{3.79}$$

The Saha equation (3.75) is successful in predicting roughly the temperature at which recombination takes place, but it does not provide an accurate description of the evolution of the ionization fraction X_e. Indeed, in the real universe, chemical equilibrium cannot be maintained throughout recombination between the four species e, p, H and γ. Hence we must employ the Boltzmann equation in order to follow the kinetics of recombination.

The exact description of recombination is considerably complicated by the fact that hydrogen can form in various excited states, and then relax to its fundamental state while emitting photons: so, there are many states and reactions to follow. In this course, we neglect these issues and do as if there was a unique hydrogen state, forming through $e^- + p \longrightarrow H + \gamma$. Considering only this reaction, the Boltzmann equation for the free electron density reads:

$$\dot{n_e} + 3Hn_e = n_e n_p \langle \sigma v \rangle \left\{ \exp\left(\frac{\mu_H - \mu_e - \mu_p}{T} \right) - 1 \right\} . \tag{3.80}$$

Using Eq. (3.74) and replacing $n_e = n_p$ and n_H in terms of $X_e = n_e/(n_e + n_H) = n_e/n_B$, we get:

$$\dot{X}_e = \langle \sigma v \rangle \left[(1 - X_e) \left(\frac{m_e T}{2\pi} \right)^{3/2} e^{-\epsilon_0/T} - X_e^2 n_B \right] . \tag{3.81}$$

The cross section and baryon number density can be computed as a function of temperature. The derivative with respect to t on the left-hand side can be replace by a derivative with respect to T using:

$$\frac{dT}{dt} = -TH = -\sqrt{\frac{8\pi G}{3}\rho T^2} . \tag{3.82}$$

In the total density ρ, one should sum over both relativistic and non-relativistic components ($\rho = \rho_R + \rho_M$), because recombination takes place close to the time of equality between matter and radiation. To see this, let us admit that the non-relativistic matter fraction today is of the order of $\Omega_m \sim 0.25$, as will be shown in the chapter on cosmological observations. Hence, the scale factor at equality obtains from equating the relativistic density:

$$\rho_R = \frac{\pi^2}{30} g_* T^4 = \frac{\pi^2}{30} g_* T_0^4 (a_0/a)^4 \tag{3.83}$$

(with $g_* = 2 + \frac{7}{8} \times 6 \times \left(\frac{4}{11}\right)^{4/3}$ for photons and neutrinos) with the non-relativistic density

$$\rho_M = \Omega_M \rho_c^0 (a_0/a)^3 = \Omega_M \rho_c^0 (T/T_0)^3 = \Omega_M \frac{3H_0^2}{8\pi G} (a_0/a)^3 . \tag{3.84}$$

So the scale factor at equality is given by

$$\frac{a_{\rm eq}}{a_0} = \frac{\pi^2}{30} g_* T_0^4 \frac{8\pi}{3 M_P^2 H_0^2} \Omega_M^{-1} . \tag{3.85}$$

For $h = 0.7$ and $\Omega_M = 0.25$ this gives $a_{\rm eq} = 3.3 \times 10^{-4}$ and $z_{\rm eq} = 3000$. So, recombination takes place slightly after matter-radiation equality.

With all these consideration, equation (3.81) can in principle be integrated numerically with respect to T. After doing so, one would find that the ionization fraction X_e becomes significantly smaller than one around $z \sim 1400$, crosses $X_e \sim 0.1$ around $z \sim 1000$, and tends to an asymptotic freeze-out value of order $X_e \to 5 \sim 10^{-4}$ for $z < 100$.

3.3.7 Photon decoupling

Till the time of recombination, photons are maintained in thermal equilibrium mainly through Compton scattering off electrons:

$$\gamma + e^- \longrightarrow \gamma + e^- \ . \tag{3.86}$$

The cross section $\langle \sigma v \rangle$ of the above reaction is the Thomson cross section, equal to $\langle \sigma v \rangle_T = 0.665 \times 10^{-24}\text{cm}^2$. Compton scattering of photons off electrons becomes inefficient roughly when the scattering rate $\Gamma = n_e \langle \sigma v \rangle_T$ equals the Hubble parameter. In order to evaluate this characteristic time, we can write $n_e = n_p = X_e n_B$ (like in the previous section, we neglect helium) and $n_B \sim \rho_b / m_p$. We obtain:

$$\frac{\Gamma}{H} = 0.07 (a_0/a)^3 X_e \Omega_b h \frac{H_0}{H} \ . \tag{3.87}$$

The Hubble rate in units of Hubble rate today can be estimated to be $H/H_0 = \Omega_M^{1/2} (a_0/a)^{3/2}$ during matter domination. Taking a to be of the order of $a_{\rm dec}$, $\Omega_b \simeq 0.4$, $\Omega_M \simeq 0.25$ and $h \simeq 0.7$, we see that photon decoupling occurs when X_e drops below $\sim 10^{-2}$ during recombination. Hence, recombination directly triggers photon decoupling. This is in fact the main reason for which recombination is important to study: it controls the decoupling of the CMB photon that we observe today. The details of recombination affect CMB anisotropies patterns. However, the temperature evolution of photons is completely unaffected by their decoupling, exactly like for neutrinos. When photons decouple, their relativistic Bose-Einstein distribution freezes-out, and only evolves at later times due to the universe expansion, which induces $p \propto a^{-1}$ and hence $T \propto a^{-1}$.

A precise calculation shows that photon decoupling arises mainly around a redshift $z_{\rm dec} = 1100$. Translating in terms of proper time, one finds photons decouple approximately 380,000 years after the initial singularity.

3.3.8 Very recent stages

From the point of view of the thermal history of the universe, very few phenomena occur after photon decoupling. Each neutrino family i becomes non-relativistic when $T_\nu < m_i$, but since they are already decoupled, this has no effect on the temperature and number density evolution ($T_\nu \propto a^{-1}$ and $n_\nu \propto a^{-3}$). Only the density and pressure of neutrinos are affected by the non-relativistic transition. The consequences of this transition on structure formation are not discussed in this course (they shall be explained next semester by Pierre Salati).

There is however another important phenomenon occurring at low redshift, than we just mention here briefly. When the first stars form, they emit a new population of photons which partially reionize hydrogen and heavier elements. However, this reionization is not sufficient for "re-coupling" photons to electrons and ionized matter: only a small fraction of CMB photons have a chance to experience Compton scattering between the time of decoupling and today.

In figure 3.3, we summarize qualitatively the main results of this section.

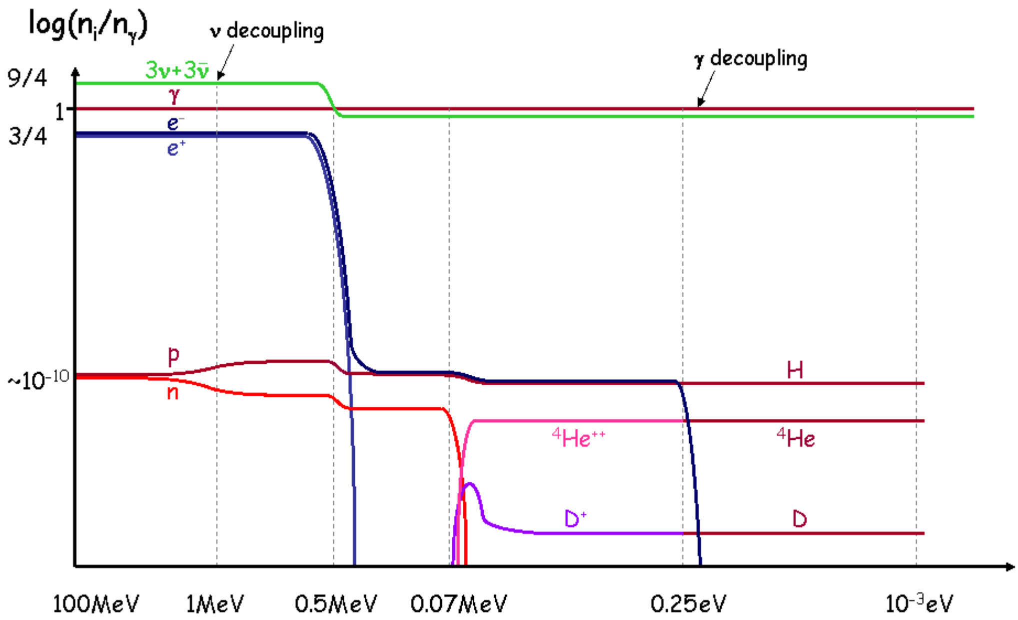

Figure 3.3: As a summary of Chapter 3, we show the qualitative evolution of n_i for each species, normalized in terms of n_γ.

Chapter 4

Dark Matter

We have many reasons to believe that in the recent universe, the non-relativistic matter is of two kinds: ordinary matter, and dark matter. One of the well-known evidences for dark matter arises from galaxy rotation curves.

Inside galaxies, the stars orbit around the center. If we can measure the redshift in different points inside a given galaxy, we can reconstruct the distribution of velocity $v(r)$ as a function of the distance r to the center. It is also possible to measure the distribution of luminosity $I(r)$ in the same galaxy. What is not directly observable is the mass distribution $\rho(r)$. However, it is reasonable to assume that the mass distribution of the *observed luminous matter* is proportional to the luminosity distribution: $\rho_{\text{lum}}(r) = b\ I(r)$, where b is an unknown coefficient of proportionality called the bias. From this, we can compute the gravitational potential Φ_{lum} generated by the luminous matter, and the corresponding orbital velocity, given by ordinary Newtonian mechanics:

$$\rho_{\text{lum}}(r) = b\ I(r), \tag{4.1}$$

$$\Delta\Phi_{\text{lum}}(r) = 4\pi\mathcal{G}\ \rho_{\text{lum}}(r), \tag{4.2}$$

$$v_{\text{lum}}^2(r) = r\frac{\partial}{\partial r}\Phi_{\text{lum}}(r). \tag{4.3}$$

So, $v_{\text{lum}}(r)$ is known up to an arbitrary normalization factor $\sqrt{b}$. However, for many galaxies, even by varying b, it is impossible to obtain a rough agreement between $v(r)$ and $v_{\text{lum}}(r)$ (see figure 2.3). The stars rotate faster than expected at large radius. We conclude that there is some non–luminous matter, which deepens the potential well of the galaxy.

A qualitatively similar argument applies to the dynamics of galaxies within galaxy clusters. Actually, the hypothesis of dark matter was formulated for the first time by Franz Zwicky in 1933, following the observation of surprisingly large galaxy velocities inside the Coma galaxy cluster.

Apart from galactic rotation curves, there are many arguments – of more cosmological nature – which imply the presence of a large amount of non–luminous matter in the universe, called dark matter. For various reasons (in particular, the distribution of CMB anisotropies, to be studied in Chapter 5 and 6), it cannot consist in ordinary matter that would remain invisible just because it is not lighten up. Dark matter has to be composed of particle that are intrinsically uncoupled with photons – unlike ordinary matter, made up of baryons. Within the standard model of particle physics, a good candidate for non-baryonic dark matter would be a neutrino with a small mass. Then, dark matter would become non-relativistic only recently,

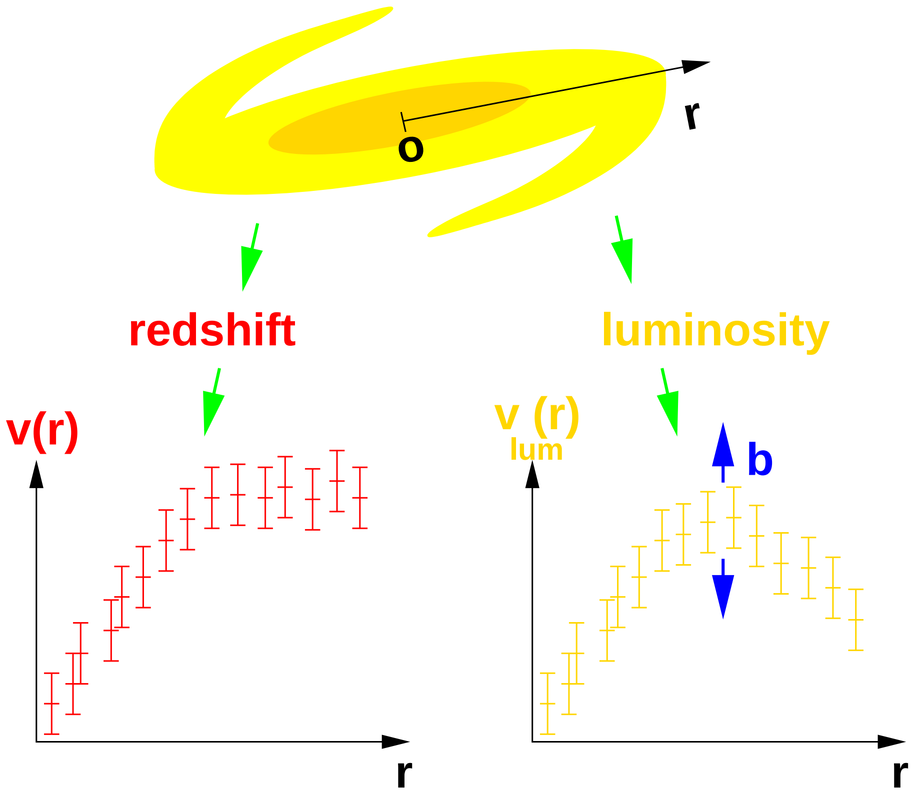

Figure 4.1: A sketchy view of the galaxy rotation curve issue. The genuine orbital velocity of the stars is measured directly from the redshift. From the luminosity distribution, we can reconstruct the orbital velocity under the assumption that all the mass in the galaxy arises form of the observed luminous matter. Even by varying the unknown normalization parameter b, it is impossible to obtain an agreement between the two curves: their shapes are different, with the reconstructed velocity decreasing faster with r than the genuine velocity. So, there has to be some non–luminous matter around, deepening the potential well of the galaxy.

and would still possess today large velocities, just a few orders of magnitude smaller than the speed of light (this hypothesis is called Hot Dark Matter or HDM). However, HDM is excluded by some types of observations: dark matter particles have to be deep inside the non-relativistic regime, otherwise galaxy could not form during matter domination. Strongly non–relativistic dark matter is generally called Cold Dark Matter (CDM).

There are a few candidates for CDM in various extensions of the standard model of particle physics: for instance, some supersymmetric partners of gauge bosons (like the neutralino or the gravitino), or the axion of the Peccei-Quinn symmetry. Despite many efforts, these particles have never been observed directly in the laboratory. This is not completely surprising, given that they are – by definition – very weakly coupled to ordinary particles. However, there are still many efforts for direct detection of dark matter particles in underground laboratories, and a discovery might occur in the next years.

We will not discuss dark matter any further, since it will be the main topic of the course by Pierre Salati, next semester. In the following, we will decompose $\Omega_{\rm M}$ in $\Omega_{\rm B}+\Omega_{\rm CDM}$. This introduces one more cosmological parameter. With this last ingredient, we have described the main features of the Standard Cosmological Model, at the level of homogeneous quantities. We will now focus on the perturbations of this background.

Chapter 5

Cosmological perturbations

In all this chapter, we will study the evolution of cosmological perturbations under the assumption that the universe is flat: this simplifies all equations considerably.

5.1 Formalism

5.1.1 Definition of perturbations, gauge transformations

In our universe, the metric and the energy-momentum tensor are inhomogeneous. Their perturbations, given by

$$\delta g_{\mu\nu}(\mathbf{x},t) = g_{\mu\nu}(\mathbf{x},t) - \bar{g}_{\mu\nu}(t) \ , \tag{5.1}$$
$$\delta T_{\mu\nu}(\mathbf{x},t) = T_{\mu\nu}(\mathbf{x},t) - \bar{T}_{\mu\nu}(t) \ , \tag{5.2}$$

are known to be small in the early universe, typically 10^5 times smaller than the background quantities, as shown by CMB anisotropies. As we shall see in this section, after photon decoupling, the matter perturbations grow by gravitational collapse and reach the non-linear regime, starting with the smallest scales. However, the linear perturbation theory is a good tool both for describing the early universe at any scales, and the recent universe on the largest scales. The most reliable observations in cosmology are those involving mainly linear (or quasi-linear) perturbations. For instance, current constraints on the spatial curvature, baryon and dark matter density based on observations of CMB anisotropies involve only linear cosmological perturbations. Therefore, our goal in this section is to describe the evolution of cosmological perturbations in the linear regime. The great advantage of linear theory is, as usual, to obtain independent equations of evolution for each Fourier mode. Note that the Fourier decomposition must be performed with respect to the comoving coordinate system: so, the quantity $(2\pi/k)$ is the *comoving wavelength* of a perturbation of wavevector $\mathbf{k}$, while the *physical wavelength* is given by

$$\lambda(t) = a(t)\frac{2\pi}{k} \ , \tag{5.3}$$

where $a(t)$ is the scale factor. For each mode $\mathbf{k}$, the amplitude of each perturbation evolves under an equation of motion (which depends only on the modulus k, since the background is isotropic), and on top of this evolution, the physical wavelength is stretched according to the universe expansion.

The perturbations defined in Eqs. (5.1) contain many degrees of freedom: the homogeneity and isotropy of the background implies that $\bar{g}_{\mu\nu}$ and $\bar{T}_{\mu\nu}$ are diagonal, but in general this is not true at the level of perturbations. However, we will see that some of these degrees of freedom are just artifacts of the relativistic perturbation theory set-up; moreover, only a fraction of the physical degrees of freedom contribute to quantities which are actually observable in the cosmic microwave background (CMB) and large scale structure (LSS) of the universe. Hence, the problem can be reduced to the integration of a small number of linear equations of evolution.

In the real universe all physical quantities (densities, curvature...) are functions of time and space. Thanks to the covariance of general relativity, they can be described in principle in any coordinate system, without changing the physical predictions. The problem is that in order to obtain simple equations of evolution, we wish to use a linear perturbation theory, in which the true physical quantities are artificially decomposed into a homogeneous background and some small perturbations. This is artificial because the homogeneous quantities are defined as spatial averages over hypersurfaces of simultaneity: $\bar{f}(t) = \langle f(t, \mathbf{x}) \rangle_{\mathbf{x}}$. Any change of coordinate system which:

1. mixes time and space (therefore, redefining hypersurfaces of simultaneity, and changing the way to perform spatial averages), and

2. remains small everywhere, so that the differences between true quantities and spatial averages are still small perturbations,

gives a new set of perturbations (new equations of evolution, new initial conditions), although the physical quantities (i.e., the total ones) are the same. This ambiguity is called the *gauge freedom* in the context of relativistic perturbation theory.

Of course, using a linear perturbation theory is only possible when there exists at least one system of coordinates in which the universe looks approximately homogeneous. We know that this is the case at least until the time of photon decoupling: in some reference frames, the CMB anisotropies do appear as small perturbations. It is a necessary condition for using linear theory to be in such a frame; however, this condition is vague and leaves a lot of gauge freedom, i.e. many possible ways to slice the spacetime into hypersurfaces of simultaneity.

We can also notice that the definition of hypersurfaces of simultaneity is not ambiguous at small distances, as long as different observers can exchange light signals in order to synchronize their clocks. Intuitively, the gauge freedom is an infrared problem, since on very large distances (larger than the Hubble distance) the word "simultaneous" does not have a clear meaning. The fact that the gauge ambiguity is only present on large scales emerges naturally from the mathematical framework describing gauge transformations.

Formally, a gauge transformation is described by a quadrivector field $\epsilon^\mu(\mathbf{x}, t)$. When the latter is infinitesimal, the Lorentz scalars, vectors and tensors describing the perturbations are shifted by the Lie derivative along ϵ^μ,

$$\delta A_{\mu\nu\ldots}(\mathbf{x}, t) \to \delta A_{\mu\nu\ldots}(\mathbf{x}, t) + \mathcal{L}_{\epsilon^\mu} [\delta A_{\mu\nu\ldots}(\mathbf{x}, t)] \ . \tag{5.4}$$

Since there are four degrees of freedom (d.o.f.) in this transformation - the four components of ϵ^μ - we see that among the ten d.o.f. of the perturbed

Einstein equation $\delta G_{\mu\nu} = 8\pi G\, \delta T_{\mu\nu}$, four represent gauge modes, and six represent physical degrees of freedom.

Let us take now a very simplistic example in order to illustrate the gauge ambiguity. We will do do as if the universe had a single spatial dimension, and could be described entirely in terms of its metric $g_{\mu\nu}(x,t)$ and energy density $\rho(x,t)$ in an arbitrary coordinate system. Let's take for instance:

$$\begin{aligned} ds^2 &= dt^2 + 2\epsilon\cos(x) dx dt + \left[\epsilon^2 \cos(2x) - 2\epsilon t \sin(x) - t^2\right] dx^2 \ , && (5.5) \\ \rho &= t + \epsilon \sin(x) \ , && (5.6) \end{aligned}$$

where ϵ is a small parameter. By averaging over x, we see that this universe can be decompose into a homogeneous background with

$$\begin{aligned} \bar{g}_{\mu\nu}(t) &= \begin{pmatrix} 1 & 0 \\ 0 & -a^2 \end{pmatrix} \ , \quad a(t) \equiv t \ , && (5.7) \\ \bar{\rho} &= t \ , && (5.8) \end{aligned}$$

plus perturbations (which are indeed small as long as $\epsilon \ll t$) with

$$\begin{aligned} \delta g_{\mu\nu}(x,t) &= \begin{pmatrix} 0 & \epsilon\cos(x) \\ \epsilon\cos(x) & \epsilon^2\cos(2x) - 2\epsilon t \sin(x) \end{pmatrix} \ , && (5.9) \\ \delta\rho(x,t) &= \epsilon \sin(x) \ . && (5.10) \end{aligned}$$

At first sight, this universe looks like a two-dimensional toy-model for the FLRW universe, with small sinusoidal perturbations. However, lets us change of coordinates system and use:

$$\begin{aligned} t' &= t + \epsilon \sin(x) \ , && (5.11) \\ x' &= x \ . && (5.12) \end{aligned}$$

In these new coordinates, this toy universe is simply described by

$$\begin{aligned} ds^2 &= dt'^2 - t'^2 dx'^2 \ , && (5.13) \\ \rho &= t' \ , && (5.14) \end{aligned}$$

and appears as perfectly homogeneous, with flat spatial curvature! In fact, this universe is indeed intrinsically homogeneous for a set of comoving observers. The perturbations were introduced artificially by the previous choice of coordinate system.

In general, a change coordinates allows to eliminate a few degrees of freedom in the perturbations (at most four function, since the field $\epsilon^\mu(\mathbf{x},t)$ is four-dimensional), but not all degrees of freedom simultaneously. When physical perturbations are present, they remain in any coordinate system, but taking different forms: sometimes a curvature perturbation, sometimes a density perturbation, sometimes both... The above example was constructed in a very special way, so that curvature and density perturbations could be eliminated simultaneously, but this is not possible when there are ten independent functions or more to eliminate!

5.1.2 Metric perturbations

Since $\delta g_{\mu\nu}$ is symmetric, it contains ten degrees of freedom (dof). Four of them can always be eliminated by a gauge transformation, so the number of physical degrees of freedom is only six. These degrees of freedom can be decomposed into scalars, 3-vectors and $3 \times 3-$tensors with respect to ordinary spatial rotations. The advantage of this decomposition is that it splits the linearized Einstein equations in three decoupled, independent sectors. Also, gauge transformations do not mix these different sectors with each other. In addition, these three sectors correspond to different physical effects: the scalars describe the generalization of Newtonian gravity, and the evolution of density perturbations; the vectors describe gravitomagnetism; the tensors describe gravitational waves.

Let us study this decomposition in details. The perturbed (flat) FLRW metric can be written as:

$$ds^2 = (1+2\phi)dt^2 + B_i dx^i dt - a^2(t)\left[(1-2\psi)\delta_{ij} + H_{ij}\right] dx^i dx^j \ , \quad (5.15)$$

where H_{ij} is a traceless tensor ($\sum_i H_{ii} = 0$). The number of degrees of freedom in $(\phi, B_i, \psi, H_{ij})$ is 1+3+1+5=10. The 3-vector B_i can be decomposed into "electric" and "magnetic" parts, called in this context "longitudinal" and "transverse":

$$\vec{B} = \vec{\nabla} b + \vec{\nabla} \times \vec{b} \quad (5.16)$$

where $\vec{b}$ is a transverse 3-vector, obeying to $\vec{\nabla}\vec{b} = 0$ and containing only two independent degrees of freedom. This can also be written as

$$B_i = \partial_i b + \epsilon_{ijk} \partial_j b_k \quad (5.17)$$

where ϵ_{ijk} is the maximally antisymmetric tensor. Similarly, the traceless tensor H_{ij} can be decomposed into three traceless parts:

$$\begin{aligned} H_{ij} \ &= \ \left(\partial_i \partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right)\mu && (5.18) \\ &+ \ (\partial_i a_j + \partial_j a_i) \qquad \text{with} \ \sum_i \partial_i a_i = 0 && (5.19) \\ &+ \ h_{ij} \qquad \text{with} \ \sum_i \partial_i h_{ij} = 0 \ \text{ and } \ \sum_i h_{ii} = 0 \ . && (5.20) \end{aligned}$$

At the end of the day, we obtain four scalar degrees of freedom: (ϕ, ψ, b, μ), two of which can be eliminated by gauge transformations; plus four vector degrees of freedom inside (b_i, a_i), two of which can be eliminated by gauge transformations; plus finally two tensor degrees of freedom inside h_{ij}, which are completely gauge-invariant.

In cosmology, vector perturbations are usually neglected, since they decay with the universe expansion. Gravitational waves might play a role, but we will return to this issue when studying inflation in the last chapter. In the rest of this section, we will only study scalar modes, which represent the most relevant degrees of freedom and allow to study the evolution of density perturbations, pressure perturbations, temperature perturbations, etc.

It is possible to build some gauge-invariant combinations of the perturbations, and to reduce the Einstein equation into a set of gauge-invariant equations. This is not the most economic way to proceed: one can simply choose an arbitrary gauge-fixing condition, i.e., a prescription that will limit the number of effective degrees of freedom to that of physical modes only, and make all calculations inside this gauge. When the same problem is studied in two different gauges, the solutions can look very different on large wavelengths; for instance, the total energy density perturbation $\delta\rho(t,k)$ for a given time and wavenumber can appear as growing in one gauge, and constant in another gauge, although the two solutions describe the same universe. However, physical observables - like the matter density perturbations probed by galaxy surveys or temperature/polarization anisotropies probed by CMB experiments - are always limited to small scales, at most of the order of the Hubble length. On those scales, the predictions arising from different gauge choices always coincide with each other.

Throughout our discussion, we choose to work in the *longitudinal gauge*, which is probably the most popular one for studying cosmological perturbations. In this gauge, one requires that the non-diagonal metric perturbations vanish. This eliminates two scalar degrees of freedom b and μ, and the metric including scalar perturbations reads:

$$\begin{aligned} ds^2 &= (1+2\phi)dt^2 - a^2(t)(1-2\psi)\delta_{ij}dx^i dx^j & (5.21) \\ &= a^2(\tau)[(1+2\phi)d\tau^2 - (1-2\psi)\delta_{ij}dx^i dx^j] & (5.22) \end{aligned}$$

in terms of proper time t or conformal time τ. At sufficiently small distances (smaller than the Hubble radius R_H), the scalar sector of general relativity can be approximated by Newtonian gravity. In this case, the above quantity ϕ can be identified with the usual gravitational potential. So, this gauge is also called the *Newtonian gauge*.

5.1.3 Energy-momentum tensor perturbations

The energy-momentum tensor of each species i also has four scalar degrees of freedom, which can be identified in a similar way than for metric perturbations. First,

$$\delta T^0_0 = \delta\rho \ , \tag{5.23}$$

where $\delta\rho$ is the energy density perturbation. Next, the energy flux δT^0_i has one longitudinal degree of freedom usually defined as θ:

$$\sum_i \partial_i \delta T^0_i \equiv (\bar{\rho}+\bar{p})\theta \ . \tag{5.24}$$

The $3\times 3-$tensor δT^i_j contains two scalar degrees of freedom: one along the diagonal, which is just the isotropic pressure perturbation δp; and one built from the other components, which represents the scalar component of the anisotropic pressure or anisotropic stress. This quantity is usually defined in the following way:

$$\delta T^i_j = -\delta p \ \delta^i_j + T^i_j\big|_{\text{anisotropic}} \ , \tag{5.25}$$

with

$$\sum_{i,j}\left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right) T^i_j\big|_{\text{anisotropic}} \equiv (\bar{\rho}+\bar{p})\nabla^2\sigma \ , \tag{5.26}$$

where σ represents the scalar anisotropic pressure (or anisotropic stress). Note that the scalar components of the energy-momentum tensor are not gauge-invariant (apart from σ): if one defines the above quantities in two different gauges, then for instance the two different $\delta\rho$'s would be related through a non-trivial equation.

For a perfect fluid, one can show that σ vanishes at first order in perturbations: a perfect fluid has an isotropic pressure in very good approximation (this statement can be proved by taking the expression of the energy-momentum tensor for a perfect fluid, $T^\mu_\nu = -pg^\mu_\nu + (\rho+p)U^\mu_\nu$, and expanding it at first order in perturbations).

5.1.4 Equations of motion

Einstein equations

We can now write the linearized Einstein equations $\delta G_{\mu\nu} = 8\pi G\, \delta T_{\mu\nu}$:

- for scalar perturbations only,
- in the longitudinal/Newtonian gauge,
- using conformal time: here and in the rest of this chapter, *the dot denotes a derivative with respect to conformal time* τ.

The four equations:

$$\begin{aligned}
-3\left(\frac{\dot{a}}{a}\right)^2 \phi - 3\frac{\dot{a}}{a}\dot{\psi} + \Delta\psi &= 4\pi G\; a^2\; \delta\rho\;, \\
\Delta\left(\frac{\dot{a}}{a}\phi + \dot{\psi}\right) &= 4\pi G\; a^2\; (\bar{\rho} + \bar{p})\; \theta\;, \\
\left(2\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2\right)\phi + \frac{\dot{a}}{a}(\dot{\phi} + 2\dot{\psi}) + \ddot{\psi} + \frac{1}{3}\Delta(\phi - \psi) &= 4\pi G\; a^2\; \delta p\;, \\
\Delta(\psi - \phi) &= 12\pi G\; a^2\; (\bar{\rho} + \bar{p})\; \sigma\;,
\end{aligned}$$

can easily be derived after computing the Christoffel symbols for the perturbed metric of Eq. (5.22), keeping only linear terms in ϕ and ψ (i.e. only first-order perturbations). The terms on the right-hand side are the total density, pressure, energy flux and anisotropic pressure perturbations, obtained by summing over the δT^{μ}_{ν}'s of all species. In Fourier space, these equations become:

$$\begin{aligned}
-3\left(\frac{\dot{a}}{a}\right)^2 \phi - 3\frac{\dot{a}}{a}\dot{\psi} - k^2\psi &= 4\pi G\; a^2\; \delta\rho\;, \\
-k^2\left(\frac{\dot{a}}{a}\phi + \dot{\psi}\right) &= 4\pi G\; a^2\; (\bar{\rho} + \bar{p})\; \theta\;, \\
\left(2\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2\right)\phi + \frac{\dot{a}}{a}(\dot{\phi} + 2\dot{\psi}) + \ddot{\psi} - \frac{k^2}{3}(\phi - \psi) &= 4\pi G\; a^2\; \delta p\;, \\
k^2(\phi - \psi) &= 12\pi G\; a^2\; (\bar{\rho} + \bar{p})\; \sigma\;,
\end{aligned}$$

Note that whenever σ vanishes, the last equation imposes $\phi = \psi$. In fact, we can use this approximation for the simplified calculations of this course. Indeed, perfect fluids cannot contribute to σ, as we just saw. Before recombination, the thermal plasma made of photons, baryons and electrons forms a perfect fluid, while cold dark matter is strongly non-relativistic and hence pressureless with $\delta p = \sigma = 0$. So, only neutrinos can contribute to σ. After recombination, we are deep inside the matter dominated regime, and the perturbations of the subdominant relativistic species (photons and neutrinos) are irrelevant; cold dark matter and baryons (now in the form of atoms) are both strongly non-relativistic and pressureless with $\delta p = \sigma = 0$. In conclusion, the only reason for a non-negligible σ term is the presence of neutrinos during radiation domination. This is only a small effect. In this course, for simplicity, we will neglect the anisotropic pressure perturbation of neutrinos, and work in the $\sigma = 0$ approximation. Hence the two metric perturbations are equal to each other, and the Einstein equations reduce to:

$$
\begin{aligned}
-3\left(\frac{\dot a}{a}\right)^2\phi - 3\frac{\dot a}{a}\dot\phi - k^2\phi &= 4\pi G\ a^2\ \delta\rho\ , &(5.27)\\
-k^2\left(\frac{\dot a}{a}\phi + \dot\phi\right) &= 4\pi G\ a^2\ (\bar\rho+\bar p)\ \theta\ , &(5.28)\\
\left(2\frac{\ddot a}{a} - \left(\frac{\dot a}{a}\right)^2\right)\phi + 3\frac{\dot a}{a}\dot\phi + \ddot\phi &= 4\pi G\ a^2\ \delta p\ . &(5.29)
\end{aligned}
$$

Equations of motion for each fluid

The Einstein equations provide the relation between geometry and matter, i.e. between metric and density/pressure perturbations. Since Einstein equations imply Bianchi identities $D_\mu T^\mu_\nu = 0$, equations (5.27-5.29) lead to two equations of motion: the conservation of the total density perturbation, and the Euler equation for the time-derivative of the total energy flux. However, if the different species "s" are isolated from each other (non-interacting), there is an independent set of conservation equations for each of them. The relation $D_\mu T_{(s)}{}^\mu_0 = 0$ corresponds to the energy conservation or continuity equation for each species:

$$\dot\delta_s = (1+w)(\theta_s + 3\dot\psi)\ , \tag{5.30}$$

where we defined the relative density perturbation

$$\delta_s \equiv \frac{\delta\rho_s}{\bar\rho_s}\ . \tag{5.31}$$

Here we also assumed that the species "s" has a linear equation of state $p_s/\rho_s = w$, with $\dot w = 0$. This will be the case for all species at the level of this course. This equation of state holds both at the level of the background and of perturbation. So, $\bar p_s/\bar\rho_s = w$, and the sound speed in the fluid "s", defined as usual through $c_s^2 \equiv \delta p_s/\delta\rho_s$, is equal to w. Next, the relation $D_\mu T_{(s)}{}^\mu_i = 0$ implies one scalar equation for each species (called the Euler equation like in usual fluid mechanics):

$$\dot\theta_s \equiv \frac{\dot a}{a}(3w-1)\theta_s - k^2\phi - k^2\sigma_s - \frac{w}{1+w}k^2\delta_s\ . \tag{5.32}$$

For perfect fluids or non-relativistic matter, σ_s vanishes and the fluid is described by two independent variables δ_s and θ_s (since $\delta p_s = w\delta\rho_s = w\bar\rho_s\delta_s$). Then, the continuity and Euler equations are sufficient for computing the evolution of the fluid for known ϕ (but ϕ can be derived from one of the Einstein equations). Hence the full system of differential equations is closed. For imperfect fluids or free-streaming species like decoupled neutrinos, or like photons around and after recombination, things are much more complicated (the equation of motion is then the perturbed Boltzmann equation), but in this course we will use various simplifications and avoid these very technical aspects.

Newtonian limit

We know that the Newtonian theory of gravitation should be recovered on distances smaller than the Hubble radius R_H. A given Fourier mode is associated with a comoving wavelength $2\pi/k$ and with a physical wavelength:

$$\lambda(t) = a(t)\frac{2\pi}{k}\ . \tag{5.33}$$

Since $R_H = 1/H$, the condition for a wavelength to be much smaller than the Hubble radius reads:

$$k \gg 2\pi\, a\, H \ . \tag{5.34}$$

This condition is usually written without the (2π) factor, since only orders of magnitudes are important here. When conformal time is employed, $H = da/(adt) = da/(a^2 d\tau)$. So, the condition becomes:

$$k \gg 2\pi \frac{\dot{a}}{a} \tag{5.35}$$

with $\dot{a} \equiv da/d\tau$. We see immediately that when this inequality is satisfied, the left-hand side in the first Einstein equation (5.27) is dominated by the term involving k^2:

$$-k^2 \phi = 4\pi G\ a^2\ \delta\rho \ . \tag{5.36}$$

Going back to real space ($-k^2 \longrightarrow \Delta$), this gives:

$$\frac{\Delta\phi}{a^2} = 4\pi G\ \delta\rho \ . \tag{5.37}$$

Here, Δ is the comoving Laplacian. Since it has the dimension of x^{-2}, the physical Laplacian is Δ/a^2 and the above equation is just the usual Poisson equation of Newtonian gravity: the Laplacian of the gravitational field is equal to the density times $4\pi G$. However, we see that in the context of cosmology, the gravitational potential is not sourced by the *total density*, but by the *density perturbations* with respect to the homogeneous FLRW background. Note also that in the above equation, $\delta\rho$ is the perturbed *energy* density, while the Newtonian Poisson equation involves the mass density. However, for non-relativistic matter, the mass and energy density are equal to each other in our $c = 1$ units.

5.2 General principles

5.2.1 A stochastic theory

It is important to understand that the theory of cosmological perturbations is a stochastic theory, i.e. a theory for the evolution of random quantities. Our purpose is not to predict the value of each perturbation in each point, and what is the exact position of each galaxy around us. The goal is only to understand the statistical properties of the fluctuations at each given time, the way in which typical objects can form, and how these statistical properties are related to a physical model describing the global properties of the universe.

Hence, the theory of cosmological perturbations differs, for instance, from usual fluid mechanics by the fact that the state of the universe at a given time is not described by a definite spatial distribution of various functions, but by the statistical properties of these distributions. In a homogeneous universe, the statistical properties of the spatial distributions should be invariant by translation: hence, it is convenient to go to Fourier space and to discuss the statistical properties of the Fourier modes. For instance, these properties can be formulated as the two-point correlation function, three-point correlation function and higher momenta of the Fourier modes.

Let us consider a variable $\delta(\mathbf{x}, \tau)$ (standing e.g. for the relative density perturbation $\delta\rho/\bar{\rho}$ of a given component). Since $\delta(\mathbf{x}, \tau)$ is real, it can be expanded in (comoving) Fourier modes $\delta(\mathbf{k}, \tau)$ which are complex numbers, but with a symmetry $\delta(\mathbf{k},t)^* = \delta(-\mathbf{k},\tau)$. As long as the theory is linear, different modes $\mathbf{k} \neq \mathbf{k}'$ obey to independent equations of motions.

Let us now view $\delta(\mathbf{x},\tau)$ or $\delta(\mathbf{k},\tau)$ as a stochastic field, characterized by statistical properties. A simple case is that of Gaussian isotropic fluctuations. In that case, at any time, each Fourier mode $\delta(\mathbf{k},\tau)$ is statistically independent from other modes, obeys to a Gaussian probability distribution, and has a variance depending on the modulus k but not on the direction $\mathbf{k}/k$ (this follows from the isotropy of the background). Hence, the properties of $\delta(\mathbf{k},\tau)$ at a given time τ are entirely described by the variance $\sigma(k,\tau)$. The square of this variance is called the power spectrum, $P(k,\tau) = \sigma(k,\tau)^2$:

$$\langle \delta(\mathbf{k},\tau)\delta(\mathbf{k},\tau)^* \rangle = P(k,\tau) \tag{5.38}$$

where the average holds over all possible realization of the stochastic number $\delta(\mathbf{k},\tau)$.

The fact that the perturbations fulfill this set of properties at any time (i.e., the fact that they keep being Gaussian and isotropic) is compatible with the equations of motion of the system: the equations are linear, so each mode $\mathbf{k}$ remains independent of the others; the coefficient which appear in the equations only depend on the modulus k, no $\mathbf{k}$-dependence can be induced in the probability distribution; and finally, the linearity of the equations of motion imply that the shape of the probability distribution of each mode (in our case, a Gaussian) is preserved: only the variance can increase with time. In order to see this better, let us recall that for a homogeneous second-order differential equation, the general solution can be written in terms of two initial conditions – typically, one initial amplitude and one initial time-derivative:

$$\delta(\mathbf{k},\tau) = \delta(\mathbf{k},\tau_i)\, T(k,\tau) + \dot{\delta}(\mathbf{k},\tau_i)\, S(k,\tau) \tag{5.39}$$

where τ_i is the initial time and $\{T(k,\tau),\, S(k,\tau)\}$ are two independent solutions of the equation of motion normalized to $T(k,\tau_i) = 1$, $S(k,\tau_i) = 0$. The solution of the equations of evolution describing cosmological perturbations are of this type (here, we simplified everything by considering a single variable and a single equation of motion; in reality, when the universe contains N fluids, the evolution can be described by N variables obeying to a system of N coupled, linear equations of motion). Note that T and S only depend on the wavenumber k: this is a consequence of the isotropy of the background, i.e. of the fact that the equations of motion only depend on the wavenumber, not on the wave-vector.

Next, let us assume that the initial state is such that $\dot{\delta}(\mathbf{k},\tau_i) = 0$, so that the solution for each mode can be written as

$$\delta(\mathbf{k},\tau) = \delta(\mathbf{k},\tau_i)\, T(k,\tau) \ . \tag{5.40}$$

Typically, such a simplification will arise in the realistic cases described later (the reader will see later that we can always neglect one of the two independent solutions, called a “decaying mode”). We end up with a simple linear relation between the Fourier modes at initial time and at a later time. The evolution is given by a multiplicative factor $T(k,\tau)$ called the *transfer function.*

If we now think of $\delta(\mathbf{k},\tau)$ as a stochastic number, it is clear that the linear relation (5.40) preserves the shape of the probability distribution. In particular, if the initial state is Gaussian-distributed, this remains true for $\delta(\mathbf{k},\tau)$ at any time. However the variance evolves according to the square of the transfer function. The power spectrum at some final time is given by:

$$\langle|\delta(\mathbf{k},\tau_f)|^2\rangle = \langle|\delta(\mathbf{k},\tau_i)|^2\rangle\, T^2(k,\tau_f) \tag{5.41}$$

Hence, it is sufficient to compute the properly normalized solutions of the equation of motion $T(k,\tau)$ for expressing the power spectrum at an arbitrary time τ_f in function of the primordial spectra $\langle|\delta(\mathbf{k},\tau_i)|^2\rangle$ at some initial time τ_i. In summary, if the universe contains Gaussian isotropic perturbations, we can entirely describe these perturbations if we know:

1. the primordial power spectrum for each variable: these spectra should be formulated as starting assumptions, or given by some theory for the generation of primordial perturbation in the early universe (the theory of inflation discussed in the last chapter of this course is the most favored theory of this type, and it predicts indeed Gaussian perturbations with a particular, calculable primordial spectrum).

2. the transfer function $T(k,\tau)$ for each variable. These functions can be obtained by integrating the equations of motion. They tell us which modes grow or decay, at which time, at which rate, etc. They allow us to compute the primordial spectrum of a given quantity at any time, by taking the product of the primordial spectrum with the square of the transfer function.

On could make an analogy with quantum field theory. For a free field (with quadratic potential), the classical equations of motion are linear and Fourier modes are independent from each other. The initial state is described by a wave functional (a wave-function for each Fourier mode), and the time evolution starting from a given initial state is given by the Hamiltonian, i.e. by the solutions of the classical equation of motion. Here, we are dealing with a classical stochastic theory. The formalism is similar, excepted that wave functions are replaced by true probabilities, and that there is no need for introducing non-commuting operators. However, there is an analogy in the sense that the initial state is given by a probability function for each Fourier mode, and the time evolution by the solution of the equations of motion.

5.2.2 Horizons

Causal horizon

Let us define $d_H(\tau)$ as the distance travelled by a photon between the very early universe (say, the Big Bang) and a given time τ. At that time, two points separated by a distance larger than d_H are not in *causal contact*: they cannot share any common information, and if their properties are random, they should be uncorrelated. Going to Fourier space, this means that on wavelengths $\lambda(\tau) \gg d_H(\tau)$, the power spectrum is expected to vanish (we will see later that this is not exactly the case, and we will explain why in the last chapter on inflation). Anyway, the scale $d_H(\tau)$, called the *causal horizon*, plays an important role; it is a characteristic distance in the universe which inevitably appears when solving differential equations, since wavelengths smaller or larger than $d_H(\tau)$ are in two different regimes, called respectively “causal” or “acausal”.

The causal horizon is easy to compute from the geodesics equation of radial photons, which implies (using as usual units such that $c = 1$)

$$\frac{dr}{\sqrt{1-kr^2}} = \frac{dt}{a(t)} = d\tau \ . \tag{5.42}$$

Integrating this relation, we get

$$d_H = a(t) \int_0^r \frac{dr}{\sqrt{1-kr^2}} = a(t) \int_0^t \frac{dt'}{a(t')} = a(\tau)\ \tau \qquad (5.43)$$

where in the last equality we assumed that τ is normalized in such way that $\tau \longrightarrow 0$ near the initial singularity. Let us compute the evolution of d_H during radiation and matter domination. In both cases, $a(t) \propto t^n$ with $n < 1$ (during RD, $n = 1/2$; during MD, $n = 2/3$). This gives in terms of proper time t:

$$d_H(t) = \frac{t}{1-n} \ . \qquad (5.44)$$

Note that the Hubble radius is given by $R_H(t) = a/\dot{a} = t/n$, so that

$$d_H(t) = \frac{n}{1-n} R_H(t) \ . \qquad (5.45)$$

We see that the causal horizon and the Hubble radius are almost equal to each other, apart from a factor $n/(1-n)$ of order one. This result is true as long as the expansion of the universe is given by $a(t) \propto t^n$ with $n < 1$ at any time between the initial singularity and now. In fact, one could show that $d_H \simeq R_H$ holds as long as the universe expansion is always *decelerated* since the initial singularity.

Observable universe

We now address the following question: how can we compute the radius $d_{\rm obs}(t_0)$ of the observable universe, and the size of the largest observable wavelength? For this calculation, let us first assume that the universe was transparent at any time, and that we can observe directly photons emitted near the initial singularity. Such photons would have travelled from the singularity till now by a distance (expressed in units of today) $d_{\rm obs}(t_0) = d_H(t_0)$. So, the causal horizon also represents the radius of the observable universe at a given time, assuming that it is transparent at any time.

In fact, the universe is only transparent to photons since the time of photon decoupling $t_{\rm dec}$, and the true observable radius is

$$d_{\rm obs}(t_0) = a(t_0) \int_{t_{\rm dec}}^{t_0} \frac{dt'}{a(t')} \ . \qquad (5.46)$$

In practise, the integral is dominated by late times, and rather insensitive to the lower bound of integration, provided that it is much smaller than the upper bound. Since $t_{\rm dec} \sim 380,000$ years while the age of the universe is of the order of ten billion years, it doesn't make a difference to integrate from zero or from $t_{\rm dec}$. In addition, the fact of integrating from $t_{\rm dec} \sim 380,000$ can be seen as an "experimental limitation". Indeed, suppose that we would be able to build an instrument measuring the cosmological background of neutrinos. The universe is transparent to neutrinos since the time of neutrino decoupling, which is a fraction of second after the Big Bang. If we could see gravitational waves emitted by the Big bang, we could choose an even smaller lower bound of integration. So, in principle, the radius of the observable universe is really $d_H(t_0)$.

We can get a rough approximation of $d_{\rm obs}(t_0)$ by neglecting radiation domination (which contribution to the integral is negligible) and a possible recent domination of a cosmological constant or of the spatial curvature term (we will see in the next chapter that there is indeed a recent stage

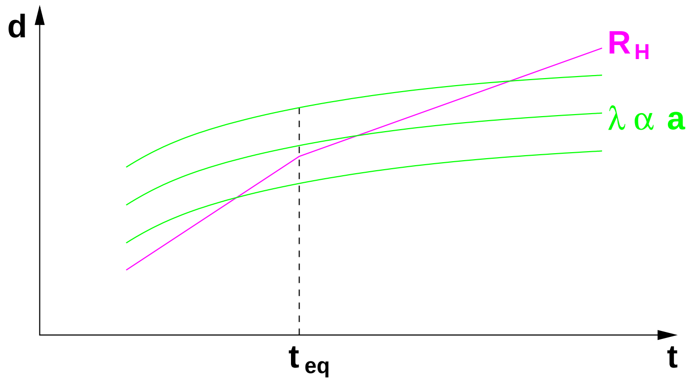

Figure 5.1: The wavelengths of some observable cosmological perturbations compared with the Hubble radius, during radiation domination and matter domination. Because all wavelengths $\lambda(t) = 2\pi a(t)/k$ grow with negative acceleration and $d_H(t) \sim R_H(t)$ grows linearly, the modes enter one after each other inside the horizon. The smallest perturbations enter during radiation domination, the others during matter domination. The largest cosmological perturbation observable today (in CMB observations) has $\lambda(t_0) \simeq R_H(t_0)$.

of cosmological constant domination, which does affect the calculation of $d_{\text{obs}}(t_0)$. So, the calculation below only indicates the correct order of magnitude). With these simplification, we can use $a(t) \propto t^{2/3}$, which gives

$$d_{\text{obs}}(t_0) = d_H(t_0) \sim 3t_0 \sim 2R_H(t_0) \ . \tag{5.47}$$

However the Hubble radius today is given in terms of the reduced Hubble parameter h by

$$R_H(t_0) = \frac{c}{H_0} = \frac{3 \times 10^5 \text{m.s}^{-1}}{100h \, \text{km.s}^{-1}\text{Mpc}^{-1}} = 3000h^{-1}\text{Mpc}. \tag{5.48}$$

So, the radius of the observable universe is of the order of $6000h^{-1}$Mpc, i.e. approximately 8000 Mpc for $h \sim 0.7$. This is only 8000 times the typical distance between two galaxies! But this corresponds to a huge volume: there should be of the order of $\sim \frac{4\pi}{3} 8000^3 \sim 10^{12}$ galaxies in the observable universe at time t_0... We recall that the result $d_{\text{obs}}(t_0)$ ~8000 Mpc was derived here for a matter-dominated universe; in case of cosmological constant or spatial curvature domination today, there should be corrections to this estimate.

Evolution of observable wavelengths versus causal horizon

So, observable wavelengths obey to $\lambda(t_0) \leq d_H(t_0) \simeq R_H(t_0)$ today, and in the past the evolution of wavelengths is always given by $\lambda(t) = 2\pi a(t)/k$. We have seen that as long as $a(t) \propto t^n$ with $n < 1$, the Hubble radius evolves like t. So, during radiation and matter domination, the Hubble radius increases *faster* than the wavelengths of all perturbations. This is illustrated in figure 5.1. So, the observable Fourier modes enter one after each other inside the causal horizon. The modes which enter earlier are those with the smallest wavelength. The wavelengths of the perturbations observed today (in the form large scale matter inhomogeneities, CMB anisotropies, etc.) were necessarily *outside* the Hubble radius in the early universe. This striking property has been puzzling cosmologists for many years. As we said before, common intuition says that outside the horizon perturbations should be uncorrelated, and the power spectrum should vanish. So, it would be natural to start from the initial condition $\delta(k,\tau) = 0$, $\dot{\delta}(k,\tau) = 0$ for any kind of perturbation with $\lambda \gg R_H$. But all equations of evolution are linear in the regime of interest. Linear equations with vanishing initial conditions imply that perturbations remain null at any time. So the universe should still be perfectly homogeneous with no galaxies, CMB anisotropies, etc. We will see in the last chapter that this contradiction can be overcame by assuming a stage called "inflation", which generates primordial fluctuations that remain coherent on very large scale, such that $P(k) \neq 0$ even for Fourier modes such that $\lambda \gg R_H$ at the beginning of radiation domination. In this case, the causal horizon still plays a role during RD and MD. Fluctuations do not vanish beyond d_H, but they are *not affected by any physical mechanism* (a given mechanism is associated with a force mediated by a boson – for instance, the photon for electromagnetism – travelling at most at the speed of light). Hence, they are "frozen" until $\lambda \sim R_H$. Then, they start to evolve, according to electromagnetic or gravitational forces for instance.

Sound horizon

The causal horizon is not the only interesting horizon. In principle, each physical mechanism can propagate at a maximum speed $v(t)$ which is not necessarily as large as the speed of light ($v = 1$). The corresponding horizon is given by integrating over $[v(t)/a(t)]dt$ instead of $[c/a(t)]dt = dt/a(t)$. For instance, we will define later the sound horizon in the photon-baryon-electron fluid, c_s, which is close to $1/\sqrt{3}$ well before photon recombination (still in units where $c = 1$, otherwise $c_s \simeq c/\sqrt{3}$). The corresponding horizon is called the sound horizon:

$$d_s = a(t) \int_0^t \frac{c_s(t')\, dt'}{a(t')} \simeq \frac{d_H(t)}{\sqrt{3}} \ . \tag{5.49}$$

We will see that the sound horizon plays a crucial role in the evolution of perturbations. Intuitively, it represents the maximum distance over which the wavefront of acoustic waves propagating at a speed c_s can travel between the early universe and some later time.

5.2.3 Initial conditions

Let us consider some early initial time between nucleosynthesis and recombination. At that time, observable wavelengths are expected to be outside the Hubble radius, $\lambda \gg R_H$, and the corresponding Fourier modes are not affected by physical mechanisms: they are frozen. However, in real space and in any single point, microscopic interactions maintain thermal equilibrium and we can apply results from thermodynamics locally.

At that time the universe contains photons, electrons and baryons in thermal equilibrium, plus decoupled neutrinos and dark matter particles. Photons are relativistic with $n_i \propto T^3$. So, the local photon abundance $n_\gamma(\mathbf{x},\tau) = \bar{n}_\gamma(\tau) + \delta n_\gamma(\mathbf{x},\tau)$ is given by the local value of the temperature $T(\mathbf{x},\tau) = \bar{T}(\tau) + \delta T(\mathbf{x},\tau)$, and at linear order:

$$\frac{\delta n_\gamma}{\bar{n}_\gamma} = 3\frac{\delta T}{\bar{T}} \tag{5.50}$$

where we omitted the arguments $(\mathbf{x}, \tau)$ for concision. Note that this relation is also true in Fourier space, i.e. using the arguments $(\mathbf{k}, \tau)$. Neutrinos are decoupled, but we have seen that their distribution is still of the Fermi-Dirac type with $\mu_\nu \simeq 0$ in the simplest models, and $T_\nu = (4/11)^{1/3} T$. This results is true in every point, so

$$\frac{\delta n_\nu}{\bar{n}_\nu} = 3\frac{\delta T_\nu}{\bar{T}_\nu} = 3\frac{\delta T}{\bar{T}} = \frac{\delta n_\gamma}{\bar{n}_\gamma} \ . \tag{5.51}$$

Hence, γ and ν share a common inhomogeneity $\delta n_i/\bar{n}_i = 3\delta T/\bar{T}$. A priori, this is not obviously true for baryons and for cold dark matter particles (we will label this last component with an index *cdm*), which could either obey to initial conditions such that

$$\frac{\delta n_b}{\bar{n}_b} = \frac{\delta n_{cdm}}{\bar{n}_{cdm}} = \frac{\delta n_i}{\bar{n}_i} = 3\frac{\delta T}{\bar{T}} \tag{5.52}$$

called *adiabatic initial conditions*, or to more general initial conditions

$$\frac{\delta n_{cdm}}{\bar{n}_{cdm}} \neq 3\frac{\delta T}{\bar{T}} \quad \text{and/or} \quad \frac{\delta n_b}{\bar{n}_b} \neq 3\frac{\delta T}{\bar{T}} \tag{5.53}$$

in which case *entropy perturbations* would be present in the early universe. It is beyond the level of this course to prove that the simplest models for baryon and dark matter production in the early universe imply exact adiabatic initial conditions for wavelengths larger than the Hubble radius: we will admit it, and hence assume that $\delta n_i/\bar{n}_i = 3\delta T/\bar{T}$ is true for *all relevant species* with $\lambda \gg R_H$. We will see later that observations are perfectly compatible with such an assumption.

Note that here we discussed the case of γ, ν, b, *cdm* but not that of electrons. In fact adiabatic initial conditions are also shared by electrons, but when studying cosmological perturbations electrons can always be neglected, since they trace baryons (as requested by electric neutrality) while their mass is negligible with respect to that of baryons. They play no role apart from maintaining photons and baryons in thermal equilibrium before recombination.

We can infer from $\delta n_i/\bar{n}_i = 3\delta T/\bar{T}$ the relation between the various density perturbations $\delta_i = \delta\rho_i/\bar{\rho}_i$. Non-relativistic species (b, *cdm*) are such that $\rho_i = m n_i$ in any point, hence $\delta_i = \delta n_i/\bar{n}_i$ for baryons and cold dark matter. Relativistic species (γ, ν) satisfy $n_i \propto T_i^3$ and $\rho_i \propto T_i^4$ locally (even for decoupled neutrinos), so $\delta_i = (4/3)\delta n_i/\bar{n}_i$ for photons and neutrinos. We conclude that in the early universe, as long as $\lambda \gg R_H$, and assuming adiabatic initial conditions,

$$\delta_b = \delta_{cdm} = \frac{3}{4}\delta_\gamma = \frac{3}{4}\delta_\nu = 3\frac{\delta T}{\bar{T}} \ . \tag{5.54}$$

This equality holds in Fourier space for modes with $k \ll aH$ (super-Hubble wavelengths). On those scales, the modes are frozen: the density perturbations are constant and the velocity divergences θ_i are negligible.

We conclude from these relations that initial conditions do not consist in one primordial power spectrum for *each* species. All perturbations are related to each others, and are characterized by a single primordial spectrum from which other power spectra can be immediately computed: $\langle|\delta_b|^2\rangle = \langle|\delta_{cdm}|^2\rangle = \frac{9}{16}\langle|\delta_\gamma|^2\rangle = \frac{9}{16}\langle|\delta_\nu|^2\rangle$.

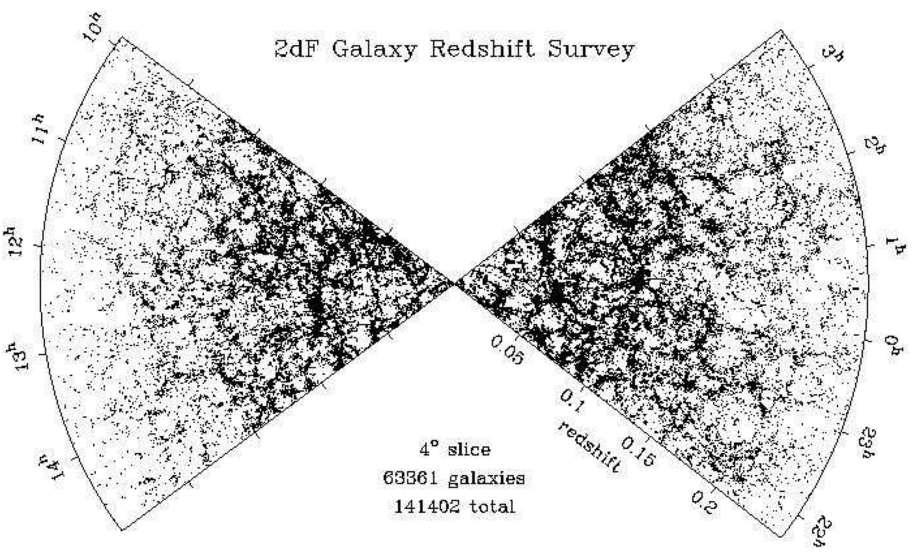

Figure 5.2: The distribution of galaxies in two thin slices of the neighboring universe, obtained by the 2dF Galaxy Redshift Survey (see J. A. Peacock et al., Nature 410 (2001) 169-173). The radial coordinate is the redshift, or the distance between the object and us.

5.2.4 Observable quantities

Power spectrum of large scale structure

If we have a large three–dimensional map of the galaxy distribution in the universe (see e.g. Fig. 5.2), we can smooth it on very large scales and reconstruct a smooth function representing the fluctuations in the luminous galactic matter (lgm) density on large scales, $\delta_{\rm lgm}(\mathbf{x}, \tau_0)$. The corresponding Fourier distribution $\delta_{\rm lgm}(\mathbf{k}, \tau_0)$ is then readily obtained. Finally, the average of $|\delta_{\rm lgm}(\mathbf{k}, \tau_0)|^2$ for all Fourier modes $\mathbf{k}$ with a fixed wavenumber k can be compared with the theoretical power spectrum $P_{\rm lgm}(k)$ computed with cosmological perturbation theory for a given cosmological model:

$$\langle|\delta_{\rm lgm}(\mathbf{k}, t)|^2\rangle_{\mathbf{k}/k} \longrightarrow P_{\rm lgm}(k) \ . \tag{5.55}$$

If $\lambda = 2\pi a_0/k$ is small with respect to the physical size L of the survey, this average includes many independent terms and should be very close to the theoretical power spectrum. In the other limit, only few independent modes can contribute to the average: in this case the average and the theoretical prediction can differ significantly by a *sample variance* called also *cosmic variance.*

Let us define the non-relativistic matter density $\rho_m = \rho_b + \rho_{cdm}$ and its perturbation $\delta_m = \delta\rho_m/\bar{\rho}_m$. The theory of cosmological perturbations can predict accurately the total matter power spectrum $P(k, \tau_0) = \langle|\delta_m(\mathbf{k}, \tau_0)|^2\rangle$, not the luminous galactic matter power spectrum $P_{\rm lgm}(k, \tau_0) = \langle|\delta_{\rm lgm}(\mathbf{k}, \tau_0)|^2\rangle$. However, various calculations (based on the modeling of galaxy formation) prove that on large scales, the distribution of luminous galactic matter "traces" the distribution of mass. This means that the inhomogeneities in luminous galactic matter is proportional to the total density inhomogeneities (i.e., the total inhomogeneities of non-relativistic matter, baryons and cold dark matter). The unknown coefficient of proportionality is called the light-to-mass bias b:

$$\delta_{\rm lgm}(\mathbf{k},\tau) \simeq b\ \delta_m(\mathbf{k},\tau)\ . \tag{5.56}$$

It follows that $P_{\rm lgm}(k,\tau_0) = b^2 P(k,\tau_0)$. Hence, at least the *shape* of the luminous matter power spectrum can be compared with the predictions of the theory of cosmological perturbations. However, galaxy maps are not the only way to probe the matter power spectrum. For instance, a statistical study of the weak lensing of galaxy images probes directly the total matter power spectrum (without the light-to-mass bias uncertainty) much further in the past than galaxy maps; it allows a measurement of $P(k,\tau)$ on a wide range of scales and for various times or redshifts, ranging typically up to $z \sim 2$ for future experiments.

Power spectrum of CMB temperature anisotropies

Before recombination and photon decoupling, the thermal plasma (photons, electrons and baryons) has small temperature/density inhomogeneities. In the approximation of instantaneous decoupling, photons decouple at some time $\tau_{\rm dec}$ and then propagate without interactions. Hence, photons reaching us from a direction $\mathbf{n}$ (with $|\mathbf{n}| = 1$) have decoupled from the plasma at a time $\tau = \tau_{\rm dec}$ and at a spatial coordinate $\mathbf{x} = r_{\rm dec}\mathbf{n}$, with $r_{\rm dec} = f_k(\chi_{\rm dec}) = f_k(\tau_0 - \tau_{\rm dec})$ (see section 2.2.4). In this chapter, we mainly focus on the case of a flat/Euclidean universe, for which $r_{\rm dec} = \chi_{\rm dec} = [\tau_0 - \tau_{\rm dec}]$. This point is located on the Last Scattering Surface (LSS), defined as the ensemble of points where a photon last scattered before reaching us today: obviously this surface is by definition a sphere centered on us, with comoving radius equal to $[\tau_0 - \tau_{\rm dec}]$ in the Euclidean case.

Since photons have a a Planckian distribution even after decoupling, the information that we receive from the LSS in a given direction $\mathbf{n}$ consists in one value of the temperature, $T_{\rm obs}(\mathbf{n})$. Actually, there is some extra information contained in the photon polarization in a given direction, but for simplicity we will not discuss polarization in this course.

We can define $\bar{T}$ as the temperature averaged over all directions (measured to be 2.726 K). The temperature anisotropy map is then defined as

$$\left.\frac{\delta T}{\bar{T}}\right|_{\rm obs}(\mathbf{n}) = \frac{T_{\rm obs}(\mathbf{n}) - \bar{T}}{\bar{T}}\ . \tag{5.57}$$

In cosmological perturbation theory, if all perturbations in the universe are stochastic and Gaussian, then the quantity $\delta T/\bar{T}$ should also be stochastic and Gaussian. It is then fully characterized by its two-point correlation function. This function can be defined in real space, on the sphere, as

$$\left\langle \frac{\delta T}{\bar{T}}(\mathbf{n})\frac{\delta T}{\bar{T}}(\mathbf{n}') \right\rangle = F(\theta) \qquad \text{with} \quad \cos\theta = \mathbf{n}.\mathbf{n}'. \tag{5.58}$$

This function, called the "angular correlation function", depends only on the angle between the two directions $(\mathbf{n},\mathbf{n}')$ as a consequence of homogeneity and isotropy of the background. If $\delta T/\bar{T}$ was a three-dimensional quantity, we would also be interested in its two-point correlation function in Fourier space, namely the Fourier power spectrum. However $\delta T/\bar{T}$ is a two dimensional function depending only on a unit vector $\mathbf{n}$ (i.e. on two angles). Hence, it should be expanded in spherical harmonics rather than Fourier modes:

$$\left.\frac{\delta T}{\bar{T}}\right|_{\rm obs}(\mathbf{n}) = \sum a_{l,m}\ Y^{l,m}(\mathbf{n}) \tag{5.59}$$

with $a_{l,m} = a^*_{l,-m}$ since $\delta T/\bar{T}$ is real. The equivalent of the Fourier power spectrum in Fourier space, $\langle|\delta(\mathbf{k})|^2\rangle$, is the harmonic power spectrum in harmonic space, defined as:

$$\langle|a_{l,m}|^2\rangle = C_l \ . \tag{5.60}$$

We recall that we are dealing with a stochastic perturbation theory, in which $\delta T/\bar{T}$ and hence also $a_{l,m}$ are stochastic numbers. The average holds over all possible realization of these stochastic, Gaussian numbers. In Fourier space, isotropy implies that the Fourier power spectrum only depends on k, not $\mathbf{k}$. In harmonic space, for the same reason, the harmonic power spectrum $\langle|a_{l,m}|^2\rangle$ only depends on l (the angular scale of the multipole), not on m (the orientation of the multipole). Note that the angular correlation function $F(\theta)$ and the harmonic power spectrum C_l contain the same amount of information, since they are related through

$$F(\theta) = \sum_l \frac{2l+1}{4\pi} C_l P_l(\cos\theta) \tag{5.61}$$

where P_l is the Legendre polynomial of order l. In principle they can be used indifferently, but usually C_l is preferred, because it gives more weight to small angular scales where the function is non-trivial and contains interesting features. In summary, the predictions of linear cosmological perturbation theory for CMB anisotropies consists in one function $F(\theta)$ or one set of multipoles C_l, which can in principle be computed for a given cosmological model.

Observers can compare the observed map with the theoretical power spectrum in the following way. The observed map can be expanded in multipoles $a_{l,m}$. If the theoretical model is correct, the average of all $a_{l,m}$'s with l fixed should be close to the theoretical prediction for C_l:

$$\langle|a_{lm}|^2\rangle_m \longrightarrow C_l \ . \tag{5.62}$$

Like for the matter power spectrum, we can raise the following argument: if l is large, there are many independent terms in the average (since $-l \leq m \leq l$) and the left-hand side should be very close to the theoretical power spectrum; in the opposite limit, there are very few independent terms in the average and the outcome can differ significantly from the theoretical prediction, by an amount called *sample variance* or *cosmic variance*.

Finally, we note that a multipole l corresponds to anisotropies with an angular scale $\theta = \pi/l$ (for the dipole $l = 1$, the angular scale is π as expected). In the limit of small angles / large l's, an angle subtends a length on the LSS equal to $\lambda = d_A(z_{\rm dec}) \times \theta$, by definition of the angular diameter distance d_A at the redshift of decoupling. On the last scattering surface, this length λ is the wavelength of Fourier modes such that $\lambda = 2\pi a_{\rm dec}/k$. We conclude that the correlation function for a given θ or the power spectrum for a given multipole l contains information about the Fourier modes of perturbations on the LSS at time $\tau_{\rm dec}$ and for a wavenumber

$$k \ \sim \ \frac{2\pi a_{\rm dec}}{d_A(z_{\rm dec})\ \theta} \ \sim \ \frac{2a_{\rm dec}\ l}{d_A(z_{\rm dec})} \ . \tag{5.63}$$

5.3 Evolution and results

5.3.1 Global evolution during radiation and matter domination

Understanding the full evolution of cosmological perturbations is rather involved. Here, we will only compute some approximate solutions in a few limits and match them to each other in order to understand roughly what is going on.

For simplicity, in section 5.3, we will *neglect the presence of neutrinos.* In this way, we can neglect the anisotropic pressure σ and work in the approximation where $\phi = \psi$[1]. The Einstein equations then reduce to (5.27), (5.28), (5.29). If during some stage we can write a relation between the total pressure and density perturbations of the type $\delta p = c_s^2\, \delta\rho$, it it useful to combine Eqs. (5.27), (5.29) into:

$$\left(2\frac{\ddot{a}}{a} - (1-3c_s^2)\left(\frac{\dot{a}}{a}\right)^2\right)\phi + k^2 c_s^2 \phi + 3(1+c_s^2)\frac{\dot{a}}{a}\dot{\phi} + \ddot{\phi} = 0 \; . \tag{5.64}$$

Radiation domination

Between nucleosynthesis and matter-radiation equality, the total energy density is dominated by the contribution of ultra-relativistic photons (since we decided to neglect neutrinos here): $\bar{\rho} \simeq \bar{\rho}_\gamma \propto a^{-4}$. In this case, the Friedmann equation implies that $\dot{a}$ is constant, hence $a \propto \tau$. The condition for a mode to be above the Hubble scale reads:

$$\lambda \gg R_H \;\Leftrightarrow\; k \ll 2\pi\frac{\dot{a}}{a} \;\Leftrightarrow\; k\tau \ll 2\pi \; . \tag{5.65}$$

During this stage, the total pressure and density perturbations $(\delta p, \delta\rho)$ are those from relativistic photons, so $\delta p = c_s^2\, \delta\rho$ with $c_s^2 = w = 1/3$. In this limit, we can write Eq. (5.64) as:

$$\frac{k^2}{3}\phi + \frac{4}{\tau}\dot{\phi} + \ddot{\phi} = 0 \; . \tag{5.66}$$

Performing the change of variable $\phi = \tau^{3/2} u$, we get

$$\ddot{u} + \frac{1}{\tau}\dot{u} + \left(\frac{k^2}{3} - \frac{9}{4\tau^2}\right) u = 0 \; , \tag{5.67}$$

which solutions are Bessel functions

$$u = J_{\pm 3/2}(z) \qquad \text{with } z \equiv \frac{k\tau}{\sqrt{3}} = k c_s \tau \; . \tag{5.68}$$

These Bessel functions have a simple analytic expression leading finally to:

$$\phi = C_1\,(k\tau)^{-2}\left(\frac{\sin z}{z} - \cos z\right) + C_2\,(k\tau)^{-2}\left(-\frac{\cos z}{z} - \sin z\right) \; , \tag{5.69}$$

where C_1 and C_2 are two constants of integration for each mode $\mathbf{k}$. Since we are dealing with a stochastic theory, C_1 and C_2 can be seen as two

[1] After photon decoupling, photons also acquire some anisotropic pressure σ, since they are not playing the role of a perfect fluid anymore. However, after photon decoupling (i.e. during matter domination and later), the density of photons is very small compared to that of baryons and cold dark matter, and the tiny σ induced by photons plays a negligible role.

random numbers taking a different value for each $\mathbf{k}$. If the perturbations are Gaussian and isotropic, all the information concerning the statistical properties of $\{C_1(\mathbf{k}),\, C_2(\mathbf{k})\}$ is contained in the power spectra $\langle |C_1(\mathbf{k})|^2\rangle$ and $\langle |C_2(\mathbf{k})|^2\rangle$, which depend only on the modulus k. The evolution of the photon density perturbations is then given by Eq. (5.27) with $\delta\rho = \bar{\rho}_\gamma \delta_\gamma$ and, using the Friedmann equation, $4\pi G a^2 \bar{\rho}_\gamma = \frac{3}{2}\left(\frac{\dot{a}}{a}\right)^2$. Eq. (5.27) then reads:

$$-\frac{3}{\tau^2}\phi - \frac{3}{\tau}\dot{\phi} - k^2\phi = \frac{3}{2\tau^2}\delta_\gamma \; . \tag{5.70}$$

Inserting the solution (5.69) for ϕ, we obtain:

$$\begin{aligned} \delta_\gamma &= -2C_1(k\tau)^{-2}\left[2\left(z^2-1\right)\frac{\sin z}{z} - \left(z^2-2\right)\cos z\right] \\ &\quad +2C_2(k\tau)^{-2}\left[2\left(z^2-1\right)\frac{\cos z}{z} - \left(z^2-2\right)\sin z\right] \; . \end{aligned} \tag{5.71}$$

Let us review the main implications of these results.

Outside the Hubble radius. First, we write the asymptotic solution outside the Hubble radius, when $k\tau \ll 1$. In this limit, Eq. (5.69) gives:

$$\phi \longrightarrow \frac{1}{9}C_1 - \frac{\sqrt{3}}{(k\tau)^3}C_2 \; . \tag{5.72}$$

Hence, ϕ is the sum of two modes, the first one being asymptotically constant, and the second one decaying like τ^{-3}. The constants C_1 and C_2 should be fixed by some prescription for the initial conditions at some very early time τ_i, when $k\tau_i$ is extremely small with respect to one. This prescription will arise from the theory of inflation, that we will study later. However, in this chapter, it is reasonable to admit that initial conditions are such that the second mode is at most of the same order as the first one at initial time: so, $C_2 \leq C_1(k\tau_i)^3$. At some later time, the decaying mode gets suppressed by a factor $(\tau_i/\tau)^3$ and becomes vanishingly small with respect to the first one. So, it is safe to neglect the decaying mode and keep only the non-decaying one proportional to C_1:

$$\phi \longrightarrow \frac{C_1}{9} \; . \tag{5.73}$$

The corresponding initial solution for photon perturbations is:

$$\delta_\gamma \longrightarrow -\frac{2C_1}{9} \; . \tag{5.74}$$

Hence, both ϕ and $\delta_\gamma = 4\,\delta T/\bar{T}$ are constant outside the Hubble radius. Finally, for adiabatic initial conditions, we know that: $\delta_b = \delta_{cdm} = \frac{3}{4}\delta_\gamma$. We conclude that the full initial conditions are given by *constant perturbations outside the Hubble radius during radiation domination*, related through:

$$\delta_\gamma = 4\,\frac{\delta T}{\bar{T}} = \frac{4}{3}\,\delta_b = \frac{4}{3}\,\delta_d = -2\,\phi = -\frac{2C_1}{9} \; , \tag{5.75}$$

where C_1 is a function of $\mathbf{k}$ (yet arbitrary).

Inside the Hubble radius. Still keeping only the non-decaying mode proportional to C_1, we can write the asymptotic solution for ϕ and δ_γ inside the Hubble radius and during radiation domination:

$$\phi \longrightarrow -C_1(k\tau)^{-2}\cos z \; , \tag{5.76}$$

$$\delta_\gamma \longrightarrow \frac{2}{3}C_1\cos z \; , \tag{5.77}$$

still with $z \equiv (k\tau/\sqrt{3}) = (kc_s\tau)$. These solutions show that once the modes enter inside the Hubble radius, or more precisely, inside the sound horizon (when $kc_s\tau$ is of the order of one), they start to oscillate as a result of the competition between pressure and gravity. Each mode behaves roughly like a harmonic oscillator placed initially away from its equilibrium point by the initial conditions of Eq. (5.75). The largest wavelengths enter later inside the sound horizon: hence, they remain frozen for a longer time and start to oscillate later. During this era, the photons and the baryons remain tightly coupled. Hence it is possible to show that $\delta_\gamma = \frac{4}{3}\delta_b$ remains true even inside the Hubble radius before decoupling. Finally, the solution for δ_{cdm}, the perturbations of cold dark matter, is difficult to compute during radiation domination. In this era, cold dark matter particles behave like test particles, in the sense that they feel metric perturbations, but do not influence them (since the universe is dominated by relativistic matter). Cold dark matter does not experience pressure forces, so nothing prevents it from accumulating inside potential wells (this is called *gravitational clustering*). Hence, δ_{cdm} grows with time. A detailed investigation of Einstein equations and continuity/Euler equations for cold dark matter would show that inside the Hubble radius, δ_{cdm} grows proportionally to $C_1 \ln(k\tau)$.

Matter domination after photon decoupling

We leave the discussion of the intermediate stage (around equality and decoupling) for the next subsection, and switch directly to the stage of matter domination after decoupling, in order to find again convenient approximations and simple solutions. Now, the total energy density is dominated by the contribution of baryons and cold dark matter, which are both non-relativistic: in good approximation, $\bar{\rho} \simeq (\bar{\rho}_b + \bar{\rho}_{cdm}) \propto a^{-3}$. The Friedmann equation shows that $a^{1/2}\dot{a}$ is constant, hence a is proportional to τ^2. The condition for a mode to be outside the Hubble radius now reads:

$$\lambda \gg R_H \;\Leftrightarrow\; k \ll 2\pi\frac{\dot{a}}{a} \;\Leftrightarrow\; k\tau \ll 4\pi \; . \tag{5.78}$$

During this era, we can write the total density perturbation as $\delta\rho = \bar{\rho}_m\delta_m$, where m stands for total non-relativistic matter (baryons plus cold dark matter), for which δp is vanishingly small. Hence we can write Eq. (5.64) with $c_s^2 = 0$ and $a \propto \tau^2$:

$$\frac{6}{\tau}\dot{\phi} + \ddot{\phi} = 0 \; . \tag{5.79}$$

The general solution of this equation reads:

$$\phi = D_1 + (k\tau)^{-5} D_2 \; , \tag{5.80}$$

where D_1 and D_2 are two constants of integration for each mode, hence two unknown functions of $\mathbf{k}$. The corresponding solution for δ_m can easily be inferred from Eq. (5.27) with $\delta\rho = \bar{\rho}_m\delta_m$, using $4\pi G a^2 \bar{\rho}_m = \frac{3}{2}\left(\frac{\dot{a}}{a}\right)^2$ and $a \propto \tau^2$:

$$-\left(\frac{12}{\tau^2} + k^2\right)\phi - \frac{6}{\tau}\dot{\phi} = \frac{6}{\tau^2}\delta_m \; . \tag{5.81}$$

Inserting the solution (5.80) for ϕ, we obtain:

$$\delta_m = -\left(2 + \frac{k^2\tau^2}{6}\right) D_1 + \left(3 - \frac{k^2\tau^2}{6}\right)(k\tau)^{-5} D_2 \; . \tag{5.82}$$

We will come back later to the way to compute D_1 and D_2, but it is clear already that even if the two modes have the same order of magnitude initially, the second one will decay like τ^{-5}, and the full solution will be dominated by the other term:

$$\begin{aligned}\phi &\longrightarrow D_1 \ , \\ \delta_m &\longrightarrow -\left(2 + \frac{k^2\tau^2}{6}\right) D_1 \ .\end{aligned} \tag{5.83}$$

These solutions show that once the modes enter inside the Hubble radius, matter perturbations start to grow, as a simple result of gravitational clustering: $|\delta_m| \propto \tau^2 \propto a$ (there is a minus sign in Eq. (5.83) because in real space, overdensities with $\delta_m > 0$ correspond to potential wells with $\phi < 0$). Hence, during matter domination, the structures observed in the universe (galaxies, clusters, etc.) can form. Structures become non-linear on a given scale when $\delta_m \sim k^2\tau^2 D_1 \sim 1$. So, the smallest wavelengths (with the largest k) become non-linear first.

Note that gravitational clustering had started already during radiation domination for dark matter. However, in that regime, dark matter behaved as a test fluid: it did not influence the gravitational potential. During matter domination, baryons and dark matter cluster, but they are not anymore test fluids: they do influence the gravitational potential. Hence, it is not a surprise to find that during radiation domination, ϕ decays and δ_{cdm} grows only like the logarithm of the scale factor, while during MD, ϕ does not decay and δ_m grows as fast as the scale factor.

Poisson equation

The Poisson equation is the limit of Einstein equations on small distances, when any kind of relativistic matter can be neglected. Hence, we expect the Poisson equation to be applicable during matter domination on scales with $k\tau \gg 1$. Indeed, on such scales, we have seen that the Einstein equations imply

$$4\pi G \delta\rho = -\frac{k^2}{a^2}\phi \tag{5.84}$$

(see equation (5.36)). Using $\delta\rho = \bar{\rho}_m \delta_m$, the Friedmann equation $4\pi G a^2 \bar{\rho}_m = \frac{3}{2}\left(\frac{\dot{a}}{a}\right)^2$ and $a \propto \tau^2$, this relation becomes:

$$\delta_m = -\frac{k^2\tau^2}{6}\phi \ , \tag{5.85}$$

in agreement with the solutions in Eq. (5.83) for $k\tau \gg 1$.

In this section, we found that during matter domination, gravitational clustering causes the growth of matter perturbations inside the Hubble radius, while ϕ remains constant. It sounds like a paradox that ϕ remains constant during structure formation (at least within the linear theory): intuitively, gravitational clustering should imply that ϕ grows (i.e., in real space, gravitational potential wells should get deeper and deeper). This intuition is wrong and based on Newtonian mechanics in a static universe. In the expanding universe, the Poisson equation says that $\delta\rho$ is proportional to the physical Laplacian $\Delta\phi/a^2$. When ϕ is constant, the comoving gradient $\Delta\phi$ is constant, and the physical gradient $\Delta\phi/a^2$ decays like a^{-2}; hence, it does not decay as fast as energy is diluted, since $\bar{\rho}_m \propto a^{-3}$; so, the density contrast $\delta\rho_m/\bar{\rho}_m$ grows like a. Intuitively, the fact that ϕ is constant during matter domination is the result of an exact cancellation between the competing effects of gravitational clustering (which tends to concentrate matter) and of the universe expansion (which tends to dilute matter).

Summary and transition between the two stages

During radiation domination, we have seen that all perturbations δ_γ, δ_b, δ_{cdm} and ϕ are constant outside the Hubble radius, and related to each other by factors of order one (Eq. (5.75)). Inside the Hubble radius, the competition between pressure and gravity in the photon-electron-baryon thermal plasma is responsible for acoustic oscillations (Eq. (5.71)); as a result, the gravitational potential experiences damped oscillations (Eq. (5.69)). Cold dark matter does not feel pressure forces and experiences gravitational clustering, however growing as slowly as $C_1 \ln(k\tau)$, i.e. as the logarithm of the scale factor.

During the intermediate stage between matter-radiation equality and photon decoupling, analytic solutions are difficult to find because no simple approximation holds. Of course, the transition from full RD ($a \propto \tau$) to full MD ($a \propto \tau^2$) is smooth. Just before photon decoupling, photon and baryons still form a tightly coupled fluid with $\delta_\gamma = \frac{4}{3}\delta_b$, but this fluid is not anymore ultra-relativistic at the beginning of the matter dominated era. Actually, the sound speed in this fluid is easily found to be:

$$c_s^2 \equiv \frac{\delta p_\gamma}{\delta\rho_\gamma + \delta\rho_b} = \frac{\frac{1}{3}\bar{\rho}_\gamma \delta_\gamma}{\bar{\rho}_\gamma \rho_\gamma + \frac{3}{4}\bar{\rho}_b \delta_\gamma} = \frac{1}{3}\left(1 + \frac{3\bar{\rho}_b}{4\bar{\rho}_\gamma}\right)^{-1} . \tag{5.86}$$

As $\bar{\rho}_b/\bar{\rho}_\gamma$ grows from zero to infinity, c_s^2 decays from 1/3 to zero. During the stage between equality and decoupling, the perturbations remain almost constant outside the Hubble radius (although a detailed study would show a very small time-evolution). Inside the Hubble radius, the photon-baryon fluid experiences damped acoustic oscillations. The evolution of the gravitational potential and of cold dark matter perturbations depend very much on the ratio $\bar{\rho}_b/\bar{\rho_{cdm}}$. If this ratio is negligible, baryons behave like a test fluid, while dark matter drives gravity. In this case, the solutions for δ_{cdm} and ϕ become similar to those found during full matter domination (see Eq. (5.83)): δ_{cdm} experiences a smooth transition between the $\ln(k\tau)$ behavior and the $k^2\tau^2$ behavior, while ϕ freezes-out to a constant value. The relation between ϕ and δ_{cdm} is given by the Poisson equation. If instead $\bar{\rho}_b/\bar{\rho_{cdm}} \gg 1$, the dark matter particles behave like a test fluid while baryons drive gravity. In this case, during the intermediate stage, the gravitational potential follows the damped oscillations of the photon-baryon fluid while δ_{cdm} grows very slowly.

At recombination time, photons decouple gradually from baryons. This is the most difficult stage to study precisely: it is necessary to integrate over time the full perturbed Boltzmann equation, which is far beyond the level of this course. However, we don't need entering into such details for understanding the main results of the next sections. After decoupling, photon travel freely (the universe is transparent): we don't need to follow them anymore, since in first approximation the observable CMB patterns are related to the photon perturbations *at decoupling.* After decoupling, baryons are non-interacting and non-relativistic particles, exactly like dark matter. So, inside the Hubble radius, they fall inside the same gravitational potential wells as dark matter. Soon after decoupling, one has $\delta_b = \delta_{cdm}$: a gravitational equilibrium settles between baryons, dark matter and the gravitational potential, summarized by the Poisson equation

$$\delta_b = \delta_{cdm} = -\frac{k^2\tau^2}{6}\phi . \tag{5.87}$$

But how is this potential ϕ related to the one *before photon decoupling*? Intuitively, this depends again on the ratio $\bar{\rho}_b/\bar{\rho_{cdm}}$. If this ratio is negligible, baryons behave like a test fluid while dark matter drives gravity. In this case, $-(6/k^2\tau^2)\,\delta_{cdm} = \phi$ remains constant and unaffected by photon decoupling while δ_b changes rapidly to reach δ_{cdm}. If the ratio $\bar{\rho}_b/\bar{\rho_{cdm}}$ is large, dark matter behaves like a test-fluid; $-(6/k^2\tau^2)\,\delta_b = \phi$ freezes-out to the value reached just before decoupling, while δ_{cdm} changes rapidly to reach δ_b. If the ratio $\bar{\rho}_b/\bar{\rho_{cdm}}$ is of order one, the system finds an intermediate equilibrium between these two limits.

During matter domination and after decoupling, all perturbations are still constant outside the Hubble radius. Inside the Hubble radius, the gravitational potential ϕ is also constant, while $\delta_b = \delta_{cdm}$ grows proportionally to $(k\tau)^2$, i.e. linearly with the scale factor. In this stage, we are not interested in following the decoupled photons, since we can estimate observable CMB anisotropies essentially from photon perturbations at decoupling (see the next sections).

Exact numerical solutions. Various numerical codes have been developed for integrating the full set of equations for all perturbations. They provide solutions which are precise up to a fraction of percent. For instance, one can have a look at the public codes downloadable from the websites:

`http://camb.info/`

or:

`http://cfa-www.harvard.edu/~mzaldarr/CMBFAST/cmbfast.html`.

Even better, one can play with an web interface at:

`http://lambda.gsfc.nasa.gov/cgi-bin/cmbfast_form.pl`

which computes the CMB and matter power spectra for any choice of cosmological parameters. In Figure 5.3, we used such codes for plotting the actual evolution of ϕ, δ_γ and δ_{cdm} in a cosmological model with $\bar{\rho}_b/\bar{\rho}_{cdm} = 0.2$. On the top panel, it is clear that the metric perturbation experiences damped oscillations during radiation domination and inside the Hubble radius (for the largest k shown here, there is only one oscillation; going to larger k we would see more of them). After equality, we see that ϕ freezes out, more or less to the value reached at equality (since we are in the case of a small ratio $\bar{\rho}_b/\bar{\rho}_{cdm}$). In the middle panel, we see the acoustic oscillations in δ_γ, which start for each k when the mode enters inside the Hubble radius, and get rapidly damped after equality. Finally, in the lower plot, we see the growth of δ_{cdm} inside the Hubble radius - first logarithmic, then linear with a after equality.

5.3.2 Impact of a possible cosmological constant domination stage

Finally, let us assume a possible stage of cosmological constant domination after matter domination. In this case, the Friedmann equation implies

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} a^2 (\bar{\rho}_m + \bar{\rho}_\Lambda) \ , \tag{5.88}$$

where the cosmological constant energy density $\bar{\rho}_\Lambda$ is time-independent by construction, and the non-relativistic matter density $\bar{\rho}_m = \bar{\rho}_b + \bar{\rho}_{cdm}$ scales

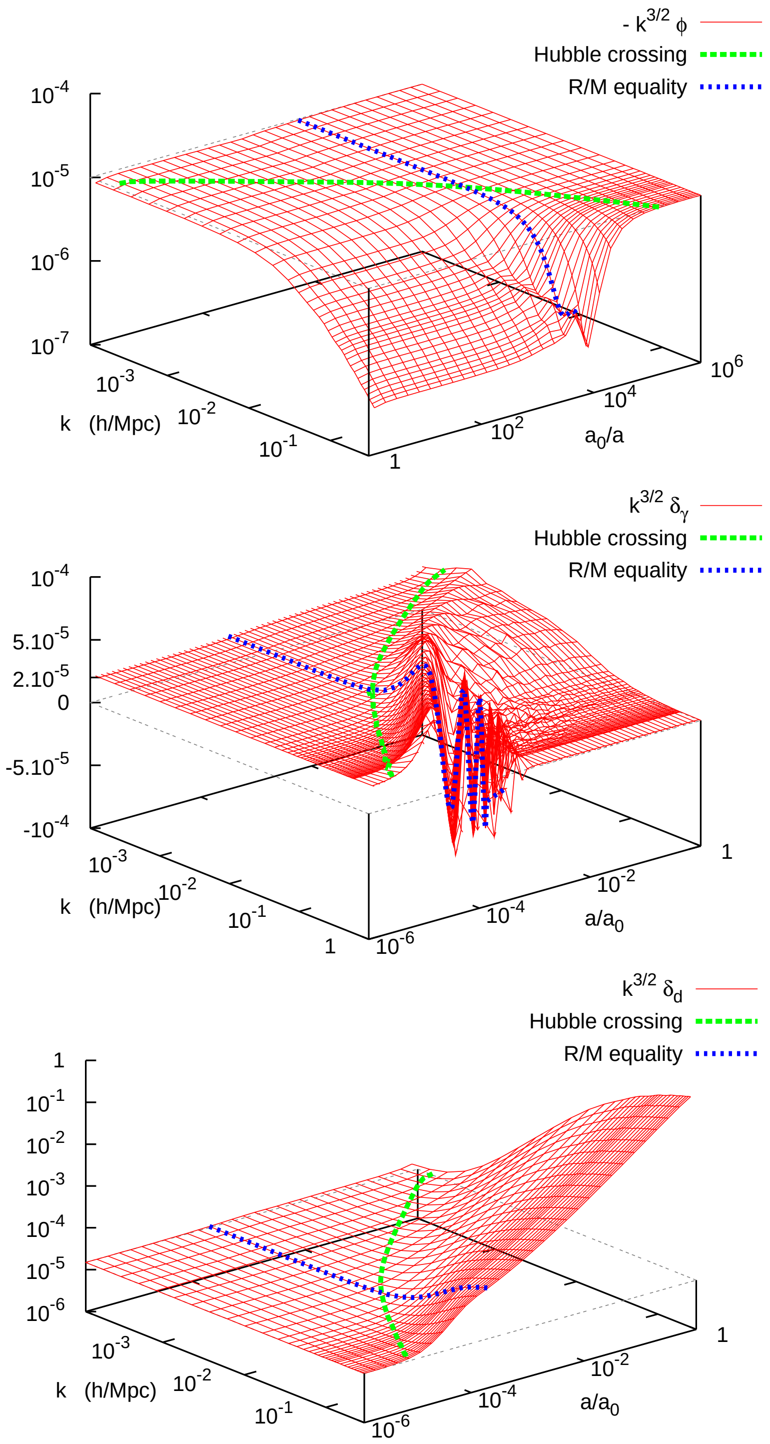

Figure 5.3: Evolution of the metric $\phi = \psi$ (top), photon density δ_γ (middle) and dark matter density perturbations $\delta_{\rm d}$ (bottom) in a neutrinoless model with $\bar{\rho}_b/\bar{\rho}_{cdm} = 0.2$, obtained numerically as a function of the scale factor a and Fourier wavenumber k. In order to get a better view, time is evolving from back to front in the upper plot, and from front to back in the lower two plots. The initial condition was set arbitrarily to $k^{3/2}\phi = -10^{-5}$. The blue line corresponds to radiation/matter equality, and the green line to Hubble crossing $(\lambda = R_H)$ for each mode.

like a^{-3}. The derivative of Eq. (5.88) gives a usefull relation:

$$\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G}{3} a^2 (2\bar{\rho}_\Lambda - \bar{\rho}_m) \ . \tag{5.89}$$

The pressure of baryons and dark matter is negligible, and the cosmological constant is homogeneous. So, the total pressure perturbation δp during this stage can be neglected, and Eq. (5.29) gives:

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} + \left(2\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2\right)\phi = 0 \ . \tag{5.90}$$

Using Eqs. (5.88, 5.89), we finally obtain:

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} + 8\pi G a^2 \rho_\Lambda \phi = 0 \ . \tag{5.91}$$

Comparing with the same equation during matter domination (5.79), we see that the effective squared mass term in front of ϕ vanishes during matter domination, but not when the cosmological constant starts to dominate: hence, ϕ=constant is a solution in the first case but not in the second one. A detailed integration of Eqs. (5.88, 5.91) would show that when $\bar{\rho}_\Lambda/\bar{\rho}_m$ starts to be sizeable, the gravitational potential starts to decay. Note that the coefficients of Eq. (5.91) are independent of k, so the decay is the same for all scales (above and below the Hubble radius). The Poisson equation (5.36) still holds inside the Hubble radius, so the decay of ϕ implies that δ_m does not grow as fast as the scale factor during Λ domination.

We conclude that a possible cosmological constant domination stage would slow down the growth of stuctures on all scales. This can be seen in the numerical solutions of Figure 5.3. The model used there has a cosmological constant corresponding today to $\Omega_\Lambda = 0.7$. In the upper plot, we see that when a approaches one, ϕ decays slightly on all scales. This effect remains very small, and it is even difficult to see it in the middle plot for δ_{cdm}.

5.3.3 Computing the matter power spectrum

We have seen in section 5.2.4 that the power spectrum of matter perturbations today, $P(k, \tau_0) = \left\langle |\delta_m(\mathbf{k}, \tau_0)|^2 \right\rangle$, is an observable quantity. Here, we will try compute it analytically (at least in an approximate way), using the results of the previous section. We must start from some initial conditions at a time τ_i, when all modes observable today are *outside* the Hubble radius during radiation domination. At that time, Eq. (5.75) gives the relation between the various perturbations and the random number $C_1(\mathbf{k})$ for each mode, but does not specify the primordial power spectrum $\langle |C_1(\mathbf{k})|^2 \rangle$ as a function of k.

The most natural initial condition would be a "scale-invariant spectrum", i.e. an initial condition for which the universe looks scale-free, roughly like a fractal distribution. This is the case if the primordial spectrum of each density perturbation multiplied by k^3 (i.e., the product $k^3 \langle |\delta_j(\mathbf{k}, \tau_i)|^2 \rangle$) is independent of k. This requires $\langle |C_1(\mathbf{k})|^2 \rangle$ to be proportional to k^{-3}: in this case,

$$\langle |\phi(\mathbf{k}, \tau_i)|^2 \rangle = \frac{1}{4} \langle |\delta_\gamma(\mathbf{k}, \tau_i)|^2 \rangle = \frac{4}{9} \langle |\delta_{b,cdm}(\mathbf{k}, \tau_i)|^2 \rangle = \frac{1}{81} \langle |C_1|^2 \rangle \propto k^{-3} \ . \tag{5.92}$$

Such a scale-invariant primordial spectrum is often called a *Harrison-Zel'dovitch spectrum*, from the name of two famous cosmologists. We will see in the chapter on inflation that the favorite mechanism for the generation of fluctuations predicts a nearly scale-invariant spectrum. Such initial conditions are also strongly indicated by observations. Hence, we will assume that the primordial spectrum is nearly scale-invariant, and that the deviation with respect to the Harrison-Zel'dovitch case can be represented by a power-law:

$$\langle|\phi(\mathbf{k},\tau_i)|^2\rangle = ... = Ak^{n-4} \ , \tag{5.93}$$

where A is some normalization factor for the primordial fluctuations, and the *spectral index* n is close to one by assumption.

Let us assume first that $\bar{\rho}_b/\bar{\rho}_m \ll 1$, and follow the evolution of $\delta_{cdm}(\mathbf{k},\tau)$, starting from some initial condition $\phi(\mathbf{k},\tau_i)$ for each mode. At the end of radiation domination, when $\tau=\tau_{eq}$:

1. some of the modes observable today are still outside the *Hubble radius at that time*. We have seen that for these modes:

$$\delta_{cdm}(\mathbf{k},\tau_{eq}) = -2\phi(\mathbf{k},\tau_i) \qquad \text{for } k\tau_{eq} \ll 1 \ . \tag{5.94}$$

2. some of the modes observable today are already inside the Hubble radius, and started to grow in such way that

$$\delta_{cdm}(\mathbf{k},\tau_{eq}) \propto \phi(\mathbf{k},\tau_i)\ln(k\tau_{eq}) \qquad \text{for } k\tau_{eq} \gg 1 \ . \tag{5.95}$$

In the limit $\bar{\rho}_b/\bar{\rho}_m \ll 1$, the dark matter drives gravity after equality. Hence, after $\tau=\tau_{eq}$, the gravitational potential freezes-out and the dark matter density grows like $(2+k^2\tau^2/6)$. In first approximation, we can do a matching between the solution of Eq. (5.83) and that from Eqs. (5.94, 5.95):

$$\begin{aligned} -\left(2+\frac{k^2\tau_{eq}^2}{6}\right)D_1(\mathbf{k}) &= -2\phi(\mathbf{k},\tau_i) \qquad \text{for } k\tau_{eq} \ll 1 \ , & (5.96)\\ &\propto \phi(\mathbf{k},\tau_i)\ln(k\tau_{eq}) \qquad \text{for } k\tau_{eq} \gg 1 \ . & (5.97)\end{aligned}$$

This gives the value of the coefficient $D_1(\mathbf{k})$:

$$\begin{aligned} D_1(\mathbf{k}) &\simeq \phi(\mathbf{k},\tau_i) \qquad \text{for } k\tau_{eq} \ll 1 \ , & (5.98)\\ &\propto \frac{\ln(k\tau_{eq})}{k^2\tau_{eq}^2}\,\phi(\mathbf{k},\tau_i) \qquad \text{for } k\tau_{eq} \gg 1 \ . & (5.99)\end{aligned}$$

Later on, during matter domination, the solution of Eq. (5.83) still applies. If we are still during matter domination today, we find:

$$\delta_{cdm}(\mathbf{k},\tau_0) = -\left(2+\frac{k^2\tau_0^2}{6}\right)D_1(\mathbf{k}) \ , \tag{5.100}$$

with $D_1(\mathbf{k})$ given by Eqs. (5.98, 5.99) in the two limits $k\tau_{eq} \ll 1$ and $k\tau_{eq} \gg 1$. However, we should remember that modes observable today are all inside the current Hubble radius, so $k\tau_0 \gg 1$. Hence the last result can be simplified in:

$$\begin{aligned} \delta_{cdm}(\mathbf{k},\tau_0) &\simeq -\frac{k^2\tau_0^2}{6}\,\phi(\mathbf{k},\tau_i) \qquad \text{for } k\tau_{eq} \ll 1 \ , & (5.101)\\ &\propto -\frac{k^2\tau_0^2}{6}\,\frac{\ln(k\tau_{eq})}{k^2\tau_{eq}^2}\,\phi(\mathbf{k},\tau_i) \qquad \text{for } k\tau_{eq} \gg 1 \ . & (5.102)\end{aligned}$$

Finally, using Eq. (5.93), we can write the matter power spectrum observable today as:

$$P(k) \simeq \frac{k^4\tau_0^4}{36} A k^{n-4} \qquad \text{for } k\tau_{eq} \ll 1 \,, \tag{5.103}$$

$$\propto \left(\frac{\tau_0^2}{\tau_{eq}^2}\ln(k\tau_{eq})\right)^2 A k^{n-4} \qquad \text{for } k\tau_{eq} \gg 1 \,. \tag{5.104}$$

We infer the two asymptotic behaviors of the matter power spectrum:

$$P(k) \propto k^n \qquad \text{for } k\tau_{eq} \ll 1 \,, \tag{5.105}$$

$$\propto \ln(k\tau_{eq})^2 k^{n-4} \qquad \text{for } k\tau_{eq} \gg 1 \,. \tag{5.106}$$

The behavior in the intermediate region ($k\tau_{eq} \sim 1$) can only be found with a detailed numerical integration, using the codes mentioned at the end of the last section. In Figure 5.4, the power spectrum $P(k)$ has been computed numerically for a model with $\bar{\rho}_b/\bar{\rho}_m = 0.18$ and $n = 1$, and is compared with the two above asymptotes, which turn out to be excellent approximations away from $k\tau_{eq} \sim 1$.

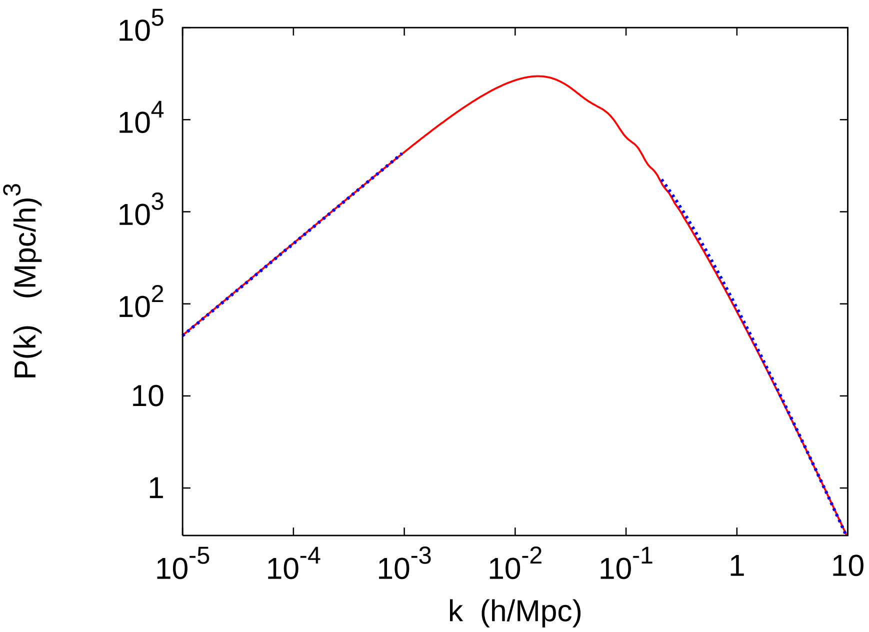

Figure 5.4: In red: matter power spectrum $P(k)$ computed numerically for a simple cosmological model in a flat universe, with $\bar{\rho}_b \ll \bar{\rho}_{cdm}$ (or in other words: $\Omega_b \ll \Omega_{cdm}$). In this case, the imprint of acoustic oscillations is very small (although still visible around $k \sim 0.1h$/Mpc) and the asymptotic approximations of Eq. (5.105, 5.106), corresponding to the blue dashed curves on the figure, work very well.

This was the result assuming $\bar{\rho}_b/\bar{\rho}_m \ll 1$. In the opposite limit $\bar{\rho}_b/\bar{\rho}_m \gg 1$, the gravitational potential follows the baryon perturbations during the intermediate stage, i.e. experiences damped acoustic oscillations inside the Hubble radius. At decoupling, ϕ freezes-out with an imprint of these oscillations, in equilibrium with the baryons: $\delta_b = -(k^2\tau^2/6)\,\phi$. Cold dark matter behaves like a test fluid and shares the same value $\delta_{cdm} = \delta_b$. During matter domination, $\delta_b = \delta_{cdm}$ grows in the usual way inside the Hubble radius, while ϕ remains frozen. So, in this case, the branch $k\tau_{eq} \ll 1$ of the potential is exactly the same, but the branch $k\tau_{eq} \gg 1$ drops more sharply as a function of k, with additional oscillations imprinted in the spectrum and frozen since that time.

In the intermediate case ($\bar{\rho}_b/\bar{\rho}_m \sim 1$), the power spectrum is in between these two cases: the solution for $k\tau_{eq} \gg 1$ departs from Eq. (5.102) through a smaller slope and a small imprint of acoustic oscillations.

Finally, if there is a recent stage of cosmological constant domination after matter domination, the amplitude of the power spectrum goes down at late times, but the shape is unchanged.

In summary, the matter power spectrum today should depend on the following quantities:

1. the overall normalization depends on the primordial spectrum amplitude A, on the age of the universe τ_0, and on the cosmological constant.

2. the overall slope depends on the primordial spectrum index n.

3. the scale k_{max} of the maximum in $P(k)$ depends on the time of equality τ_{eq}, i.e. of the matter-to-radiation ratio $(\bar{\rho}_b + \bar{\rho}_{cdm})/\bar{\rho}_r$. The radiation density is fixed by the CMB temperature, so in fact τ_{eq} only depends on the matter density today, i.e. on $\omega_b + \omega_{cdm} = (\Omega_b + \Omega_{cdm})h^2$.

4. the shape of the spectrum for $k > k_{max}$ depends on n, but also on $\bar{\rho}_b/\bar{\rho}_{cdm} = \Omega_b/\Omega_{cdm}$ (a high baryon density implies a lower amplitude for $k > k_{max}$, as well as additional oscillations).

The last three effects are illustrated in Figure 5.5. For simplicity, we do not discuss here the effect of neutrinos and of a possible spatial curvature on the power spectrum.

5.3.4 Computing the temperature anisotropy spectrum

In Chapter 2, we have seen that in the homogeneous case the photons have a Planckian distribution before and after decoupling:

$$f_\gamma(p,t) = \frac{1}{e^{\frac{p}{T(t)}} - 1} \tag{5.107}$$

with $T(t) \propto 1/a(t)$. Here, in perturbed cosmology, the distribution depends on many more arguments, namely $(\mathbf{p}, \mathbf{x}, t)$. However, thermal equilibrium still holds locally before decoupling. Hence, the dependence with respect to $p = |\mathbf{p}|$ must still be of the Planckian type. If the unit vector $\mathbf{n}$ denotes the direction of the momentum ($\mathbf{n} = \mathbf{p}/p$), we can write the perturbed photon distribution as:

$$f_\gamma(p, \mathbf{n}, \mathbf{x}, t) = \frac{1}{e^{\frac{p}{\bar{T}+\delta T}} - 1} \tag{5.108}$$

where the average $\bar{T}$ is a function of t only, while the perturbation δT depends on all arguments excepted p: namely, $\delta T = \delta T(\mathbf{n}, \mathbf{x}, t)$.

Let us work in the limit of instantaneous decoupling: we assume that for $\tau \leq \tau_{\rm dec}$, the photons are tightly coupled with baryons and electrons, while for $\tau > \tau_{\rm dec}$ they are fully decoupled.

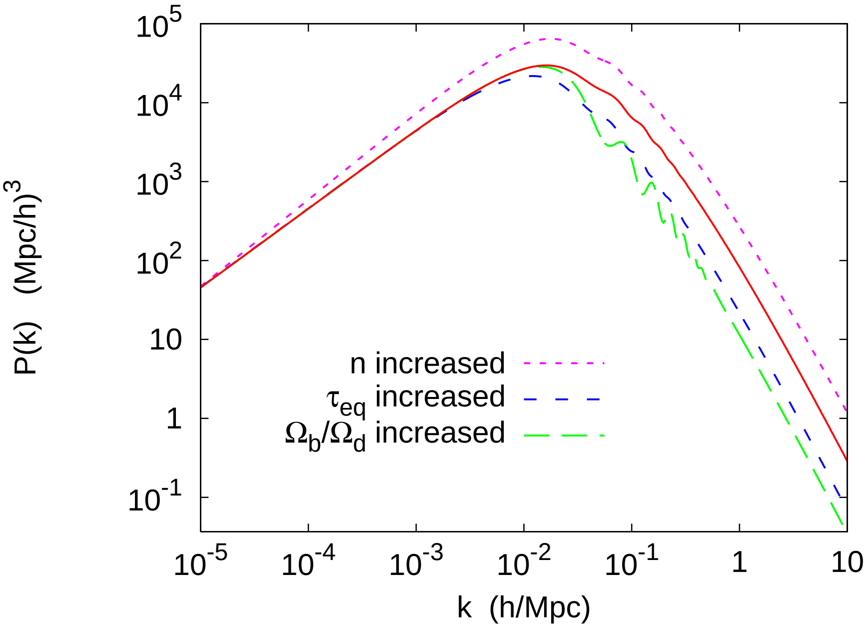

Figure 5.5: The red solid line shows the same reference matter power spectrum $P(k)$ as in the previous figure. The other lines show the effect on $P(k)$ of increasing one of the following quantities: n (the overall slope of the power spectrum increases); τ_{eq} (the scale of the maximum k_{max} decreases); and finally ω_b/ω_{cdm} (the power on scales $k > k_{max}$ is suppressed and the imprint of acoustic oscillations is more visible).

- **Before decoupling.** In the previous sections, we said that in this regime the photons can be treated as a perfect fluid, with only two independent variables $\{\delta_\gamma, \theta_\gamma\}$. Can we reconcile this simple fluid description with the more complicated framework introduced above? We can, because as long as photons are strongly coupled, the function $(\delta T/\bar{T})(\mathbf{n}, \mathbf{x}, t)$ takes a very particular form. First, we notice that an observer comoving with the photon fluid would see the same value of the temperature in all directions. This follows from the fact that in thermal equilibrium, the relation $\rho_\gamma \propto T^4$ holds locally, hence $\delta_\gamma = 4\delta T/\bar{T}$. Hence, the observer comoving with the photon fluid would see a temperature perturbation

$$\frac{\delta T}{\bar{T}}(\mathbf{n}, \mathbf{x}, t) = \frac{1}{4}\delta_\gamma(\mathbf{x}, t) \tag{5.109}$$

with no dependence on $\mathbf{n}$. However, in general, the tightly-coupled photon fluid is not at rest with respect to the coordinate system: it has a bulk velocity $\mathbf{v}_\gamma(\mathbf{x}, t)$. So, an observer at rest with the coordinate system will see a Doppler effect induced by the motion of the fluid with respect to him. The Doppler effects shifts the wavelength of each photon according to $\delta\lambda/\lambda = (\mathbf{v}_\gamma.\mathbf{n})$. Since the blackbody temperature is inversely proportional to the average photon wavelength, the Doppler effect induces a correction $\delta T/T = -\delta\lambda/\lambda = -(\mathbf{v}_\gamma.\mathbf{n})$ to the temperature perturbation. We conclude that the observer at rest with the coordinate system would see a temperature given by

$$\frac{\delta T}{\bar{T}}(\mathbf{n},\mathbf{x},t) = \frac{1}{4}\delta_\gamma(\mathbf{x},t) - \mathbf{n}.\mathbf{v}_\gamma(\mathbf{x},t) \ . \tag{5.110}$$

In other words, if the temperature anisotropy in a given point was expanded in spherical harmonics, there would only be a monopole and a dipole. We conclude that before decoupling, the photons need not be described by a complicated function $(\delta T/\bar{T})(\mathbf{n},\mathbf{x},t)$, but just by one function of space $\delta_\gamma(\mathbf{x},t)$ and one vector field $\mathbf{v}_\gamma(\mathbf{x},t)$. If we go to Fourier space and introduce as usual the velocity divergence θ (other terms in the velocity field contribute to vector perturbations), we conclude that in Fourier space, tightly coupled photons are just described by the two functions $\delta_\gamma(\mathbf{k},t)$ and $\theta_\gamma(\mathbf{k},t)$. This is exactly what we assumed in the previous sections of this chapter. Actually, it is even possible to show that the perturbed Boltzmann equation giving the time-evolution of $f_\gamma(p,\mathbf{n},\mathbf{x},t)$ reduces to the continuity and Euler equations for $\delta_\gamma(\mathbf{k},t)$ and $\theta_\gamma(\mathbf{k},t)$ in this regime.

- **After decoupling.** In this regime, thermal equilibrium does not hold anymore. It is possible to show that for decoupled photons, the distribution remains Planckian and is still in the form of eq. (5.108). However $\delta T/\bar{T}$ really becomes a complicated function of $(\mathbf{n},\mathbf{x},t)$ and eq. (5.110) is not true anymore. We can avoid complications if we admit one simple equation (deriving this equation would take a few pages, but is not too difficult). First, we recall that at the level of homogeneous cosmology, photons traveling from the LSS to us and observed in a direction $-\mathbf{n}$ (where $\mathbf{n}$ is the direction of their momentum) travel along a geodesic simply parametrized by $\mathbf{r}(\tau) = (\tau - \tau_0)\mathbf{n}$. This trajectory starts from the coordinate $\mathbf{r}(\tau_{\rm dec}) = (\tau_{\rm dec} - \tau_0)\mathbf{n}$ and reaches us in straight line. In perturbed cosmology, we expect deviations from this straight line, due to gravitational lensing effects (between the LSS and us, the photons travel through structures associated to gravitational potential maxima and minima). This is true, but these deviations only matter for the computation of the CMB spectrum at order two in perturbations. At order one in perturbations, we can still treat the geodesics as straight lines with $\mathbf{r}(\tau) = (\tau - \tau_0)\mathbf{n}$. Let us go back to the photon distribution. The evolution of $f_\gamma(p,\mathbf{n},\mathbf{x},t)$ is given by a complicated perturbed Boltzmann equation. However, if we assume that photons are decoupled and if we integrate this Boltzmann equation along a given photon geodesic $\mathbf{r}(\tau)$, we will find that

$$\frac{d}{d\tau}\left(\frac{\delta T}{\bar{T}}(\mathbf{n},\mathbf{x},t) + \phi(\mathbf{x},t)\right) = \frac{\partial}{\partial\tau}\left(\phi(\mathbf{x},t) + \psi(\mathbf{x},t)\right) \ , \tag{5.111}$$

where $\frac{d}{d\tau}$ is the total derivative along the geodesic. We already assumed that the metric perturbations ϕ and ψ are equal to each other. This differential equation can be formally integrated between $\tau_{\rm dec}$ and τ_0, giving:

$$\left[\frac{\delta T}{\bar{T}}(\mathbf{n},\mathbf{x},t) + \phi(\mathbf{x},t)\right]_{\rm O}^{\rm LSS} = \int_{\tau_{\rm dec}}^{\tau_0} d\tau\, 2\dot{\phi}([\tau - \tau_0]\mathbf{n},\tau) \tag{5.112}$$

where LSS means: in the point of coordinates $(\mathbf{x} = [\tau_{\rm dec} - \tau_0]\mathbf{n}, \tau = \tau_{\rm dec})$, and O means: in the point of coordinates $(\mathbf{x} = \vec{o}, \tau = \tau_0)$. Note that the integral over $\dot{\phi}$ is not equal to $[\phi]_{\rm O}^{\rm LSS}$ because $\dot{\phi}$ is only the partial derivative of ϕ with respect to τ. According to equation (5.110), we can replace $\frac{\delta T}{\bar{T}}$ in LSS by $(\frac{1}{4}\delta_\gamma - \mathbf{n}.\mathbf{v}_\gamma)$. Finally, we notice that $\frac{\delta T}{\bar{T}}$ in the point 0 is the observable temperature anisotropy in the direction $-\mathbf{n}$ (photons traveling in the direction $\mathbf{n}$ are seen by us as coming from a direction $-\mathbf{n}$). We get:

$$\frac{1}{4}\delta_\gamma|_{\mathrm{LSS}} - \mathbf{n}.\mathbf{v}_\gamma|_{\mathrm{LSS}} + \phi|_{\mathrm{LSS}} - \left.\frac{\delta T}{\bar{T}}\right|_{\mathrm{obs}}(-\mathbf{n}) - \phi|_0 = 2\int_{\tau_{\mathrm{dec}}}^{\tau_0} d\tau\, \dot{\phi}. \quad (5.113)$$

After a transformation $\mathbf{n} \longrightarrow -\mathbf{n}$ which makes the final result more readable, we get the following expression for the observable temperature anisotropy in direction $\mathbf{n}$:

$$\begin{aligned}\left.\frac{\delta T}{\bar{T}}\right|_{\mathrm{obs}}(\mathbf{n}) \quad = \quad & \frac{\delta T}{\bar{T}}([\tau_{\mathrm{dec}} - \tau_0]\mathbf{n}, \tau_{\mathrm{dec}}) + \phi([\tau_{\mathrm{dec}} - \tau_0]\mathbf{n}, \tau_{\mathrm{dec}}) - \phi(\vec{o}, \tau_0) \\ & + \mathbf{n}.\mathbf{v}_\gamma([\tau_{\mathrm{dec}} - \tau_0]\mathbf{n}, \tau_{\mathrm{dec}}) \\ & + 2\int_{\tau_{\mathrm{dec}}}^{\tau_0} d\tau\, \dot{\phi}([\tau_0 - \tau]\mathbf{n}, \tau)\ . \end{aligned} \quad (5.114)$$

On the right-hand side, the terms on the first line stand for the so-called Sachs-Wolfe contribution to observable anisotropies. They account for the fact that when we measure temperature fluctuations today, we see the intrinsic temperature fluctuation on the last scattering surface, given by $\frac{1}{4}\delta_\gamma|_{\mathrm{LSS}}$, corrected by a "gravitational Doppler effect": indeed, the total amount by which the photon is redshifted between the last scattering surface and now does not depend only on the ratio a_0/a_{dec}, but also on the local value of the gravitational potential at the emission and reception points. This "gravitational Doppler effect" is also called the Sachs-Wolfe effect. The second term is really important because $\phi([\tau_0 - \tau_{\mathrm{dec}}]\mathbf{n}, \tau_{\mathrm{dec}})$ is different in each point of the last scattering surface. Let us consider an overdensity on the last scattering surface, i.e. a point where the intrinsic temperature fluctuation is positive, $\delta_\gamma > 0$ (this place is called a "hot spot"). This overdensity is responsible for a gravitational potential well: $\phi < 0$. When climbing out of this gravitational potential well, photons will loose energy and get redshifted. The second effect wins: $\frac{1}{4}\delta_\gamma + \phi < 0$. This means that a hot spot on the last scattering surface is actually seen as a cold spot in the CMB anisotropy map. Conversely, photons coming from a cold spot are blueshifted and appear as hot spots in the maps.

The third term on the right-hand side of eq. (5.114), $-\phi(\vec{o}, \tau_0)$, is irrelevant in practice, because it is the same for all observed direction: so, it contributes to the average temperature (by a negligible fraction) and observers automatically absorb it in $\bar{T}$.

The term in the second line stands for the usual Doppler effect, as we already discussed.

The last term in the third line is called the integrated Sachs-Wolfe term. It depends on the history of the photons along each line of sight. If the gravitational potential was static, this term would vanish. The reason is that when photons travel through maxima/minima of a static gravitational potential, the redhsifts and blueshifts experienced along the trajectory cancel each other apart from a net effect $\phi|_{\mathrm{LSS}} - \phi|_0$, already included in the Sachs-Wolfe term. Instead, when the gravitational potential is non-static, these redhsifts/blueshifts accumulate along the line of sight, and produce extra temperature anisotropies called "secondary anisotropies" (since they are not related to the last scattering surface).

Overall shape of the temperature power spectrum C_l. The angular correlation function $F(\theta)$ and the harmonic power spectrum C_l represent the two-point correlation function of $\delta T/\bar{T}$, respectively in real and harmonic space. We showed already that for a given θ or a given l, they receive mainly contributions from Fourier modes in the LSS, with a correspondence between θ, l and k given by eq. (5.63). Looking at equation (5.114) and neglecting for the moment the integrated Sachs-Wolfe term, we see that $F(\theta)$ or C_l must be related to the Fourier power spectra of $[\frac{1}{4}\delta_\gamma + \phi]$ and of θ_γ, evaluated at the time $\tau_{\rm dec}$ and for

$$k \sim \frac{2\pi a_{\rm dec}}{d_A(z_{\rm dec})\,\theta} \sim \frac{2a_{\rm dec}\,l}{d_A(z_{\rm dec})} . \tag{5.115}$$

Actually, most representations of the CMB two-point correlation function show the quantity $l^2 C_l$ as a function of l. It is possible to show that $l^2 C_l$ is related in first approximation to the spectra

$$l^2 C_l \longleftrightarrow \left\{ k^3 \langle | \frac{1}{4}\delta_\gamma + \phi |^2 \rangle \quad , \quad k^3 \langle | \frac{\theta_\gamma}{k} \rangle \right\} . \tag{5.116}$$

The power spectrum of $[\frac{1}{4}\delta_\gamma + \phi]$ at the time of radiation/matter equality is given approximately by Eqs. (5.69, 5.71) with $\tau = \tau_{eq}$, neglecting the decaying mode ($C_2 = 0$) and adopting the same primordial spectrum as in the previous section: $\langle |\phi(\mathbf{k}, \tau_i)|^2 \rangle = Ak^{n-4} = \langle |C_1(\mathbf{k})|^2 \rangle / 81$, with a spectral index n assumed to be close to one. For modes which did not enter into the sound horizon at the time of equality ($kc_s\tau_{eq} \ll 1$), the result is:

$$\langle | \frac{1}{4}\delta_\gamma(\mathbf{k}, \tau_{eq}) + \phi(\mathbf{k}, \tau_{eq}) |^2 \rangle = \left(\frac{(-\frac{1}{2} + 1)}{9} \right)^2 \langle |C_1(\mathbf{k})|^2 \rangle = \frac{1}{4} Ak^{n-4} , \tag{5.117}$$

while for modes which entered into the sound horizon before equality ($kc_s\tau_{eq} \gg 1$):

$$\langle | \frac{1}{4}\delta_\gamma(\mathbf{k}, \tau_{eq}) + \phi(\mathbf{k}, \tau_{eq}) |^2 \rangle = \frac{9}{4} \cos^2(kc_s\tau_{eq}) Ak^{n-4} . \tag{5.118}$$

According these approximate solutions, $k^3\langle|\frac{1}{4}\delta_\gamma + \phi|^2\rangle$ is nearly flat for $kc_s\tau_{eq} \ll 1$ (if $n \simeq 1$), and behaves like $\cos^2(kc_s\tau_{eq})$ for $kc_s\tau_{eq} \gg 1$. This just reflects the fact that at equality, modes outside the sound horizon are still frozen and given by the initial (nearly flat) spectrum, while modes inside the sound horizon did have time to oscillate one, two, three or more times between horizon crossing and equality. The number of oscillations depends on the ratio between the sound horizon and the wavelength of the mode.

The two-point correlation function of CMB anisotropies is related to $k^3\langle|\frac{1}{4}\delta_\gamma + \phi|^2\rangle$ at decoupling, not at equality. We have already mentioned that between equality and decoupling, δ_γ undergoes a rather complicated evolution, with damped oscillations, depending very much on the baryon-to-dark matter density ratio.

In addition, the observable anisotropies should be corrected by the Doppler effect. For a harmonic oscillator, the velocity is out of phase with the position. Here, similarly, the bulk velocity of the photon-baryon fluid and hence the Doppler effect is out of phase with the oscillations of δ_γ and ϕ. So, for values of k such that $\cos(kc_s\tau_{eq}) = 0$, the Doppler effect is not null, and the total observable spectrum of CMB anisotropies never reaches zero.

The next effect to take into account is the integrated Sachs-Wolfe effect (the last term in Eq. (5.114)), that depends on the time-variation of the gravitational potential $\dot{\phi}$ along the photon geodesics. We have seen that during full matter domination, when the total energy density scales like $\bar{\rho} \propto a^{-3}$, the potential ϕ is constant in time, both inside and outside the Hubble radius: hence, the integrated Sachs-Wolfe effect cancels during this stage. A non-zero effect can occur either at the beginning of matter domination, when radiation is not completely negligible and $\bar{\rho}$ is not yet exactly proportional to a^{-3}, or at the end of matter domination if a cosmological constant (or a spatial curvature term) starts to dominate today. The first case leads to the so-called *early integrated Sachs-Wolfe effect*, which is strong when equality takes place just before decoupling, and weak when equality takes place a long time before decoupling. This effect is known to enhance the *first* observable acoustic peak, i.e. the first oscillation of the modes entering inside the sound horizon just before decoupling. The second case (due to a cosmological constant or to curvature in the recent universe) is called the *late integrated Sachs-Wolfe effect*, and is known to enhance the anisotropy spectrum only for the *largest* observable wavelength, which entered recently inside the Hubble radius.

The last important effect affecting the CMB temperature spectrum is the so-called *Silk damping* effect, named after a famous English cosmologist. It is related to the fact that decoupling is not instantaneous. The mean free path of the photon does not go from zero to infinity instantaneously: it increases progressively during recombination. Before decoupling, in the tightly coupled photon-baryon fluid, the photons are distributed according to the plasma temperature in each point. If decoupling was instantaneous, the photons would travel freely from a point in the tightly-coupled fluid to us; hence photons arriving from a direction $\mathbf{n}$ would carry information about $\delta T/T([\tau_0 - \tau_{\rm dec}]\mathbf{n}, \tau_{\rm dec})$ (as we assumed throughout section 5.2.4). In fact, after leaving thermal equilibrium, each photon can experience a small number of interactions (elastic scatterings with electrons or baryons) changing their trajectory. Hence, when they reach us, they carry information not exactly about the plasma temperature in the point $\mathbf{x} = [\tau_0 - \tau_{\rm dec}]\mathbf{n}$, but in a slightly different point and direction. The "aberration angle" is of the order of the mean free path of the photon around the time of decoupling divided by the angular diameter distance of the last scattering surface, $d_A(z_{\rm dec})$. On angles smaller than this characteristic angle, we cannot relate the observed temperature anisotropies to the Fourier spectrum of the fluid temperature at decoupling: hence, the angular correlation function $\left\langle \frac{\delta T}{T}(\mathbf{n}) \frac{\delta T}{T}(\mathbf{n}') \right\rangle$ falls down and reaches asymptotically zero for small angular separation (same for C_l at large l). In summary, Silk damping produces a cut-off in the observable spectrum of CMB anisotropies.

All these effects are taken into account when the CMB temperature spectrum is computed numerically. In figure 5.6, we show a typical result for standard cosmological parameters. The vertical axis corresponds to $l^2 C_l$ in some units. The horizontal axis corresponds to l, i.e. to the inverse of the angular separation θ. On the figure, we can clearly see the plateau on large scales (left) and the acoustic peaks on small scales (right). The plateau is not exactly flat due to the late integrated Sachs-Wolfe effect. The first peak is much higher than the other ones due to the early integrated Sachs-Wolfe effect. The relative amplitude of the next peaks depends on many parameters, in particular on the time of equality and on the baryon-to-dark matter density ratio. Finally, the spectrum is suppressed exponentially on small scales due to the Silk damping effect.

Parameter dependence of the temperature power spectrum.

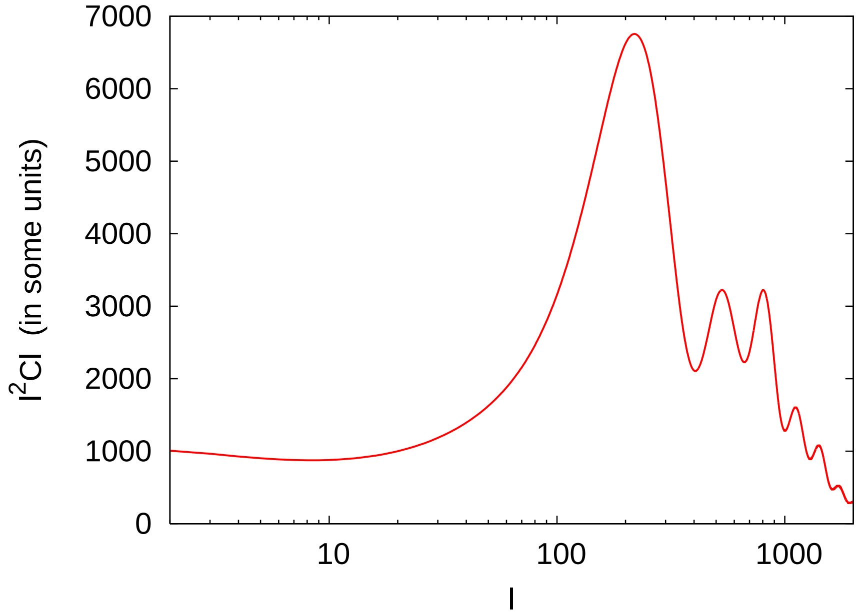

Figure 5.6: The CMB temperature power spectrum $l^2 C_l$ as a function of $l \sim \pi/\theta$, computed numerically for a simple cosmological model in a flat universe, with $\Omega_b/\Omega_{cdm} = 0.18$, $\Omega_\Lambda = 0.73$ and $n = 1$.

Let us now summarize the effect of the main cosmological parameters on the CMB temperature spectrum. For simplicity, we go on assuming a flat universe and neglecting the effect of neutrinos. We also neglect the so-called "reionisation of the universe" at late times. However, we leave the possibility of a cosmological constant parametrized by Ω_Λ and dominating the energy density of the recent universe. With such assumptions, the cosmological model can be described with five parameters: the cosmological constant fraction Ω_Λ, the total non-relativistic matter density $\omega_\mathrm{m} = \Omega_\mathrm{m} h^2 = (\Omega_b + \Omega_{cdm})h^2$, the baryon density $\omega_b = \Omega_b h^2$, the primordial spectrum amplitude A and finally its spectral index n. Since we assume a flat universe, the reduced Hubble parameter h can be inferred from ω_m and Ω_Λ:

$$1 = \Omega_\Lambda + \Omega_m = \Omega_\Lambda + \frac{\omega_m}{h^2} \qquad \Rightarrow \qquad h = \sqrt{\frac{\omega_m}{1-\Omega_\Lambda}} \, . \tag{5.119}$$

These parameters control various physical effects which are responsible for the shape of the CMB temperature power spectrum:

1. **The time of Radiation/Matter equality**. In the above parameter basis, the time of equality between $\bar{\rho}_\mathrm{m}$ and $\bar{\rho}_\mathrm{r}$ is fixed by ω_m only, since $(a_\mathrm{eq}/a_0) = \bar{\rho}^0_\mathrm{r}/\bar{\rho}^0_\mathrm{m}$ (the script ^{0} means "evaluated today"), and $\bar{\rho}^0_\mathrm{r}$ is fixed by the CMB temperature. A late equality implies less time between equality and decoupling, and less damping for the acoustic oscillations of modes entering inside the sound horizon during radiation domination. It also implies more early integrated Sachs-Wolfe effect. So, a later equality induces higher CMB peaks, especially for the first one.

2. **The time of Matter/Λ equality**. If the cosmological constant is larger, equality between matter and Λ takes place earlier. During Λ domination, the time-variation of the metric perturbations leads to a late integrated Sachs-Wolfe effect, and the spectrum on the largest angular scales is enhanced.

3. **The angular scale of the sound horizon at recombination**. The characteristic scale of the oscillations in the CMB power spectrum is set by $c_s \tau_{dec}$, or more precisely by the sound horizon $d_s(\tau_{\rm rec})$ at recombination. The time $\tau_{\rm rec}$ is fixed by thermodynamics, but the sound horizon is an integral over $c_s dt/a(t)$ between 0 and $\tau_{\rm rec}$. This integral depends on the time of equality, and on the baryon density at late times (through c_s). In addition, the actual observable quantity is the *angular scale* of this sound horizon on the last scattering surface, which is given by the ratio of $d_s(\tau_{\rm rec})$ over the angular diameter distance $d_A(z_{\rm rec})$. The latter also depends on h and Ω_Λ. We conclude that the observed angular scale of the peaks constrains a combination of the three parameters Ω_Λ, $\omega_{\rm m}$ and ω_b.

4. **The coupling between gravity and acoustic oscillations between the time of radiation-matter equality and photon decoupling.** We have seen that between radiation-matter equality and decoupling, the evolution of the metric perturbations can be driven either by dark matter if $\omega_b \ll \omega_{cdm}$, or by baryons if $\omega_b \gg \omega_{cdm}$, or by a combination of both if $\omega_b \sim \omega_{cdm}$. This affects considerably the evolution of acoustic oscillations during this intermediate stage: the oscillation amplitude and equilibrium point are different if the photon-baryon fluid behaves like is a test fluid or drives the evolution of the gravitational potential. The main effect of increasing the ratio ω_b/ω_{cdm} is to enhance the first peak and slightly suppress the second one.

5. **The amplitude of primordial perturbations**, defined as the parameter A in Eq. (5.93), obviously fixes the global normalization of the CMB spectrum .

6. **The spectral index of primordial perturbations**, defined as the parameter n in Eq. (5.93), fixes the overall slope of the CMB spectrum.

In Fig. 5.7, we illustrate graphically the consequence of varying either n, $\omega_{\rm m}$, Ω_Λ or $\omega_{\rm b}$, while keeping all the other cosmological parameters fixed. Increasing n just changes the overall slope of the power spectrum. When we decrease $\omega_{\rm m}$, the relevant effects are (1) and (3): equality is postponed, boosting the CMB peaks and especially the first one; simultaneously, there is a tiny horizontal shift of the spectrum corresponding (smaller angular scale of the sound horizon at recombination). When we increase Ω_Λ, we can see the effects (2) and (3): the spectrum increase on the largest angular scales (left of the figure), and the spectrum is shifted to the left (larger angular scale of the sound horizon at recombination). When $\omega_{\rm b}$ is increased, the effects (3) and (4) are taking place: the first CMB peak increases, while the second one decreases slightly; the scale of the peaks is also shifted (smaller angular scale of the sound horizon at recombination).

We have seen that A and n have very specific effects, while (Ω_Λ, $\omega_{\rm m}$, ω_b) have intricate effects. Nevertheless, since there are six physical effects for five parameters, it is in principle possible to measure all of them with CMB data, assuming a flat universe.

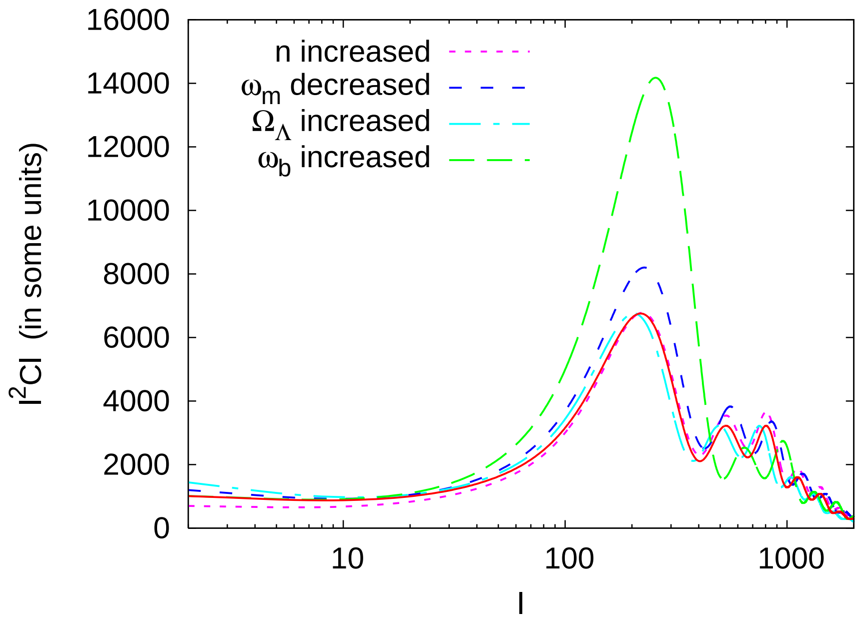

Figure 5.7: The red solid line shows the same reference CMB temperature spectrum as in the previous figure. The other lines show the effect of varying one of the following quantities: n; ω_m (and the time of radiation-matter equality τ_{eq}); ω_Λ (and the time of matter-Λ equality); and finally ω_b (and the baryon-to-dark matter density ratio). The corresponding effects are described in the text.

Finally, let us mention the consequences of varying the spatial curvature. We know that the curvature parameter has a crucial effect on the angular diameter distance – redshift relation. Hence, we expect a change in Ω_k to change the correspondence between Fourier modes on the last scattering surface and angles of observation today: this should correspond in a horizontal shift of the CMB temperature spectrum (and of the scale of all acoustic peaks). Also, if in the Friedmann equation the curvature term starts to dominate over the matter density term in the recent universe, the metric perturbation ϕ is not anymore constant at late time (the explanation is the same as for a cosmological constant: the effective mass square in Eq. (5.29) does not vanish as long as the expansion is not lead by the $\bar{\rho}_m \propto a^{-3}$ term, so ϕ decays). So, in case of spatial curvature, we also expect the late integrated Sachs-Wolfe effect to modify the large-scale CMB spectrum. These two effects are illustrated in figure 5.8.

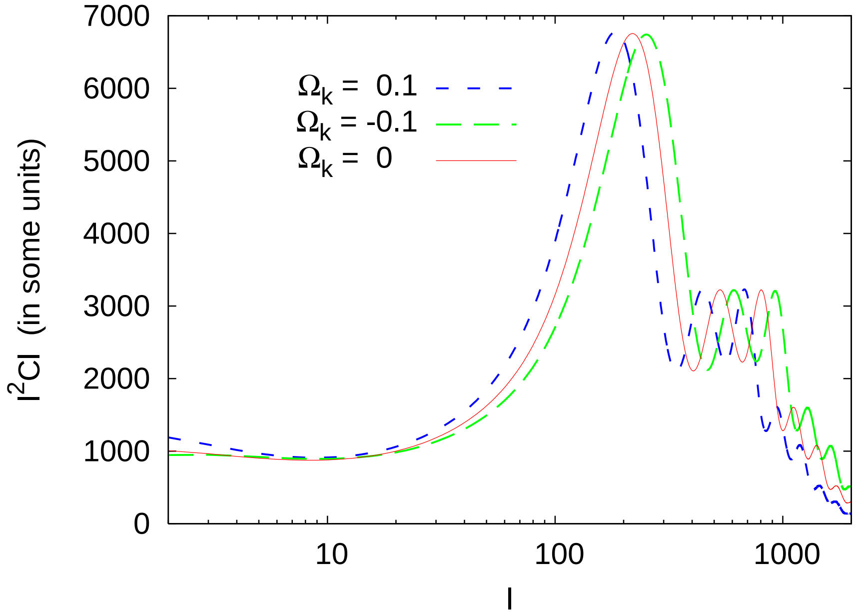

Figure 5.8: The red solid line shows the same reference CMB temperature spectrum as in the previous figure. The other lines show the effect of taking $\Omega_k = 0.1$ (positively curved universe) or $\Omega_k = -0.1$ (negatively curved universe), all other parameters being fixed. The corresponding effects are described in the text.

Chapter 6

Cosmological observations

According to the previous sections, it is reasonable to assume that the cosmological scenario can be parametrized by:

- the total matter density ω_m and the baryon density ω_b (the dark matter density is then given by $\omega_{cdm} = \omega_m - \omega_b$).
- a possible cosmological constant density fraction Ω_Λ and spatial curvature density fraction Ω_k.
- the primordial spectrum amplitude A and spectral index n (see Eq. (5.93)).

The total radiation density is not a free parameter, since the photon density is fixed by the CMB temperature today:

$$\begin{aligned} \omega_\gamma &\equiv \Omega_\gamma h^2 \\ &= \frac{\bar{\rho}_\gamma^0}{\bar{\rho}_c^0} h^2 \\ &= \left(\frac{\pi^2}{15} T_0^4\right) \left(\frac{8\pi G}{3H_0^2}\right) h^2 \\ &= \frac{8\pi^3 T_0^4}{45 (H_0/h)^2 M_P^2} \ , \end{aligned} \tag{6.1}$$

while the neutrino density relative to that of photons is fixed by the assumption that each of the three neutrino families has a Planckian distribution with $T_\nu = (4/11)^{1/3} T$, so that:

$$\begin{aligned} \omega_r &\equiv \omega_\gamma + \omega_\nu \\ &= \left[\frac{\pi^2}{15} T_0^4 + 3 \times \frac{7}{8} \times \frac{\pi^2}{15} T_{\nu 0}^4\right] \left(\frac{8\pi G}{3H_0^2}\right) h^2 \\ &= \left[1 + 3 \times \frac{7}{8} \times \left(\frac{4}{11}\right)^{4/3}\right] \omega_\gamma \\ &\sim 4 \times 10^{-5} \qquad \text{for} \ \ T_0 = 2.726 \text{ K} \ . \end{aligned} \tag{6.2}$$

For $\Omega_k = 0$, this model is usually called "flat ΛCDM" or just "ΛCDM", since besides baryons and radiation it contains two major ingredients: cold dark matter (CDM) and a cosmological constant Λ. We see that ΛCDM has five independent free parameters, e.g. $\{\omega_m, \omega_b, \Omega_\Lambda, A, n\}$. The same model with a sixth independent parameter Ω_k describing positive (resp. negative) curvature is usually called "closed ΛCDM" (resp. "open ΛCDM").

No.	Reaction	Type	No.	Reaction	Type
1	$n \longrightarrow p$	*weak*	22	^{6}Li + p $\longrightarrow \gamma$ + ^{7}Be	(p,γ)
2	^{3}H $\rightarrow \bar{\nu}_e + e^- +$ ^{3}He	*weak*	23	^{6}Li + p $\longrightarrow$ ^{3}He + ^{4}He	^{3}He Pickup
3	^{8}Li $\rightarrow \bar{\nu}_e + e^- +$ 2 ^{4}He	*weak*	24	^{7}Li + p $\longrightarrow$ ^{4}He + ^{4}He	^{4}He Pickup
4	^{12}B $\rightarrow \bar{\nu}_e + e^- +$ ^{12}C	*weak*	24 bis	^{7}Li + p $\longrightarrow \gamma$ + ^{4}He + ^{4}He	(p,γ)
5	^{14}C $\rightarrow \bar{\nu}_e + e^- +$ ^{14}N	*weak*	25	^{4}He + ^{2}H $\longrightarrow \gamma$ + ^{6}Li	(d,γ)
6	^{8}B $\rightarrow \nu_e + e^+ +$ 2 ^{4}He	*weak*	26	^{4}He + ^{3}H $\longrightarrow \gamma$ + ^{7}Li	(t,γ)
7	^{11}C $\rightarrow \nu_e + e^+ +$ ^{11}B	*weak*	27	^{4}He + ^{3}He $\longrightarrow \gamma$ + ^{7}Be	(^{3}He,γ)
8	^{12}N $\rightarrow \nu_e + e^+ +$ ^{12}C	*weak*	28	^{2}H + ^{2}H $\longrightarrow$ n + ^{3}He	^{2}H Strip.
9	^{13}N $\rightarrow \nu_e + e^+ +$ ^{13}C	*weak*	29	^{2}H + ^{2}H $\longrightarrow$ p + ^{3}H	^{2}H Strip.
10	^{14}O $\rightarrow \nu_e + e^+ +$ ^{14}N	*weak*	30	^{3}H + ^{2}H $\longrightarrow$ n + ^{4}He	^{2}H Strip.
11	^{15}O $\rightarrow \nu_e + e^+ +$ ^{15}N	*weak*	31	^{3}He + ^{2}H $\longrightarrow$ p + ^{4}He	^{2}H Strip.
12	p + n $\longrightarrow \gamma$ + ^{2}H	(n,γ)	32	^{3}He + ^{3}He $\longrightarrow$ p + p + ^{4}He	(^{3}He,$2p$)
13	^{2}H + n $\longrightarrow \gamma$ +^{3}H	(n,γ)	33	^{7}Li + ^{2}H $\longrightarrow$ n + ^{4}He + ^{4}He	(d,n α)
14	^{3}He + n $\longrightarrow \gamma$ + ^{4}He	(n,γ)	34	^{7}Be + ^{2}H $\longrightarrow$ p + ^{4}He + ^{4}He	(d,p α)
15	^{6}Li + n $\longrightarrow \gamma$ + ^{7}Li	(n,γ)	35	^{3}He + ^{3}H $\longrightarrow \gamma$ + ^{6}Li	(t,γ)
16	^{3}He + n $\longrightarrow$ p + ^{3}H	charge ex.	36	^{6}Li + ^{2}H $\longrightarrow$ n + ^{7}Be	^{2}H Strip.
17	^{7}Be + n $\longrightarrow$ p + ^{7}Li	charge ex.	37	^{6}Li + ^{2}H $\longrightarrow$ p + ^{7}Li	^{2}H Strip.
18	^{6}Li + n $\longrightarrow$ ^{3}H + ^{4}He	^{3}H Pickup	38	^{3}He + ^{3}H $\longrightarrow$ ^{2}H + ^{4}He	(^{3}H,d)
19	^{7}Be + n $\longrightarrow$ ^{4}He + ^{4}He	^{4}He Pickup	39	^{3}H + ^{3}H $\longrightarrow$ n + n + ^{4}He	(t,n n)
20	^{2}H + p $\longrightarrow \gamma$ + ^{3}He	(p,γ)	40	^{3}He + ^{3}H $\longrightarrow$ p + n + ^{4}He	(t,n p)
21	^{3}H + p $\longrightarrow \gamma$ + ^{4}He	(p,γ)			

Table 6.1: The first forty reactions used in the nucleosynthesis code PARTENOPE. *Table taken from [arXiv:0705.0290] by Ofelia Pisanti et al.*

The questions to address now are: is this ΛCDM model able to explain all cosmological observations (with ot without spatial curvature)? If yes, does the data provide a measurement of all the above parameters? If not, what kind of new physical ingredient is needed? We will review here the main cosmological observations and their implications for cosmological parameters. The order of the next sections corresponds more or less to the order in which each observations started to play a crucial role for measuring cosmological parameters over the last twenty years.

6.1 Abundance of primordial elements

In sections 3.3.5, we have seen that the theory of nucleosynthesis can predict the abundance of light elements formed in the early universe, when the energy density was of order $\rho \sim (1\,\mathrm{MeV})^4$. After nucleosynthesis, there are no more nuclear reactions in the universe, excepted in the core of stars. So, today, in regions of the universe which were never filled by matter ejected from stars, the proportion of light elements is still the same as it was just after nucleosynthesis. Fortunately, the universe contains clouds of gas fullfilling this criteria, and the abundance of deuterium, helium, etc. can be measured in such regions (e.g. by spectroscopy). The results can be directly compared with theoretical predictions.

The predictions presented in this course were based on a very simplistic description of nucleosynthesis. Precise predictions arise from codes simulating the evolution of a system of many different reactions. Table 6.1 shows, for instance, the first 40 reactions used in the public code PARTHENOPE (`http://parthenope.na.infn.it/`). In the section 3.3.5, we only studied the reactions called 1 and 12 in this table.

Numerical simulation of nucleosynthesis accurately predict all relative abundances as a function of the only free parameter in the theory, the baryon density. We remember that the temperature at which light elements start forming is fixed by equation (3.56) and depends on $\eta_b \equiv n_b/n_\gamma \sim 10^{-10}$, which precise value is given by $\eta_b = 5.5 \times 10^{-10}(\omega_b/0.020)$ (note that η_b is defined at any time between positron annihilation and today: it is constant in this range). Hence, relative abundances depend on ω_b, as already mentioned for helium in Eq. (3.70). Figure 6.1 shows the dependence of the abundance of ^{4}He, D, ^{3}He and ^{7}Li as a function of η_b.

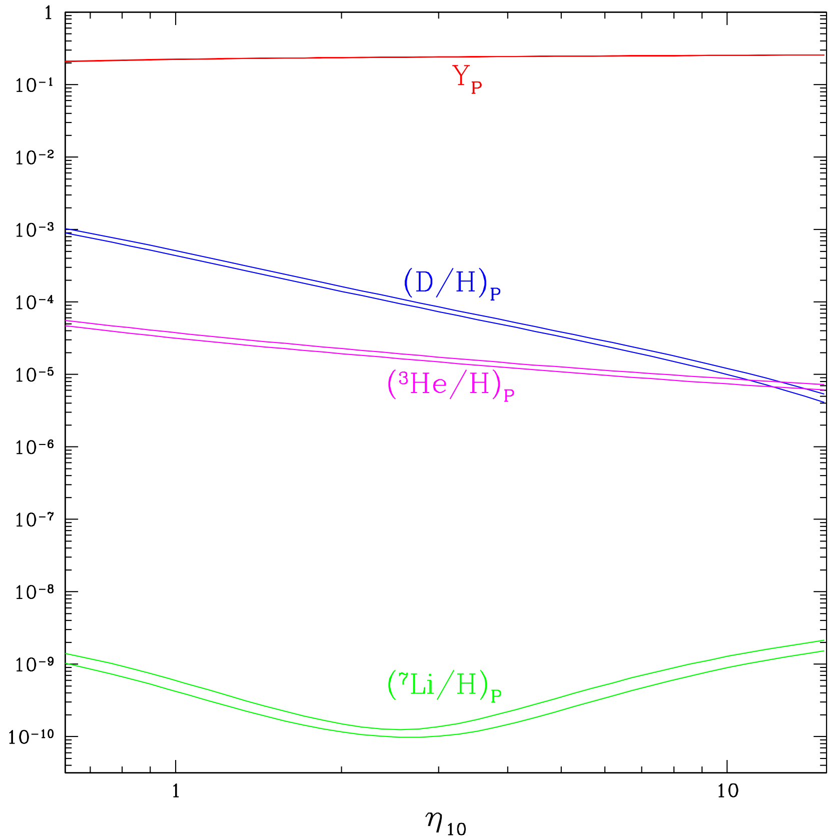

Figure 6.1: The nucleosynthesis-predicted primordial abundances of D, ^{3}He, ^{7}Li (relative to hydrogen by number), and the ^{4}He mass fraction (Y_P), as functions of the baryon abundance parameter $\eta_{10} \equiv 10^{10}\eta_b$. The widths of the bands reflect the uncertainties in the nuclear and weak interaction rates. *Plot taken from Int.J.Mod.Phys. E15 (2006) 1-36 [arXiv:astro-ph/0511534v1] by Gary Steigman.*

Current observations (mainly of ^{4}He and D) show that

$$\omega_b \equiv \Omega_b h^2 = 0.020 \pm 0.002. \tag{6.3}$$

Hence, for $h = 0.7$, the baryon fraction is of the order of $\Omega_b \sim 0.04$: approximately four percent of the universe density is due to ordinary matter. This is already more than the sum of all luminous matter, which represents one per cent: so, 75% of ordinary matter is not even visible.

Note that if ω_r was a free parameter, the outcome of nucleosynthesis would also depend crucially on ω_r. So, nucleosynthesis can also be used as a tool for testing the fact that Eq. (6.2) is correct. It turns out to be the case: primordial element abundances provide a measurement of ω_r precise at the 10% level, and perfectly compatible with Eq. (6.2).

6.2 Age of the universe

The age of the universe can be conveniently computed once the function $H(a)/H_0$ or $H(z)/H_0$ is known. This function follows from the Friedmann equation divided by H_0^2:

$$\begin{aligned}\frac{H^2}{H_0^2} &= \frac{\bar\rho_{tot}}{\bar\rho_c} - \frac{k}{a^2H_0^2} \\ &= \Omega_r\left(\frac{a_0}{a}\right)^4 + \Omega_m\left(\frac{a_0}{a}\right)^3 - \Omega_k\left(\frac{a_0}{a}\right)^2 + \Omega_\Lambda \qquad (6.4) \\ &= \Omega_r\,(1+z)^4 + \Omega_m\,(1+z)^3 - \Omega_k\,(1+z)^2 + \Omega_\Lambda\ , \qquad (6.5)\end{aligned}$$

with the constraint that $\Omega_r + \Omega_m - \Omega_k + \Omega_\Lambda = 1$ by construction. Since $H = da/(adt)$, we can write:

$$dt = \frac{da}{aH} = -\frac{dz}{(1+z)H}\ . \qquad (6.6)$$

Hence, the age of the universe can be computed from the integral

$$t = \int_0^{a_0} \frac{da}{aH} = H_0^{-1}\int_0^{a_0}\frac{da}{a}\left(\frac{H_0}{H(a)}\right)\ , \qquad (6.7)$$

or equivalently from

$$t = \int_0^{\infty}\frac{dz}{(1+z)H} = H_0^{-1}\int_0^{\infty}\frac{dz}{1+z}\left(\frac{H_0}{H(z)}\right)\ . \qquad (6.8)$$

This integral converges with respect to the boundary corresponding to the initial singularity, $a \longrightarrow 0$ or $z \longrightarrow \infty$. Actually, it is easy to show that the radiation dominated period gives a negligible contribution to the age of the universe, hence the term proportional to Ω_r can be omitted in the integral. If the universe is matter-dominated today ($\Omega_\Lambda = \Omega_k = 0$), then $\Omega_m = 1$ and the age of the universe is simply given by:

$$t = H_0^{-1}\int_0^\infty dz\,(1+z)^{-5/2} = \frac{2}{3H_0} = 6.52h^{-1}\mathrm{Gyr}\ , \qquad (6.9)$$

where 1 Gyr $\equiv$ 1 billion years. If $\Omega_\Lambda > 0$ and/or $\Omega_k < 0$ (negatively curved universe), the ratio $H(z)/H_0$ decreases with respect to the $\Omega_\Lambda = \Omega_k = 0$ case for all values of z corresponding to Λ or curvature domination. For $\Omega_k > 0$ (closed universe), it increases. Hence, the age of the universe increases with respect to $6.52h^{-1}$Gyr if $\Omega_\Lambda > 0$ and/or $\Omega_k < 0$, and decreases if $\Omega_k > 0$.

The age of of a few specific object in the universe can be evaluated with a number of techniques, e.g. by nucleochronology (studying the radioactive decay of isotopes inside an object, exactly like in the ^{14}C method used in archeology); or by measuring the cooling of stars in their final state, called "white dwarfs", and comparing with the mean evolution curve of white dwarfs; etc. If the age of an object is found to be extremely large, it provides a lower bound on the age of the universe itself. Current observations can set a reliable lower bound on the age of the universe: $t >$ 11Gyr. This is incompatible with the matter-dominated universe of Eq. (6.9) unless $h < 0.59$, while observations of the Hubble flow prefer $h \sim 0.7$. Hence, these observations provide a strong hint that that the universe is either negatively curved or Λ–dominated today. This "age problem" was already known in the 90's.

6.3 Luminosity of Type Ia supernovae

The evidence for a non–flat universe and/or a non–zero cosmological constant has increased considerably in 1998, when two independent groups studied the apparent luminosity of distant type Ia supernovae (SNIa). For this type of supernovae, astronomers believe that there is a simple relation between the absolute magnitude and the luminosity decay rate. In other words, by studying the rise and fall of the luminosity curve during a few weeks, one can deduce the absolute magnitude of a given SNIa. Therefore, it can be used in the same way as cepheids, as a probe of the luminosity distance – redshift relation. In addition, supernovae are much brighter that cepheids, and can be observed at much larger distances (until redshifts of order one or two). While observable cepheids only probe short distances, where the luminosity distance – redshift relation only gives the Hubble law (the proportionality between distance and redshift), the most distant observable SNIa's are in the region where general relativity corrections are important: so, they can provide a measurement of the scale factor evolution (see section 2.2.2).

We note first that the observation of nearby cepheids and supernovae gives the following estimate of the reduced Hubble parameter:

$$h = 0.742 \pm 0.036 \tag{6.10}$$

at the 68% confidence level *(Riess et al. [E-print: 0905.0695])*.

On figure 6.2, the various curves represent the effective magnitude–redshift relation, computed for various choices of Ω_{M} and Ω_Λ. The effective magnitude m_B plotted here is essentially equivalent to the luminosity distance d_L, since it is proportional to $\log[d_L]$ plus a constant. For a given value of H_0, all the curves are asymptotically equal at short distance. Significant differences show up only at redshifts $z > 0.2$. Each red data point corresponds to a single supernovae in the first precise data set: that of the "Supernovae Cosmology Project", released in 1998. Even if it is not very clear visually from the figure, a detailed statistical analysis of this data revealed that a flat matter–dominated universe (with $\Omega_m = 1$, $\Omega_\Lambda = 0$) was excluded. This result has been confirmed by various more recent data sets. The top panel of figure 6.3 shows the luminosity distance – redshift diagram for the SNLS data set released in 2005. The corresponding constraints on Ω_m and Ω_Λ are displayed in Figure 6.4, and summarized by:

$$(\Omega_m - \Omega_\Lambda, \Omega_m + \Omega_\Lambda) = (-0.49 \pm 0.12, 1.11 \pm 0.52)\ . \tag{6.11}$$

Hence, supernovae data strongly suggest the existence of a cosmological constant today ($\Omega_\Lambda > 0$). In fact, the small luminosity of high-redshift supernovae suggests that the universe is currently in accelerated expansion. The supernovae data does not say whether the parameter Ω_k is negligible, positive or negative.

6.4 CMB temperature anisotropies

The order of magnitude of CMB anisotropies was predicted many years before being measured. By extrapolating from the present inhomogeneous structure back to the time of decoupling, many cosmologists in the 80's expected $\delta T/\bar{T}$ to be at least of order 10^{-6} – otherwise, clusters of galaxies could not have formed today.

Many experiments were devoted to the detection of these anisotropies. The first successful one was COBE-DMR, an American satellite carrying

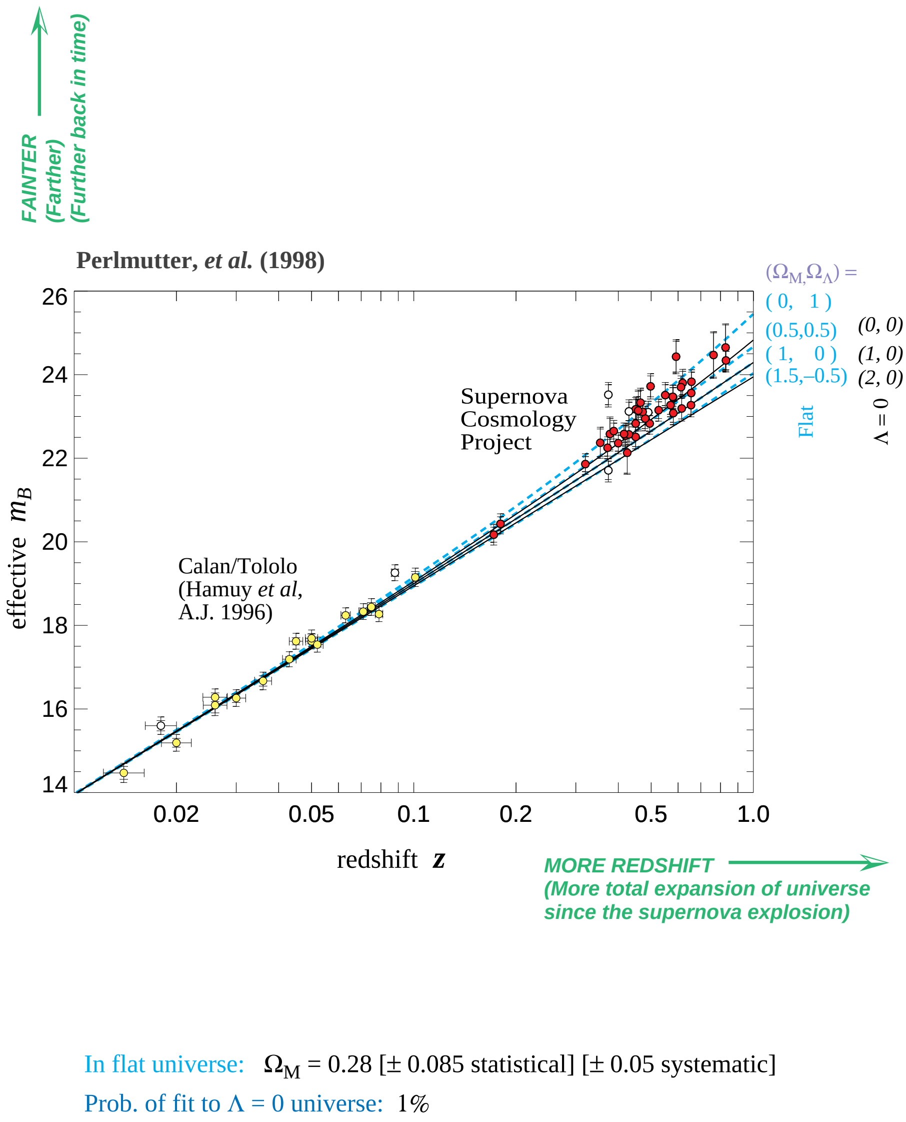

Figure 6.2: The results published by the "Supernovae Cosmology Project" in 1998 (see Perlmutter et al., Astrophys.J. 517 (1999) 565-586). The various curves represent the effective magnitude–redshift relation, computed for various choices of Ω_m and Ω_Λ. This plot is equivalent to a luminosity distance – redshift relation (effective magnitude and luminosity distance can be related in a straightforward way: $m_B \propto (\log[d_L] + \mathrm{cst})$). The solid black curves account for three examples of a universe with positive/null/negative curvature and no cosmological constant. The dashed blue curves correspond to three spatially flat universes with different values of Ω_Λ. For a given value of H_0, all the curves are asymptotically equal at short distance, probing only the Hubble law. The yellow points are short–distance SNIa's: we can check that they are approximately aligned. The red points, at redshifts between 0.2 and 0.9, show that distant supernovae are too faint to be compatible with a flat matter–dominated universe $(\Omega_m, \Omega_\Lambda)$ =(1,0).

an interferometer of exquisite sensitivity. In 1992, COBE mapped the anisotropies all over the sky, and found an average amplitude $\delta T/\bar{T} \sim 10^{-5}$ (see figure 6.5). This was in perfect agreement with the theoretical predictions – another big success for cosmology. The COBE experiment had an angular resolution of a few degrees: so, anisotropies seen under one degree or less were smoothed by the detector. In a Fourier decomposition, it means that COBE could only measure the spectrum of wavelengths larger than the sound horizon at decoupling. So, it was not probing the acoustic oscillations, but only the flat plateau. Hence, after 1992, considerable efforts were devoted to the design of new experiments with better angular resolution, in order to probe smaller wavelengths, check the existence of the acoustic peaks, compare them with theoretical predictions and measure the related cosmological parameters.

For instance, some decisive progresses were made with Boomerang, a US–Italian–Canadian balloon, carrying some detectors called bolometers. In 2001, Boomerang published the map of figure 6.6. It focuses on a small patch of the sky, but with much better resolution than COBE (a few arc–minutes). The Fourier decomposition of the Boomerang map clearly showed the first three acoustic peaks (see figure 6.7). Let us recall that the angular size of the first peak probes the angular diameter distance at the redshift of photon decoupling, and depends heavily on the spatial curvature parameter Ω_k. The position of the first peak measured by Boomerang was perfectly consistent with $\Omega_k = 0$. Boomerang brought the first convincing arguments in favor of an exactly flat or at least nearly flat universe. The combination of Boomerang data with supernovae observations started to show that the preferred values of Ω_m and Ω_Λ were around 0.3 and 0.7 respectively.

At the beginning of 2003, the NASA satellite WMAP published a full-sky CMB map shown in the lower part of figure 6.8. This was the second one after that of COBE, with a resolution increased by a factor 30. The corresponding temperature spectrum is shown in figure 6.9. It is in surprisingly good agreement with the predictions of a flat ΛCDM model. Since then, WMAP has improved its measurements by accumulating more years of observations, while ground-based experiments have produced some maps of small regions of the sky only, but with better resolution than WMAP, allowing to compute the power spectrum on smaller angular scales. Figure 6.10 shows a compilation of recent data (although all the latest data are not included here). Acoustic peaks are now clearly visible till the fifth one.

There are still many CMB experiments going on. In May 2009, the European satellite PLANCK has been be launched. It is currently collecting data. PLANCK is expected to perform the best possible measurement of the CMB temperature anisotropies, with such precision that most cosmological parameters should be measured at the per cent level.

At the moment, the position of the first acoustic peak combined with SNLS supernovae data provides the following constraint on the curvature of the universe: $\Omega_k = 0.011 \pm 0.012$. Hence, the spatial curvature is either null or tiny. Given that this measurement is compatible with zero, it is reasonable to make the assumption that $\Omega_k = 0$, since this is easier to explain theoretically than a very small value of the order of 10^{-2}.

Assuming that $\Omega_k = 0$, the combination of all CMB experiments provides the following constraints on the parameters of the ΛCDM model:

$$\begin{aligned} \omega_m &= 0.135 \pm 0.007 \\ \omega_b &= 0.0227 \pm 0.0006 \\ \Omega_\Lambda &= 0.74 \pm 0.03 \end{aligned}$$

$$n = 0.963 \pm 0.015 \tag{6.12}$$

(we don't report here the measurement of the primordial amplitude A, which is not very interesting by itself). From these estimates, one can derive some bounds on the reduced Hubble parameter: $h = 0.72 \pm 0.03$, and on the age of the universe: $t_0 = 13.6 \pm 0.1$ Gyr. These results are in remarkable agreement with other independent techniques. For instance, the measurement of h is perfectly consistent with measurements of the same parameters using Hubble diagrams. But the most striking feature is probably the remarkable agreement between the values of ω_b inferred from the CMB and from primordial element abundances. These are completely independent techniques for probing the baryon density in the universe: one relies on nuclear physics when the universe had a temperature of $\sim$1 MeV, the other on relativistic fluid mechanics at temperature of the order of $\sim$1 eV. The perfect overlap between the two constraints indicates that our knowledge of the universe is impressively good, at least during the epoch between $t \sim 1$ s and $t \sim 100\ 000$ yr. Let us finally emphasize that CMB observations alone bring some very strong evidence in favor of the existence of dark matter: from the above bounds, one can infer that

$$\omega_{cdm} = 0.110 \pm 0.006 \ . \tag{6.13}$$

6.5 Galaxy correlation function

Over the past decades, astronomers could build some very large three-dimensional maps of the galaxy distribution – on larger scales than the scale of non-linearity. On figure 6.11, we can see the galaxy distribution reconstructed by the SDSS collaboration within a thin slice of the surrounding universe. We can see that the galaxy distribution looks really homogeneous on large scales. Actually, the average density seems to decrease slightly at large z, but this is simply a so-called *selection effect*: since the telescope is limited in sensitivity, it can see most nearby galaxies, and only a selection of the brightest remote galaxies. Observers know how to model this effect, and how to correct it when computing the luminous galactic matter power spectrum $P_{\rm lgm}(k)$.

On figure 6.12, we see the Fourier spectrum reconstructed from the map of figure 6.11 for the same two galaxy samples, which have different bias factors (and hence different normalisations of $P_{\rm lgm}(k)$). The red line shows the theoretical prediction for the ΛCDM model which gives the best fit to WMAP data, for two adjusted values of the bias. The shape of the theoretical $P(k)$ agrees very well with the observed shapes. In the LRG power spectrum, the first data points (with the smallest values of k) start to probe the maximum in $P(k)$ (corresponding to scales entering inside the Hubble radius around the time of matter–radiation equality).

On the theoretical curves, the baryon acoustic oscillations are clearly visible. Are they visible also in the data? The answer is yes, but this is seen more clearly in figure 6.13 where the same data points are plotted with different units.

Since the baryon acoustic oscillations (called usually BAO) are observable and corresponds to a known comoving scale (that of the sound horizon at the time of photon decoupling), we can use it as a standard ruler. Not entering into details, let us mention that it is possible to select all galaxies with a given approximate redshift, and for these galaxies, to measure the power spectrum in angular space (i.e., as a function of angular separation

θ), exactly like for CMB anisotropies. This angular power spectrum has a clear imprint of BAO: hence, we can measure the angle under which BAO are seen for a given redshift. This technique is similar in principle to the measurement of the angular scale of the first CMB peak, but it applies to much smaller redshifts. Hence, by observing BAO, people are able to map the angular diameter–redshift relation for small redshifts (typically smaller than one). This information is complementary to that of type Ia supernovae, and tends to be more precise. This technique has emerged only recently (see *Astrophys.J.633:560-574,2005 [astro-ph/0501171] by D. Eisenstein et al.*) and provides very good complementary constraints, in particular on Ω_Λ.

We don't provide here detailed bounds on cosmological parameters from these two techniques (measurement of the galaxy power spectrum, and of the scale of BAO), but they are in very good agreement with the previously mentioned results from CMB experiments. They confirm the validity of the ΛCDM scenario with the parameters listed in Eqs. (6.12), (6.13). This ΛCDM model is often called the *concordance model*.

6.6 Other observations not discussed here

For concision, we will not address in this course other techniques which might become particularly important in the future: the study of galaxy cluster abundances as a function of redshift; surveys of peculiar velocities; analyses of Lyman-α forests in the spectrum of quasars; galaxy weak lensing and cosmic shear (discussed in the course of Pierre Salati); CMB weak lensing; the study of the 21cm absorption line in gas clouds; etc.

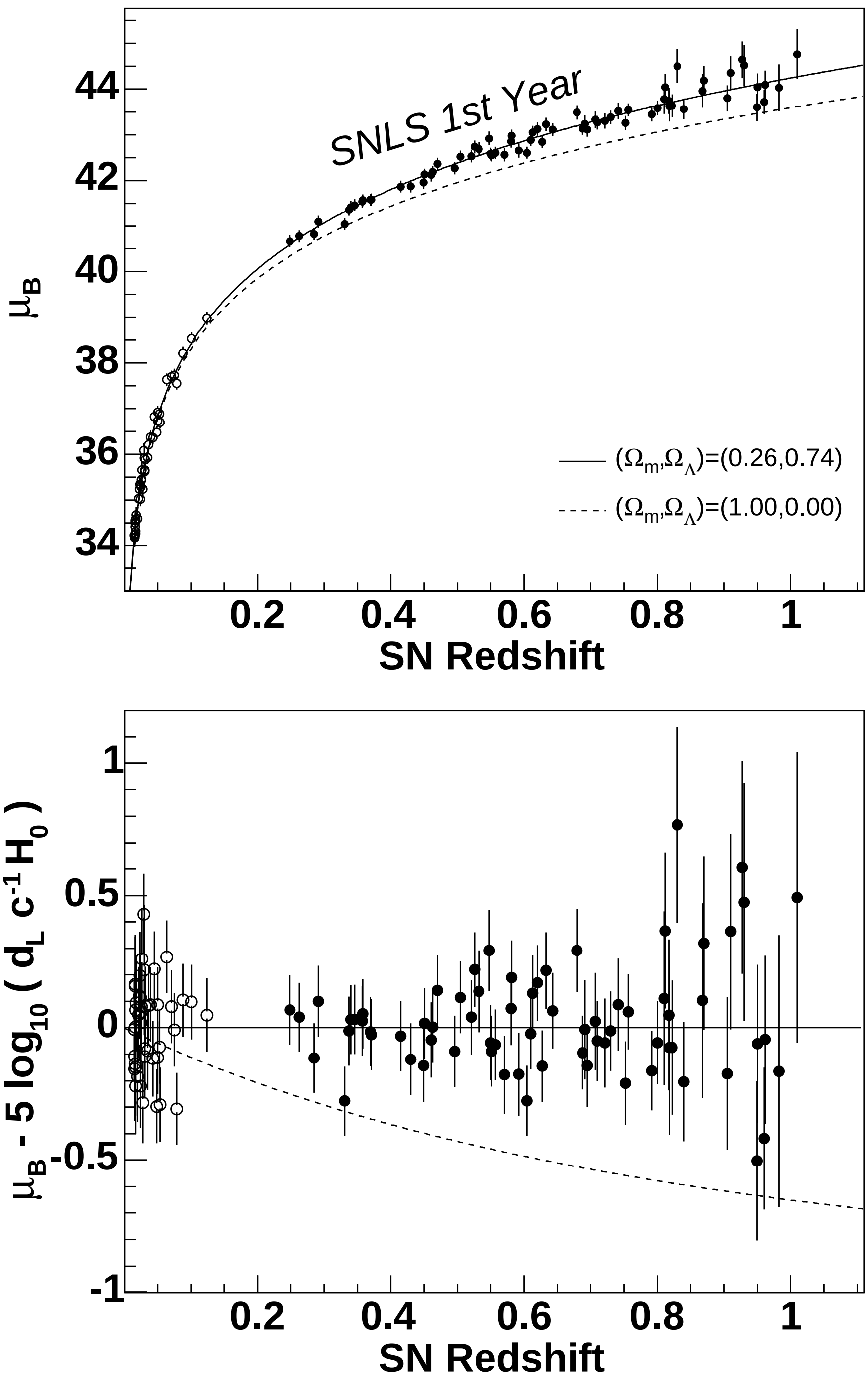

Figure 6.3: (Top panel) Same kind of luminosity distance – redshift diagram as in the previous figure, but for more recent data published by the SNLS collaboration in 2005. (Lower panel) Same data points and errors, divided by the theoretical prediction for the best fit ΛCDM model. *Plot taken from Astronomy and Astrophysics 447: 31-48, 2006 [e-Print: astro-ph/0510447] by Pierre Astier et al.*

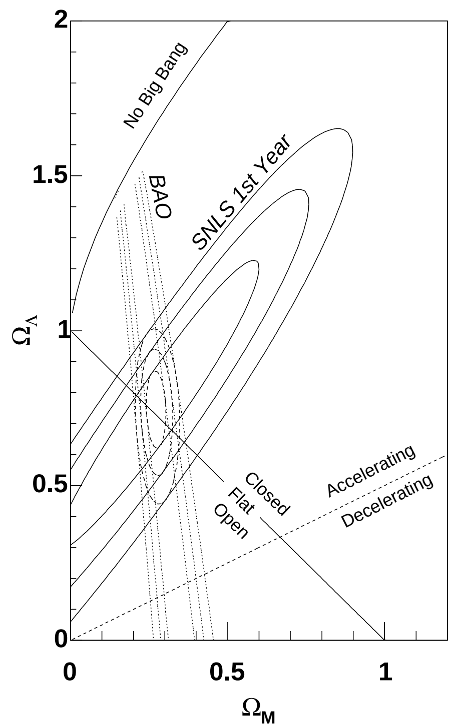

Figure 6.4: Contours at 68.3%, 95.5% and 99.7% confidence levels in the (Ω_m, Ω_Λ) plane from the SNLS supernovae data (solid contours), the SDSS baryon acoustic oscillations (see section 6.5, dotted lines), and the joint confidence contours (dashed lines). These plots are all assuming a ΛCDM cosmology, as we are doing in this chapter. *Plot taken from Astronomy and Astrophysics 447: 31-48, 2006 [e-Print: astro-ph/0510447] by Pierre Astier et al.*

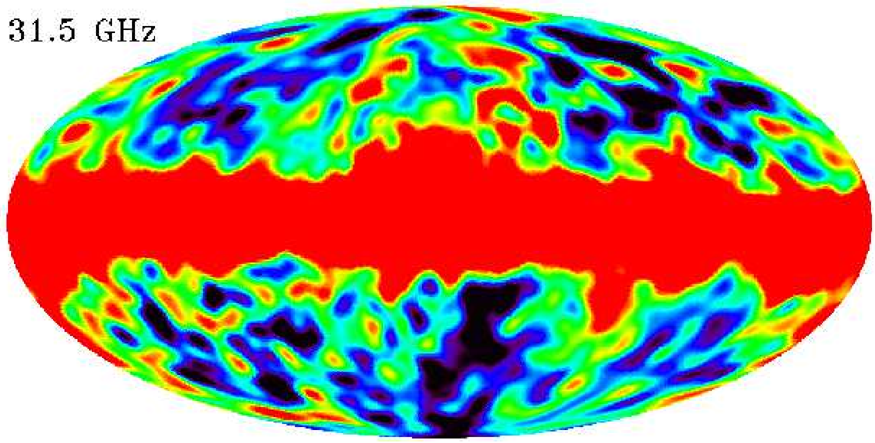

Figure 6.5: The first genuine "picture of the universe" at the time of decoupling, 370 000 years after the initial singularity, and 13.6 billion years before the present epoch. Each blue (resp. red) spot corresponds to a slightly colder (resp. warmer) region of the universe at that time. This map, obtained by the American satellite COBE in 1994 (see C. L. Bennett et al., Astrophys.J. 464 (1996) L1-L4), covers the entire sky: so, it pictures a huge sphere centered on us (on the picture, the sphere has been projected onto an ellipse, where the upper and lower points represent the direction of the poles of the Milky way). Away from the central red stripe, which corresponds to photons emitted from our own galaxy, the fluctuations are only of order 10^{-5} with respect to the average value $T_0 = 2.728$ K. They are the "seeds" for the present structure of the universe: each red spot corresponds to a small over-density of photons and baryons at the time of decoupling, that has been enhanced later, leading to galaxies and clusters of galaxies today.

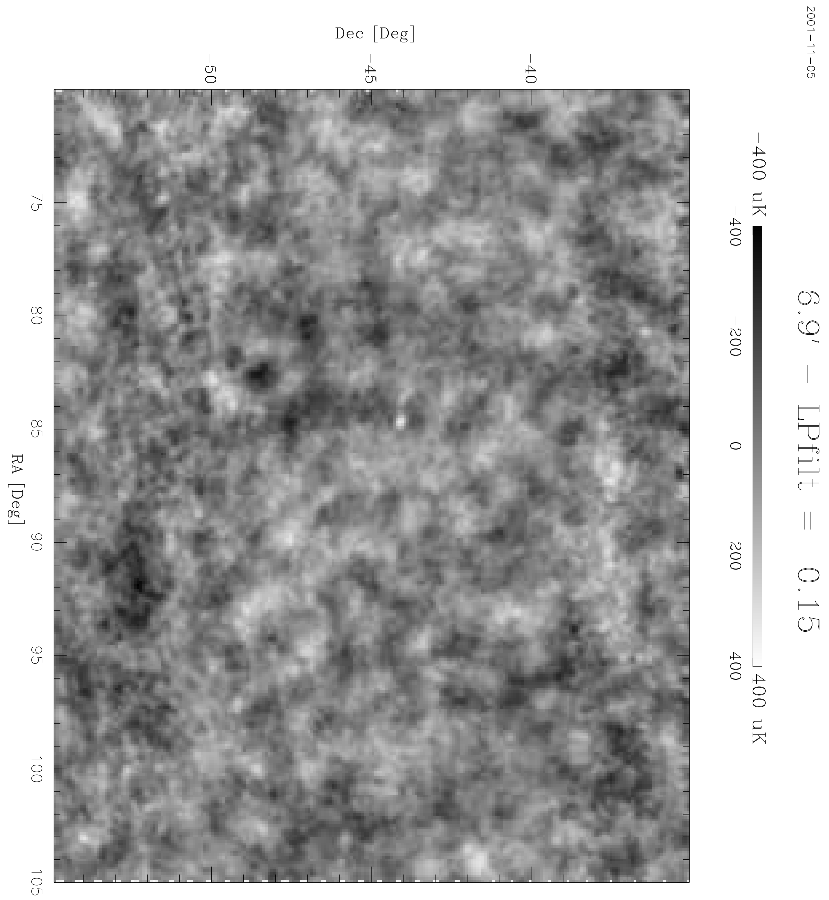

Figure 6.6: The map of CMB anisotropies obtained by the balloon experiment Boomerang in 2001 (see S. Masi et al., Prog.Part.Nucl.Phys. 48 (2002) 243-261). Unlike COBE, Boomerang only analyzed a small patch of the sky, but with a much better angular resolution of a few arc–minutes. The dark (resp. light) spots correspond to colder (resp. warmer) regions.

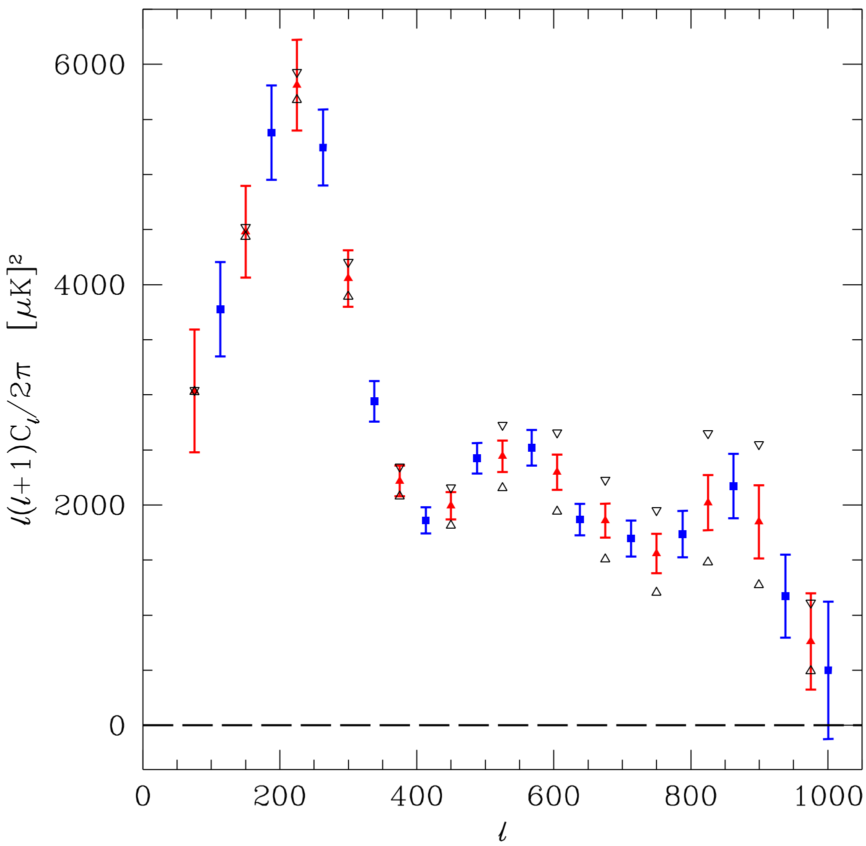

Figure 6.7: The power spectrum of the Boomerang map reveals the structure of the first three acoustic oscillations (see C. B. Netterfield et al., Astrophys.J. 571 (2002) 604-614). These data points account for the temperature power spectrum $\langle|\delta T/T|^2\rangle$ in $(\mu K)^2$, as a function of a number l which is equivalent to $1/\theta$ in some units.

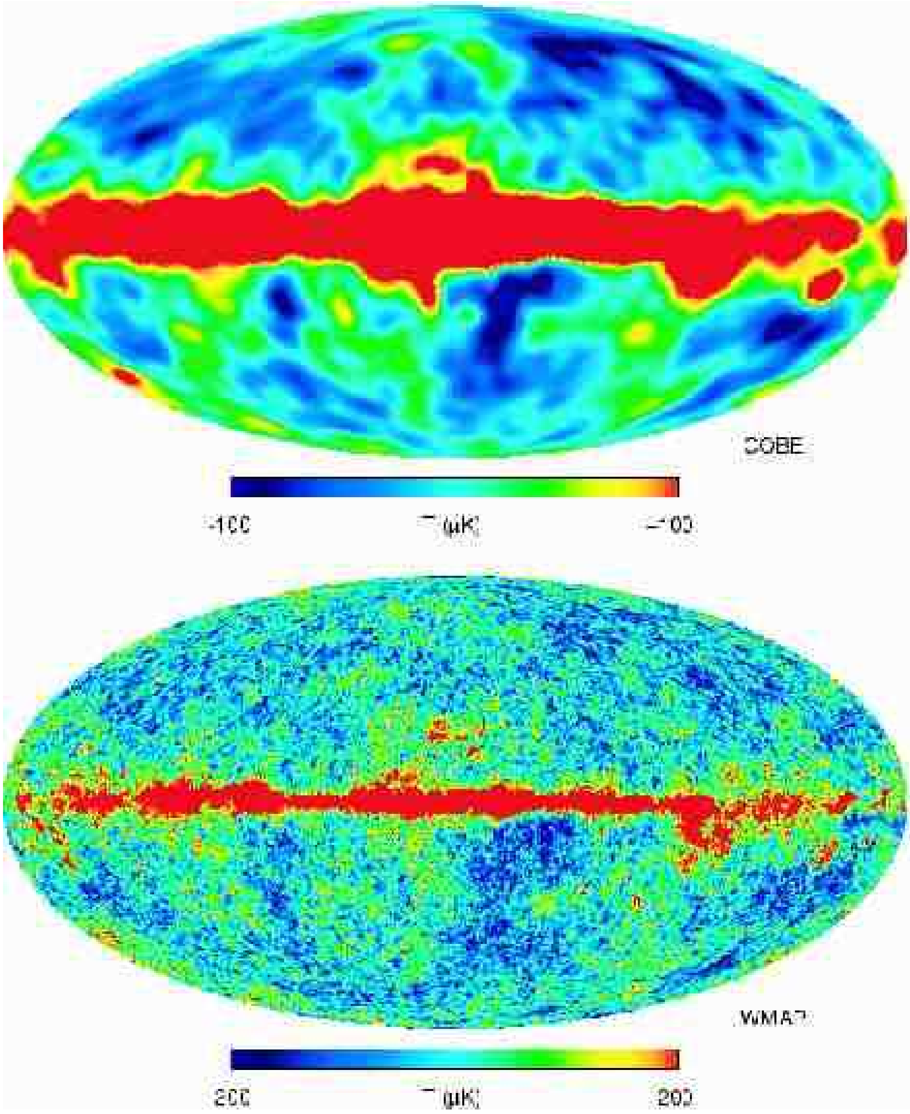

Figure 6.8: *(Bottom)* The full-sky map of CMB anisotropies obtained by the satellite WMAP in 2003 (see C. L. Bennett et al., Astrophys.J.Suppl. 148 (2003) 1). A higher-resolution image is available at `http://lambda.gsfc.nasa.gov/product/map/`. The blue (resp. red) spots correspond to colder (resp. warmer) regions. The central red stripe in the middle is the foreground contamination from the Milky Way. *(Top)* The COBE map, shown again for comparison. The resolution of WMAP is 30 times better than that of COBE, but one can easily see that on large angular scales the two experiments reveal the same structure.

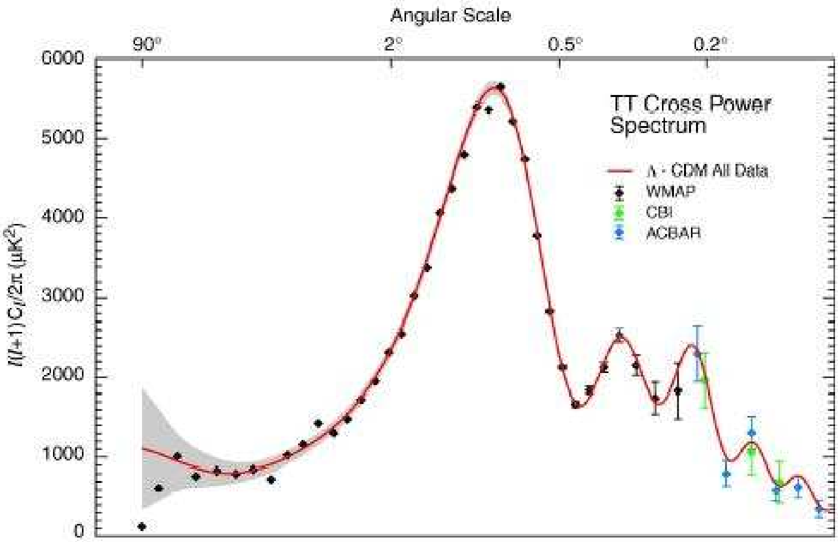

Figure 6.9: Power spectrum indicated by WMAP, perfectly fitted by the theoretical prediction of a ΛCDM model. The black dots show the WMAP measurements (see C. L. Bennett et al., Astrophys.J.Suppl. 148 (2003) 1). The error bars are so small that they are difficult to distinguish from the dots. The blue and green dots show some complementary measurements from other experiments dedicated to smaller angular scales. The red curve is one of the best theoretical fits.

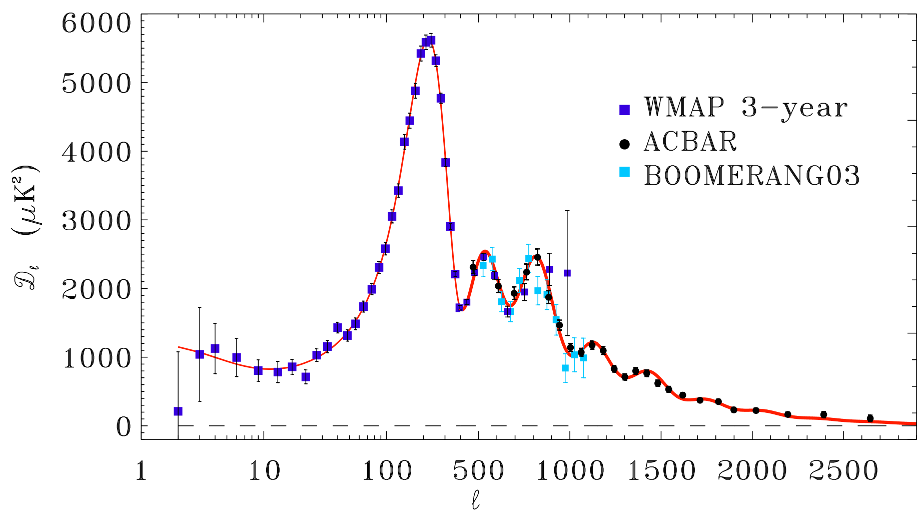

Figure 6.10: Recent results on the temperature spectrum from WMAP (large angular scales), Boomerang (intermediate angular scales) and ACBAR (small angular scales), perfectly fitted by the theoretical predictions of a ΛCDM model. The power spectrum is plotted as a function of a number l which is equivalent to $1/\theta$ in some units. *Taken from arXiv:0801.1491 [astro-ph] by C. L. Reichardt et al.*

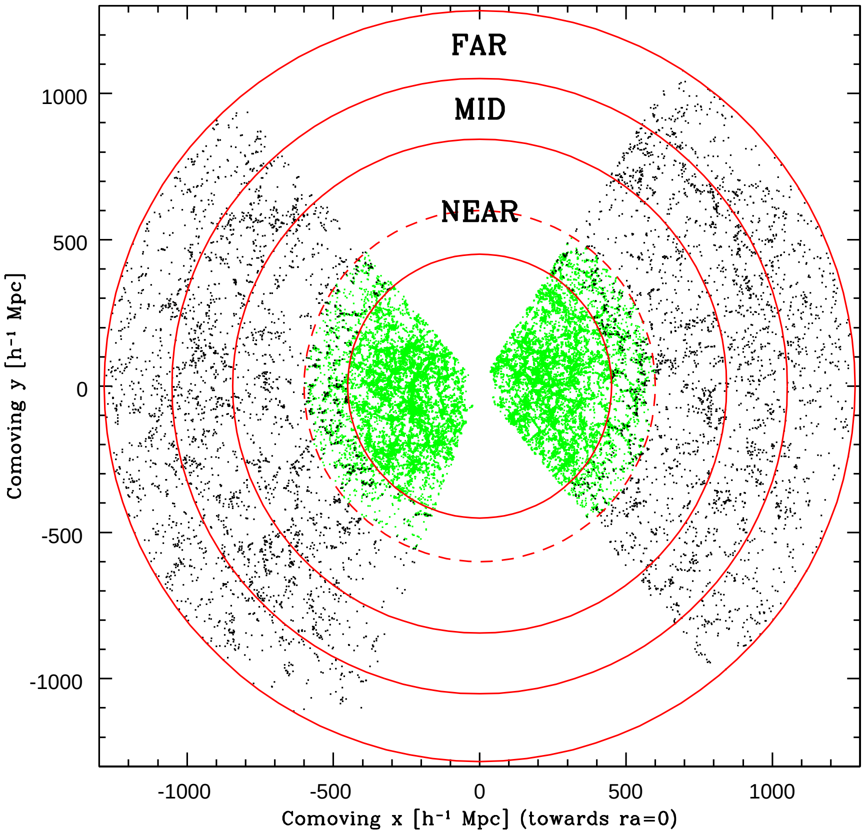

Figure 6.11: The distribution of galaxies in a thin slice of the neighboring universe centered on us, obtained by the Sloan Digital Sky Survey (SDSS). The radial coordinate is the comoving distance in units of h^{-1}Mpc. The four solid red circles correspond to the redshifts $z = 0.155, 0.3, 0.38, 0.474$. Each point represents a galaxy belonging to one of two different samples: the sample called “main galaxies” by the SDSS group (green points) and that called “Luminous Red Galaxies” (LRG, black dots), which extends further since it represents a selection of very bright galaxies only. *Taken from Phys.Rev.D74:123507,2006 [astro-ph/0608632] by M. Tegmark et al.*

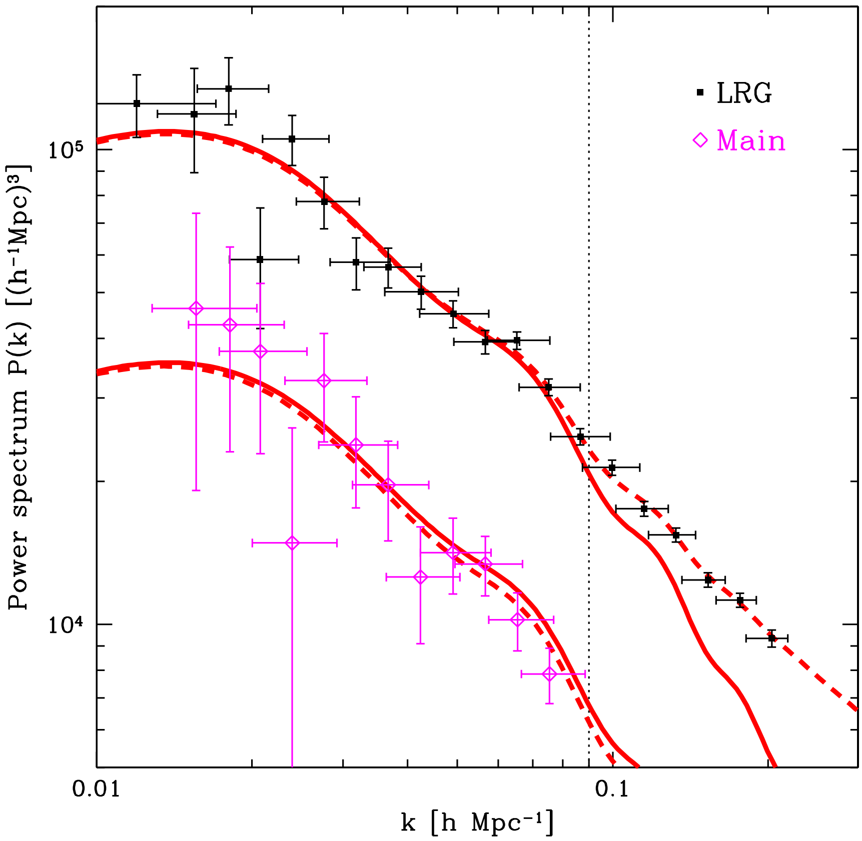

Figure 6.12: Measured power spectra for the galaxies of figure 6.11. The upper points are for luminous red galaxies, the lower one for main galaxies. The two samples don't have necessarily the same light-to-mass bias: this is why the data points indicate two different normalisations of the luminous galactic matter power spectrum $P_{\rm lgm}(k)$. The solid curves correspond to the theoretical prediction for the ΛCDM model best-fitting WMAP3 data, normalized to a light-to-mass bias $b = 1.9$ (top) and $b = 1.1$ (bottom) relative to the $z = 0$ matter power $P(k)$. The dashed curves show an estimate of the nonlinear corrections on small scales, but this aspect is beyond the scope of this course. Note however that the onset of nonlinear corrections is clearly visible for $k \geq 0.09h$/Mpc (vertical line). *Taken from Phys.Rev.D74:123507,2006 [astro-ph/0608632] by M. Tegmark et al.*

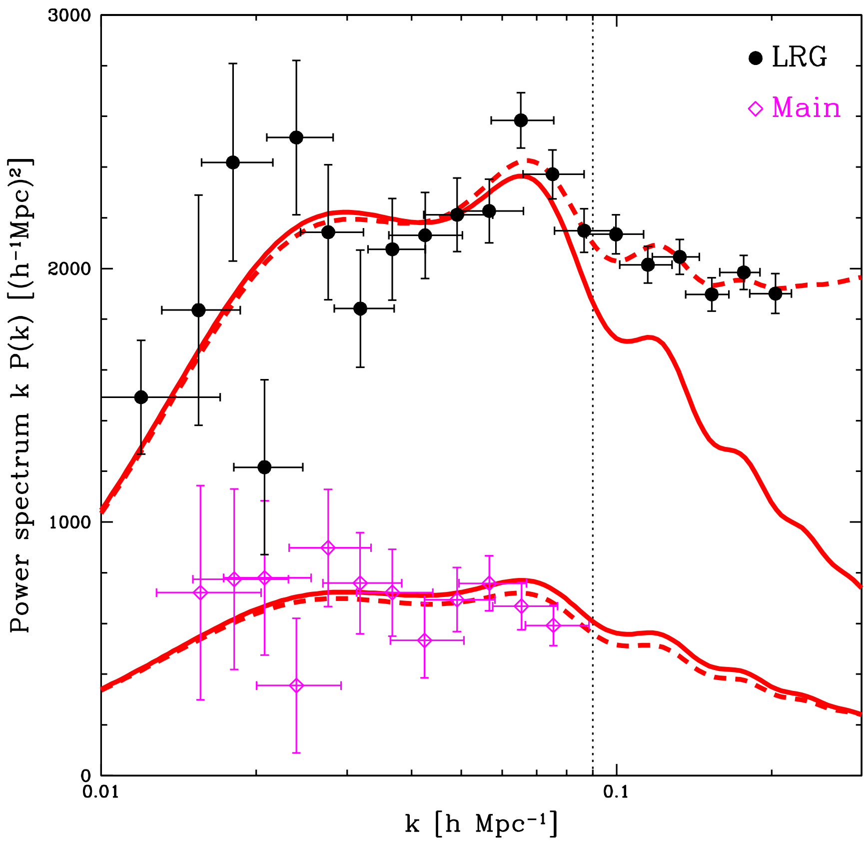

Figure 6.13: Same as 6.12, but multiplied by k and plotted with a linear vertical axis to more clearly illustrate the observation of at least the first baryon acoustic oscillation. *Taken from Phys.Rev.D74:123507,2006 [astro-ph/0608632] by M. Tegmark et al.*

Chapter 7

Inflation

7.1 Motivations for inflation

7.1.1 Flatness problem

Today, Ω_k is measured to be at most of order 10^{-2}, possibly much smaller, while $\Omega_r \equiv \rho_r/\rho_{\rm crit} \simeq \rho_r/(\rho_\Lambda + \rho_m)$ is of order 10^{-4}. Since $\rho_k^{\rm eff}$ scales like a^{-2}, while radiation scales like a^{-4}, the hierarchy between ρ_r and $\rho_k^{\rm eff}$ increases as we go back in time. If t_i is some initial time, t_0 is the time today, and we assume for simplicity that the ratio $\rho_k^{\rm eff}/\rho_r$ is at most equal to one today, we obtain

$$\frac{\rho_k^{\rm eff}(t_i)}{\rho_r(t_i)} \leq \left(\frac{a(t_i)}{a(t_0)}\right)^2 = \left(\frac{\rho_r(t_0)}{\rho_r(t_i)}\right)^{1/2} . \tag{7.1}$$

Today, the radiation energy density $\rho_r(t_0)$ is of the order of $(10^{-4}\text{eV})^4$. If the early universe reached the order of the Planck density $(10^{18}\text{GeV})^4$ at the Planck time t_P, then at that time the ratio was

$$\frac{\rho_k^{\rm eff}(t_P)}{\rho_r(t_P)} = \frac{(10^{-4}\text{eV})^2}{(10^{18}\text{GeV})^2} \sim 10^{-62} . \tag{7.2}$$

Even if the universe never reached such an energy, the hierarchy was already huge when ρ_r was of order, for instance, of $(1\,\text{TeV})^4$.

If we try to build a mechanism for the birth of the classical universe (when it emerges from a quantum gravity phase), we will be confronted to the problem of predicting an initial order of magnitude for the various terms in the Friedmann equation: matter, spatial curvature and expansion rate. The Friedmann equation gives a relation between the three, but the question of the relative amplitude of the spatial curvature with respect to the total matter energy density, i.e. of the hierarchy between $\rho_k^{\rm eff}$ and ρ_r, is an open question. We could argue that the most natural assumption is to start from contributions sharing the same order of magnitude; this is actually what one would expect from random initial conditions at the end of a quantum gravity stage. The flatness problem can therefore be formulated as: why should we start from initial conditions in the very early universe such that $\rho_k^{\rm eff}$ should be fine-tuned to a fraction 10^{-62} of the total energy density in the universe?

The whole problem comes from the fact that the ratio $\rho_k^{\rm eff}/\rho_r$ (or more generally $\Omega_k \equiv \rho_k^{\rm eff}/\rho_{\rm crit}$) increases with time: i.e., a flat universe is an unstable solution of the Friedmann equation. Is this a fatality, or can we choose a framework in which the flat universe would become an attractor solution? The answer to this question is yes, even in the context of ordinary general relativity. We noticed earlier that $|\Omega_k|$ is proportional to $(aH)^{-2}$, i.e. to $\dot{a}^{-2}$. So, as long as the expansion is decelerated, $\dot{a}$ decreases and $|\Omega_k|$ increases. If instead the expansion is accelerated, $\dot{a}$ increases and $|\Omega_k|$ decreases: the curvature is diluted and the universe becomes asymptotically flat.

Inflation is precisely defined as an initial stage during which the expansion is accelerated. One of the motivations for inflation is simply that if this stage is long enough, $|\Omega_k|$ will be driven extremely close to zero, in such way that the evolution between the end of inflation and today does not allow to reach again $|\Omega_k| \sim 1$.

We can search for the *minimal quantity of inflation* needed for solving the flatness problem. For addressing this issue, we should study a cosmological scenario where inflation takes place between times t_i and t_f such that $|\Omega_k| \sim 1$ at t_i, and $|\Omega_k| \sim 1$ again today at t_0. Let us compute the duration of inflation in this model. This will give us an absolute *lower bound* on the needed amount of inflation in the general case. Indeed, we could assume $|\Omega_k| \gg 1$ at t_i (since there could be a long stage of decelerated expansion before inflation); this would just require more inflation. Similarly, we could assume $|\Omega_k| \ll 1$ today at t_0, requiring again more inflation.

So, we assume that between t_i and t_f the scale factor grows from a_i to a_f, and for simplicity we will assume that the expansion is exactly De Sitter (i.e., exponential) with a constant Hubble rate H_i, so that the total density $\rho_{\rm inf}$ is constant between t_i and t_f. We assume that at the end of inflation all the energy $\rho_{\rm inf}$ is converted into a radiation energy ρ_r, which decreases like a^{-4} between t_f and t_0. Finally, we assume that $\rho_k^{\rm eff}$ (which scales like a^{-2}) is equal to $\rho_{\rm inf}$ at t_i and to ρ_r at t_0. With such assumptions, we can write

$$\frac{\rho_k^{\rm eff}(a_0)}{\rho_k^{\rm eff}(a_i)} = \left(\frac{a_i}{a_0}\right)^2 = \frac{\rho_r(a_0)}{\rho_{\rm inf}(a_i)} = \frac{\rho_r(a_0)}{\rho_{\rm inf}(a_f)} = \frac{\rho_r(a_0)}{\rho_r(a_f)} = \left(\frac{a_f}{a_0}\right)^4 \qquad (7.3)$$

and we finally obtain the relation

$$\frac{a_f}{a_i} = \frac{a_0}{a_f} \ . \qquad (7.4)$$

So, the condition for the minimal duration of inflation reads

$$\frac{a_f}{a_i} \geq \frac{a_0}{a_f} \ , \qquad (7.5)$$

which can be summarized in one sentence: there should be as much expansion during inflation as after inflation. A convenient measure of expansion is the so-called *e-fold number* defined as

$$N \equiv \ln a \ . \qquad (7.6)$$

The scale factor is physically meaningful up to a normalization constant, so the e-fold number is defined modulo a choice of origin. The amount of expansion between two times t_1 and t_2 is specified by the number of e-folds $\Delta N = N_2 - N_1 = \ln(a_2/a_1)$. So, the condition on the absolute minimal duration of inflation reads

$$(N_f - N_i) \geq (N_0 - N_f) \qquad (7.7)$$

i.e., the number of inflationary e-folds should be greater or equal to the number of post-inflationary e-folds $\Delta N \equiv N_0 - N_f$. There is no upper bound on $(N_f - N_i)$: for solving the flatness problem, inflation could be arbitrarily long.

It is easy to compute ΔN as a function of the energy density at the end of inflation, $\rho_r(a_f)$. We know that today $\rho_r(a_0)$ is of the order of $(10^{-4}\text{eV})^4$, and we will see in section 7.3.2 that the inflationary energy scale is at most of the order of $(10^{16}\text{GeV})^4$, otherwise current observations of CMB anisotropies would have detected primordial gravitational waves. This gives

$$\Delta N = \ln \frac{a_0}{a_f} = \ln \left(\frac{\rho_r(a_f)}{\rho_r(a_0)} \right)^{1/4} \leq \ln 10^{29} \sim 67 \; . \tag{7.8}$$

We conclude that if inflation takes place around the 10^{16}GeV scale, it should last for a minimum of 67 e-folds. If it takes place at lower energy, the condition is weaker. The lowest scale for inflation considered in the literature (in order not to disturb too much the predictions of the standard inflationary scenario) is of the order of 1 TeV. In this extreme case, the number of post-inflationary e-folds would be reduced to

$$\Delta N \sim \ln 10^{16} \sim 37 \tag{7.9}$$

and the flatness problem can be solved with only 37 e-folds of inflation.

7.1.2 Horizon problem

We recall that the causal horizon $d_H(t_1, t_2)$ is defined as the physical distance at time t_2 covered by a particle emitted at time t_1 and travelling at the speed of light. If the origin of spherical comobile coordinates is chosen to coincide with the point of emission, the physical distance at time t_2 can be computed by integrating over small distance elements dl between the origin and the position r_2 of one particle,

$$d_H(t_1, t_2) = \int_0^{r_2} dl = \int_0^{r_2} a(t_2) \frac{dr}{\sqrt{1 - k\, r^2}} \; . \tag{7.10}$$

In addition, the geodesic equation for ultra-relativistic particles gives $ds = 0$, i.e., $dt = a(t) dr / \sqrt{1 - k\, r^2}$, which can be integrated along the trajectory of the particles,

$$\int_{t_1}^{t_2} \frac{dt}{a(t)} = \int_0^{r_2} \frac{dr}{\sqrt{1 - k\, r^2}} \; . \tag{7.11}$$

We can now replace in the expression of d_H and get

$$d_H(t_1, t_2) = a(t_2) \int_{t_1}^{t_2} \frac{dt}{a(t)} \; . \tag{7.12}$$

Usually, the result is presented in this form. However, for the following discussion, it is particularly useful to eliminate the time from the integral by noticing that $dt = da/(aH)$,

$$d_H(a_1, a_2) = a_2 \int_{a_1}^{a_2} \frac{da}{a^2 H(a)} \; , \tag{7.13}$$

where the Hubble parameter is seen now as a function of a. Let us assume that t_1 and t_2 are two times during Radiation Domination (RD). We know from the Friedmann equation that during RD on has $H \propto a^{-2}$, so we can parametrize the Hubble rate as $H(a) = H_2 \; (a_2/a)^2$. We obtain

$$d_H(a_1, a_2) = a_2 \int_{a_1}^{a_2} \frac{da}{a_2^2 H_2} = \frac{1}{H_2} \frac{(a_2 - a_1)}{a_2} \; . \tag{7.14}$$

If the time t_2 is much after t_1 so that $a_2 \gg a_1$, the expression for the horizon does not depend on a_1,

$$d_H(a_1, a_2) \simeq \frac{1}{H_2} \,. \tag{7.15}$$

So, during RD, the horizon equals the Hubble radius at time t_2 (in agreement with the result of Eq. (5.45) with $n = 1/2$). During matter domination, the horizon is still close to the Hubble radius, modulo a factor of order one.

The horizon represents the causal distance in the universe. Suppose that a physical mechanism is turned on at time t_1. Since no information can travel faster than light, the physical mechanism cannot affect distances larger than $d_H(t_1, t_2)$ at time t_2. So, the horizon provides the *coherence scale* of a given mechanism. For instance, if a phase transition creates bubbles or patches containing a given vacuum phase, the scale of homogeneity (i.e., the maximum size of the bubble, or the scale on which a patch is nearly homogeneous) is given by $d_H(t_1, t_2)$ where t_1 is the time at the beginning of the transition.

Before photon decoupling, the Planck temperature of photons at a given point depends on their local density. A priori, we can expect that the universe will emerge from a quantum gravity stage with random values of the local density. The coherence length, or characteristic scale on which the density is nearly homogeneous, is given by $d_H(t_1, t_2)$. We have seen that if t_1 and t_2 are two times during radiation domination, this quantity cannot exceed $R_H(t_2)$, even in the most favorable limit in which t_1 is chosen to be infinitely close to the initial singularity. We conclude that at time t_2, the photon temperature should not be homogeneous on scales larger than $R_H(t_2)$.

CMB experiments map the photon temperature on our last-scattering-surface at the time of photon decoupling. So, we expect CMB maps to be nearly homogeneous on a characteristic scale $R_H(t_{\rm dec})$. This scale is very easy to compute: knowing that $H(t_0)$ is of the order of $(h/3000)\,{\rm Mpc}^{-1}$ with $h \simeq 0.7$, we can extrapolate $H(t)$ back to the time of equality, and find that the distance $R_H(t_{\rm dec})$ subtends an angle of order of a few degrees in the sky - instead of encompassing the diameter of the last scattering surface. So, it seems that the last scattering surface is composed of several thousands causally disconnected patches. However, the CMB temperature anisotropies are only of the order of 10^{-5}: in other words, the full last scattering surface is extremely homogeneous. This appears as completely paradoxical in the framework of the Hot Big Bang scenario.

What is the origin of this problem? When we computed the horizon, we integrated $(a^2 H)^{-1}$ over da and found that the integral was converging with respect to the boundary a_1: so, even by choosing the initial time to be infinitely early, the horizon is bounded by a function of a_2. If the integral was instead divergent, we could obtain an infinitely large horizon at time t_2 simply by choosing a_1 to be small enough. The convergence of the integral

$$\int_{a_1}^{a_2} \frac{da}{a^2 H(a)} = \int_{a_1}^{a_2} \frac{da}{a\dot{a}} \tag{7.16}$$

with respect to $a_1 \to 0$ depends precisely on the fact that the expansion is accelerated or decelerated. For linear expansion, the integrand is $1/a$, the limiting case between convergence and divergence. If it is decelerated, $\dot{a}$ decreases and the integral converges. If it is accelerated, $\dot{a}$ increases and the integral diverges in the limit $a_1 \to 0$.

So, if the radiation dominated phase is preceded by an infinite stage of accelerated expansion, one can reach an arbitrarily large value for the horizon at the time of decoupling. In fact, in order to explain the homogeneity of the last scattering surface, we only need to boost the horizon by a factor of $\sim 10^3$ with respect to the Hubble radius at that time. This can be fulfilled with a rather small amount of accelerated expansion.

Let us take an exemple and assume that between a_i and a_f, the acceleration is exponential, $a = e^{\alpha t}$. In this case, the Hubble parameter $\dot{a}/a$ is constant over this period: let's call it $H_{\rm inf}$. The horizon computed between a_i and a_f reads:

$$d_H(a_i, a_f) = a_f \int_{a_i}^{a_f} \frac{da}{a^2 H_{\rm inf}} = \frac{1}{H_{\rm inf}} \left(\frac{a_f}{a_i} - 1 \right) \simeq \frac{1}{H_{\rm inf}} \; \frac{a_f}{a_i} \; . \qquad (7.17)$$

So, at the end of inflation, the horizon is larger than the Hubble radius $R_H = 1/H_{\rm inf}$ by a factor a_f/a_i, i.e, by the exponential of the number of inflationary e-folds. After, the horizon will keep growing in the usual way,

$$\begin{aligned} d_H(a_i, a_2) &= a_2 \int_{a_i}^{a_2} \frac{da}{a^2 H(a)} \\ &= \frac{a_2}{H_{\rm inf}} \left(\frac{1}{a_i} - \frac{1}{a_f} \right) + a_2 \int_{a_f}^{a_2} \frac{da}{a^2 H(a)} \\ &\simeq \frac{1}{H_{\rm inf}} \; \frac{a_2}{a_i} + \frac{1}{H_2} \; , \end{aligned} \qquad (7.18)$$

and remains much larger than the Hubble radius $\frac{1}{H_2}$.

The condition for solving the horizon problem can be shown to be exactly the same as for solving the flatness problem: the number of inflationary e-fold should be at least equal to that of post-inflationary e-folds. If it is larger, then the size of the observable universe is even smaller with respect to the causal horizon.

7.1.3 Origin of perturbations

Since our universe is inhomogeneous, one should find a physical mechanism explaining the origin of cosmological perturbations. Inhomogeneities can be expanded in comoving Fourier space. Their physical wavelength

$$\lambda(t) = \frac{2\pi a(t)}{k} \qquad (7.19)$$

is stretched with the expansion of the universe. During radiation domination, $a(t) \propto t^{1/2}$ and $R_H(t) \propto t$. So, the Hubble radius grows with time faster than the perturbation wavelengths. We conclude that observable perturbations were originally super-Hubble fluctuations (i.e., $\lambda > R_H \Leftrightarrow k < 2\pi a H$). Actually, the discussion of the horizon problem already showed that at decoupling the largest observable fluctuations are super-Hubble fluctuations. Even if we take a smaller scale, e.g. the typical size of a galaxy cluster $\lambda(t_0) \sim 1$ Mpc, we find that the corresponding fluctuations were clearly super-Hubble fluctuations for instance at the time of nucleosynthesis. We have seen that in the Hot Big Bang scenario (without inflation) the Hubble radius $R_H(t_2)$ gives an upper bound on the causal horizon $d_H(t_1, t_2)$ for whatever value of t_1. So, super-Hubble fluctuations are expected to be out of causal contact. The problem is that it is impossible to find a mechanism for generating coherent fluctuations on acausal scales. There are two possible solutions to this issue:

- we can remain in the framework of the Hot Big Bang scenario and assume that perturbations are produced causally when a given wavelength enters into the horizon. In this case, there should be not coherent fluctuations on super-Hubble scales, i.e. the power spectrum of any kind of perturbation should fall like white noise in the limit $k \ll aH$. This possibility is now ruled out for at least two reasons. First, the observation of CMB anisotropies on angular scales greater than one degree (i.e., super-Hubble scales at that time) is consistent with coherent fluctuations rather than white noise. Second, the observations of acoustic peaks in the power spectrum of CMB anisotropies is a clear proof that cosmological perturbations are generated much before Hubble crossing, in such way that all modes with a given wavelength entering inside the Hubble radius before photon decoupling experience coherent acoustic oscillations (i.e. oscillate with the same phase).

- we can modify the cosmological scenario in such way that all cosmological perturbations observable today were inside the causal horizon when they were generated at some early time (we will study a concrete generation mechanism in section 7.3).

So, our goal is to find a paradigm such that the largest wavelength observable today, which is $\lambda_{\max}(t_0) \sim R_H(t_0)$ (see section 5.2.2), was already inside the causal horizon at some early time t_i. If before t_i the universe was in decelerated expansion, then the causal horizon at that time was of order $R_H(t_i)$. How can we have $\lambda_{\max} \leq R_H$ at t_i and $\lambda_{\max} \sim R_H$ today? If between t_i and t_0 the universe is dominated by radiation or matter, it is impossible since the Hubble radius grows faster than the physical wavelengths. However, in general,

$$\frac{\lambda(t)}{R_H(t)} = \frac{2\pi a(t)}{k} \frac{\dot{a}(t)}{a(t)} = \frac{2\pi \dot{a}(t)}{k} \,, \tag{7.20}$$

so that during accelerated expansion the physical wavelengths grow faster than the Hubble radius. So, if between some time t_i and t_f the universe experiences some inflationary stage, it is possible to have $\lambda_{\max} < R_H$ at t_i: the scale $\lambda_{\max}$ can then exit the Hubble radius during inflation and re-enter approximately today (see Figure 7.1).

It is easy to show that once again, the minimal number of inflationary e-folds requested for solving this problem should be at least equal to that of post-inflationary e-folds.

One could argue that the argument on the origin of fluctuations is equivalent to that of the horizon problem, reformulated in a different way. Anyway, for understanding inflation it is good to be aware of the two arguments, even if they are not really independent from each other.

7.1.4 Monopoles

We will not enter here into the details of the monopole problem. Just in a few words, some phase transitions in the early universe are expected to create "dangerous relics" like magnetic monopoles, with a very large density which would dominate the total density of the universe. These relics are typically non-relativistic, with an energy density decaying like a^{-3}: so, they are not diluted, and the domination of radiation and ordinary matter can never take place.

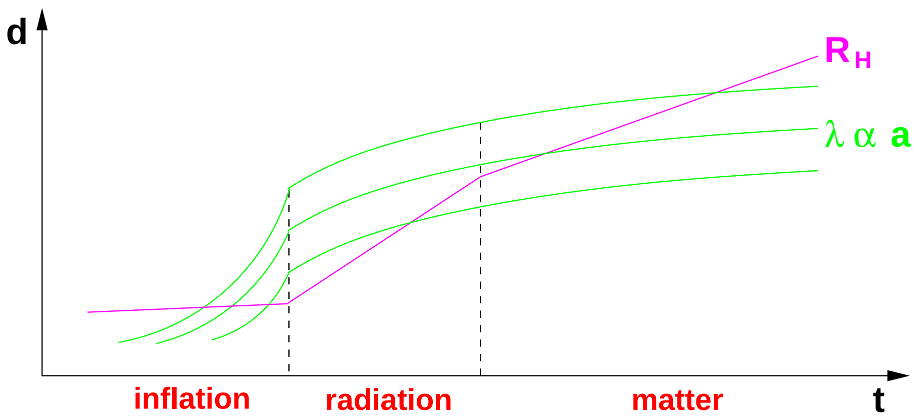

Figure 7.1: Comparison of the Hubble radius with the physical wavelength of a few cosmological perturbations. During the initial stage of accelerated expansion (called inflation), the Hubble radius grows more slowly than each wavelength. So, cosmological perturbations originate from inside R_H. Then, the wavelength of each mode grows larger than the Hubble radius during inflation and re–enters during radiation or matter domination.

Inflation can solve the problem provided that it takes place after the creation of dangerous relics. During inflation, monopoles and other relics will decay like a^{-3} (a^{-4} in the case of relativistic relics) while the leading vacuum energy is nearly constant: so, the energy density of the relics is considerably diluted, typically by a factor $(a_f/a_i)^3$, and today they are irrelevant. The condition on the needed amount of inflation is much weaker than the condition obtained for solving the flatness problem, since dangerous relics decay faster than the effective curvature density ($\rho_k^{\rm eff} \propto a^{-2}$).

7.2 Slow-roll scalar field inflation

So, the first three problems of section 7.1 can be solved under the assumption of a long enough stage of accelerated expansion in the early universe. How can this be implemented in practice?

First, by combining the Friedman equation (2.47) in a flat universe with the conservation equation (2.48), it is easy to find that

$$\ddot{a} > 0 \qquad \Rightarrow \qquad \rho + 3p < 0. \tag{7.21}$$

What type of matter corresponds to such an unusual relation between density and pressure? A positive cosmological constant can do the job:

$$p_\Lambda = -\rho_\Lambda \qquad \Rightarrow \qquad \rho_\Lambda + 3p_\Lambda = -2\rho_\Lambda < 0. \tag{7.22}$$

But since a cosmological constant is... constant, it cannot be responsible for an initial stage of inflation: otherwise this stage would go on forever, and there would be no transition to radiation domination.

Let us consider instead the case of a scalar field (i.e., a field of spin zero, represented by a simple function of time and space, and invariant under Lorentz transformations). The general action for a scalar field in curved space-time

$$S = -\int d^4x \sqrt{|g|}\, (\mathcal{L}_g + \mathcal{L}_\varphi) \tag{7.23}$$

involves the Lagrangian of gravitation

$$\mathcal{L}_g = \frac{R}{16\pi\mathcal{G}} \tag{7.24}$$

and that of the scalar field

$$\mathcal{L}_\varphi = \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - V(\varphi) = \frac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - V(\varphi) \tag{7.25}$$

where $V(\phi)$ is the scalar potential. The variation of the action with respect to $g_{\mu\nu}$ enables to define the energy-momentum tensor

$$T_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi - \mathcal{L}_\varphi g_{\mu\nu} \tag{7.26}$$

and the Einstein tensor $G_{\mu\nu}$, which are related through the Einstein equations

$$G_{\mu\nu} = 8\pi\mathcal{G}\,T_{\mu\nu}\ . \tag{7.27}$$

Instead, the variation of the action with respect to φ gives Klein-Gordon equation

$$\frac{1}{\sqrt{|g|}}\partial_\mu\left[\sqrt{|g|}\partial^\mu\varphi\right] + \frac{\partial V}{\partial\varphi} = 0\ . \tag{7.28}$$

The same equation could have been obtained using a particular combination of the components of $T_{\mu\nu}$ and their derivatives, which vanish by virtue of the Bianchi identities (in other word, the Klein-Gordon equation is contained in the Einstein equations).

Let us now assume that the homogeneous Friedmann universe with flat metric

$$g_{\mu\nu} = \mathrm{diag}\left(1, -a(t)^2, -a(t)^2, -a(t)^2\right) \tag{7.29}$$

is filled by a homogeneous classical scalar field $\bar{\varphi}(t)$. One can show that the corresponding energy-momentum tensor is diagonal, $T_\mu^\nu = \mathrm{diag}(\rho, -p, -p, -p)$, with

$$\rho = \frac{1}{2}\dot{\bar{\varphi}}^2 + V(\varphi)\ , \tag{7.30}$$

$$p = \frac{1}{2}\dot{\bar{\varphi}}^2 - V(\varphi)\ . \tag{7.31}$$

The Friedmann equation reads

$$G_0^0 = 3H^2 = 8\pi\mathcal{G}\,\rho \tag{7.32}$$

and the Klein-Gordon equation

$$\ddot{\bar{\varphi}} + 3H\dot{\bar{\varphi}} + \frac{\partial V}{\partial\varphi}(\bar{\varphi}) = 0\ . \tag{7.33}$$

These two independent equations specify completely the evolution of the system. However it is worth mentioning that the full Einstein equations provide another relation

$$G_i^i = \left(2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right) = -8\pi\mathcal{G}\,p\ . \tag{7.34}$$

The combination $\dot{G}_0^0 + 3H(\dot{G}_0^0 - G_i^i)$ vanishes (it is one of the Bianchi identities), and gives a conservation equation $\dot{\rho} + 3H(\rho + p) = 0$, which is nothing but the Klein-Gordon equation. Finally, the combination $G^i_i - G^0_0$ provides a very useful relation

$$\dot{H} = -4\pi\mathcal{G}\,\dot{\bar{\varphi}}^2 \tag{7.35}$$

which is consistent with the fact that the Hubble parameter can only decrease.

The condition $p < -\rho/3$ reads $\dot{\bar{\varphi}}^2 < V$: when the potential energy dominates over the kinetic energy, the universe expansion is accelerated. In the limit of zero kinetic energy, the energy-momentum tensor would be that of a cosmological constant, and the expansion would be exponential (this is called "De Sitter expansion") and everlasting. For a long, finite stage of acceleration we must require that the *first slow-roll condition*

$$\frac{1}{2}\dot{\bar{\varphi}}^2 \ll V(\bar{\varphi}) \tag{7.36}$$

holds over an extended period. Since the evolution of the scalar field is given by a second-order equation, the above condition could apply instantaneously but not for an extended stage, in particular in the case of oscillatory solutions. If we want the first slow-roll condition to hold over an extended period, we must impose that the time-derivative of this condition also holds (in absolute value). This gives the *second slow-roll condition*

$$|\ddot{\bar{\varphi}}| \ll \left|\frac{\partial V}{\partial \varphi}(\bar{\varphi})\right| \tag{7.37}$$

which can be rewritten, by virtue of the Klein-Gordon equation, as

$$|\ddot{\bar{\varphi}}| \ll 3H\,|\dot{\bar{\varphi}}| \ . \tag{7.38}$$

When these two conditions hold, the Friedmann and Klein-Gordon equations become

$$\begin{aligned} 3H^2 &\simeq 8\pi\mathcal{G}\,V(\bar{\varphi}) \ , & (7.39)\\ \dot{\bar{\varphi}} &\simeq -\frac{1}{3H}\frac{\partial V}{\partial \varphi}(\bar{\varphi}) \ . & (7.40) \end{aligned}$$

The two slow-roll conditions can be rewritten as conditions either on the slowness of the variation of $H(t)$, or on the flatness of the potential $V(\varphi)$.

So, a particular way to obtain a stage of accelerated expansion in the early universe is to introduce a scalar field, with a flat enough potential. Scalar field inflation has been proposed in 1979 by Guth. Starting from 1979 and during the 80's, most important aspects of inflation were studied in details by Starobinsky, Guth, Hawking, Linde, Mukhanov and other people. Finally, during the 90's, many ideas and models were proposed in order to make contact between inflation and particle physics. The purpose of scalar field inflation is not only to provide a stage of accelerated expansion in the early universe, but also, a mechanism for the generation of matter and radiation particles, and another mechanism for the generation of primordial cosmological perturbations. Let us summarize how it works in a very sketchy way.

Slow-roll. First, let us assume that just after the initial singularity, the energy density is dominated by a scalar field, with a potential flat enough for slow–roll. In any small region where the field is approximately homogeneous and slowly–rolling, accelerated expansion takes place: this small region becomes exponentially large, encompassing the totality of the present observable universe. Inside this region, the causal horizon becomes much larger than the Hubble radius, and any initial spatial curvature is driven almost to zero – so, some of the main problems of the standard cosmological model are solved. After some time, when the field approaches the minimum its potential, one of the two slow-roll conditions breaks down, and inflation ends: the expansion becomes decelerated again.

Reheating. At the end of inflation, the kinetic energy of the field is bigger than the potential energy; in general, the field is quickly oscillating around the minimum of the potential. According to the laws of quantum field theory, the oscillating scalar field will decay into fermions and bosons. This could explain the origin of all the particles filling our universe. The particles probably reach quickly a thermal equilibrium: this is why this stage is called "reheating".

Generation of primordial perturbations. Finally, the theory of scalar field inflation also explains the origin of cosmological perturbations – the ones leading to CMB anisotropies and large scale structure formation. Using again quantum field theory in curved space–time, it is possible to compute the amplitude of the small quantum fluctuations of the scalar field φ (as well as the quantum fluctuations of the metric $h_{\mu\nu}$). The physical wavelengths of these fluctuations grow quickly, like in figure 7.1. So, they are initially inside the Hubble radius, where we can apply the laws of quantum mechanics in flat space-time (as long as $k \ll aH$, the modes do not see the curvature of space-time). In the opposite limit, when a wavelength is stretched to scales larger than the Hubble length, it is possible to show that the modes experience a kind of quantum–to–classical transition, in the sense that they become indistinguishable from classical stochastic fluctuations: hence, the primordial fluctuations have a random distribution (as expected), but we don't need to employ the formalism of quantum mechanics (wave functions, etc.) in order to describe their statistics. In addition, the initial quantum fluctuations $\delta\varphi$ are assumed to be vacuum fluctuations (corresponding to the fundamental state of the field $\delta\varphi$). As a consequence, the probability distribution of each mode $\delta\varphi(\mathbf{k})$ after the transition can be showed to be a Gaussian, depending only on k. Hence, at a given time, all information about the statistics of the field is contained in the power spectrum $\langle|\varphi(\mathbf{k})|^2\rangle$, which is a function of k.

7.3 Inflationary perturbations

7.3.1 Scalar perturbations

The perturbations of the scalar field $\delta\varphi$ are coupled with those of the scalar metric fluctuations: for instance, ϕ and ψ in the longitudinal gauge. At first order in perturbation theory, it is easy to show that $\phi = \psi$, so the problem of scalar perturbations during inflation reduces to the evolution of two quantities only, $\delta\varphi$ and ϕ. In addition, the linearized Einstein equations provides a relation between $\delta\varphi$ and ϕ: they are not independent, and their evolution is dictated by a single equation of motion.

As explained above, quantum field theory allows to exactly follow the evolution and the quantum–to–classical transition of the fields $\delta\varphi$ and ϕ during inflation. So, it is possible to compute exactly the power spectrum of $\delta\varphi$ and ϕ for observable modes, i.e. on wavelengths much larger than the Hubble radius at the end of inflation. But how can we related these spectra to the initial conditions at the beginning of radiation domination, after the end of inflation?

We could fear that such a relation could be very difficult to compute, and could depend on the mechanism through which the scalar field decays into radiation and matter... Fortunately, this is not the case: the relation between the spectrum of fluctuations at the end of inflation and at the beginning of radiation domination is trivial. The reason is that when the wavelength of a mode $\phi(\mathbf{k})$ becomes much larger than the Hubble radius, the perturbation freezes out. Hence it is not affected by the decay of the scalar field during reheating. But when the radiation and matter particles are formed during reheating, they are sensitive to the gravitational potential, and more particles accumulate in the potential wells. So, the gravitational potential behaves like a mediator between the scalar field perturbations during inflation and the radiation/matter perturbations in the radiation/matter–dominated universe. If we can compute the power spectrum $\langle|\phi(\mathbf{k})|^2\rangle$ at the end of inflation, we are done, because this power spectrum remains the same at the beginning of radiation domination on super-Hubble scale; then the initial condition described in Eq. (5.75) apply, and the evolution of all radiation/matter perturbations is entirely determined.

It is far beyond the level of these notes to compute the evolution of primordial perturbations during inflation. However, we should stress that it can be studied in a very precise way using quantum field theory. The result for the primordial spectrum of scalar metric perturbations during inflation/radiation domination and on super-Hubble scales $k \ll aH$ reads:

$$\langle|\phi(\mathbf{k})|^2\rangle = 2\left(\frac{8\pi G}{3k}\right)^3 \frac{V^3}{V'^2}\ , \tag{7.41}$$

where V and V', which are both functions of φ, should be evaluated *with the value of the field corresponding to the time of Hubble crossing during inflation for each mode* $\mathbf{k}$, i.e., with the value $\bar{\varphi}(t)$ at the time t when $k = aH$. Hence, the primordial spectrum depends on k not only through the above k^{-3} factor, but also through the V^3/V'^2 factor. However, since the field is in slow-roll, V^3/V'^2 does not vary a lot between the time at which the largest and the smallest observable wavelengths cross the Hubble radius during inflation. Hence, the dependence of V^3/V'^2 on k is small, and the above spectrum is close to a scale-invariant spectrum, $\langle|\phi(\mathbf{k})|^2\rangle \propto k^{-3}$. However, the deviation from exact scale-invariance (i.e. the value of the spectral index n minus one) depends crucially on the evolution of this ratio V^3/V'^2 with time and scale. By taking the derivative of the above equation, one could show that $n-1$ is indeed related to the ratios V''/V and V'/V evaluated when observable scales cross the horizon.

In the previous chapters, we saw that CMB and large scale structure observations allow to reconstruct the cosmological evolution during radiation/matter/Λ domination, as well as the primordial spectrum $\langle|\phi(\mathbf{k})|^2\rangle$. In particular, the amplitude A and spectral index n (defined in Eq. (5.93)) of the primordial spectrum $\langle|\phi(\mathbf{k})|^2\rangle$ can be measured. According to the above results, these observations provide a measurement of V^3/V'^2 and of its evolution with φ within a small interval. Hence the potential $V(\phi)$ can be reconstructed to some extent from observations. It is quite remarkable that current observations provide a way to constrain the physical mechanism governing the evolution of the universe at extremely high energy (considerably higher than during nucleosynthesis) and extremely early times (a tiny fraction of second after the initial singularity).

7.3.2 Tensor perturbations (gravitational waves)

The same mechanism which produces stochastic fluctuations of $\delta\varphi$ and ϕ (more precisely, of the scalar metric perturbations) on cosmological scales produces also stochastic fluctuations of the tensor metric perturbations, i.e., of the tensor h_{ij} defined in Eq. (5.20). These perturbations are called gravitational waves, since inside the Hubble radius they have oscillatory solutions: there are deformation of our space-time manifold, propagating like waves with the velocity of light. Unlike scalar perturbations, they do not couple with matter fields or scalar fields within linear perturbation theory. Like electromagnetic waves, gravitational waves can propagate in the vacuum without being damped.

It is possible to compute the primordial spectrum of gravitational waves (i.e., the primordial spectrum of the components of h_{ij}) using the same formalism as for the scalar metric fluctuation ϕ. The result reads

$$\langle |h(\mathbf{k})|^2 \rangle = \frac{2}{3} (8\pi G)^2 \, k^{-3} V \ , \tag{7.42}$$

where V is evaluated like for scalar perturbations, i.e. with the value $\bar{\varphi}(t)$ at the time t when $k = aH$. Here h stands for the components of h_{ij} (we don't give the exact definition of h here for concision).

Hence, inflation is also expected to fill the universe with a random background of gravitational waves which could be detected today, at least in principle. Unfortunately, this background of gravitational waves is so low that its detection is unlikely with the current generation (and even the next generation) of gravitational wave detectors (VIRGO, LIGO, etc.) However, there is a chance to detect it in the CMB: gravitational waves of primordial origin are expected to contribute to the CMB spectrum on the largest angular scales, as shown in figure 7.2. The shape of the tensor contribution to the CMB spectrum can be computed with the same kind of numerical code as for scalar perturbations. The main uncertainty is not on the shape, but on the amplitude of this contribution. Equation (7.42) shows that the amplitude depends on V during inflation, i.e. on the energy scale of inflation. The condition for the tensor contribution to be roughly of the same order as the scalar one in the large scale CMB spectrum is roughly that $V \sim (10^{16}\text{GeV})^4$ during inflation, i.e. that the energy scale of inflation is of the order of 10^{16}GeV (coincidentally, this turns out to be the order of magnitude of GUT symmetry breaking).

So far, the observation of CMB anisotropies is consistent with a spectrum arising only from scalar perturbations. A large tensor contribution cannot be present, because it would lead to an increase in the ratio between the amplitude of the large-scale plateau region and that of the small-scale peak region, at odds with observation. Hence, CMB temperature maps allow to put an upper limit on the energy scale of inflation: roughly, it has to be smaller than 10^{16}GeV. Future observations of the CMB will be able to test the possible contribution of tensors with better precision: hence, in the next years, cosmologists hope either to push this bound further down, or to detect the background of primordial gravitational waves produced by inflation.

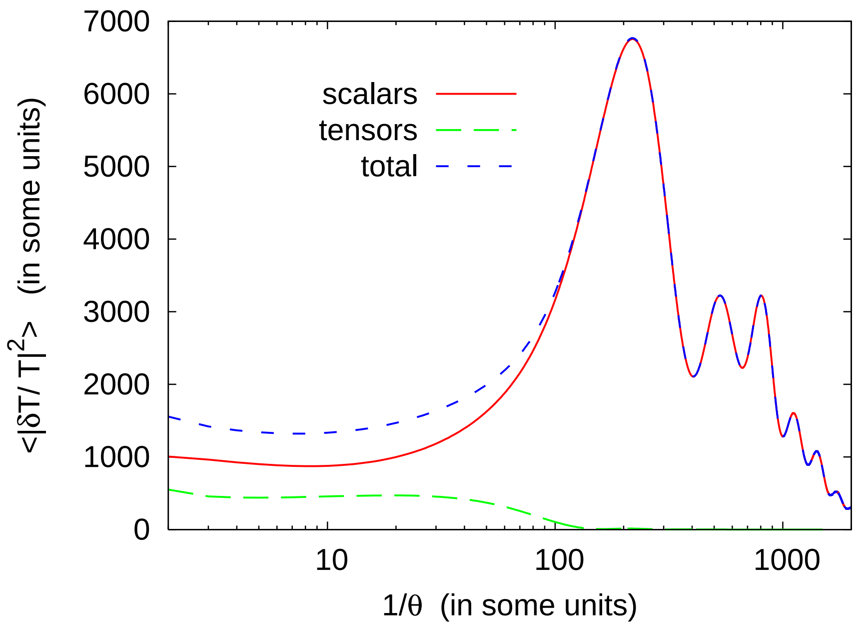

Figure 7.2: The red solid line shows the same reference CMB temperature spectrum as in the previous figures. The dashed green line shows the additional contribution from tensor perturbations for the same cosmological model, assuming a primordial tensor amplitude such that roughly one third of the observed CMB power spectrum on large angular scales would come from tensors. The dashed blue curve shows the total spectrum which would be observed in this model.

7.4 Success of the theory of inflation

Let us summarize the positive outcomes of the theory of inflation:

1. it provides a simple solution to the flatness, horizon and monopole problems.

2. it includes in some unavoidable way a mechanism for producing primordial fluctuations starting from simple initial conditions, i.e. from a perfectly homogeneous scalar field with vacuum fluctuations dictated by quantum mechanics.

3. these perturbations have almost automatically the properties which are necessary in order to explain observations: they are generated very early and on super-Hubble scales; they have a Gaussian statistics and obey to adiabatic initial conditions; they have a nearly scale-invariant primordial spectrum.

4. inflation provides a mechanism for the generation of a thermal bath of particles in the early universe (the so-called reheating phase occurring after or during the scalar field oscillations and decay). Unfortunately, this mechanism is very difficult to probe experimentally: reheating does not have clear observable signatures, unlike the mechanism for the generation of primordial fluctuations.

5. thanks to the theory of inflation, it is possible to provide a self-consitent explanation for the global properties of our universe without making any assumption about quantum gravity (during inflation, one quantizes only metric perturbations, not the metric itself: hence inflation is based on quantum field theory in curved space-time, but NOT on quantum gravity). In fact, in inflationary cosmology, what happens *before* inflation is usually not important: our universe only keeps track of what happened during the last $\sim$60 e-folds of inflation and after inflation.

The third point is the most convincing argument in favor of inflation. Before the first observations of CMB anisotropies, it was impossible to know whether our universe was described by such initial conditions (primordial perturbations on super-Hubble scales, Gaussian, adiabatic and nearly scale-invariant). So, cosmologist were studying various possible scenarios for the generation of perturbations. The main alternative would be to assume that they are generated during a phase transition (e.g. a spontaneous symmetry breaking). In this case, they would appear inside the Hubble radius and would be non-Gaussian, non-adiabatic and far from scale invariance. As we have seen before, the observation of CMB anisotropies has confirmed the four generic predictions of inflation, as far as primordial perturbations are concerned. Alternative theories are discarded (at least as a dominant mechanism for the generation of primordial perturbations) and most people agree that inflation is a very likely scenario. It is a striking example of a predictive and elegant theory (with few assumptions leading to many observable consequences validated by observations).

The negative outcomes of the theory of inflation are the following:

1. inflation is based on a scalar field (usually called the *inflaton*), but we don't know anything about its origin and its relationship with other known fields/particles. However it is possible to assume some connection between inflation and particle physics (the inflaton could be a Higgs field, the size of an extra dimension, etc.) The difficulty is then to find a good reason for which the inflaton potential would be flat enough for fulfilling the slow-roll conditionc. This argument readily excludes the possibility that the inflaton would be the usual electroweak Higgs field (in the context of standard particle physics and gravity). However, it could still be a component of the Higgs field associated with the breaking of the GUT symmetry (although this question is very subtle and related to supersymmetry, supergravity, etc.) There are many interesting research activities in this direction.

2. inflation predicts a background of gravitational waves which have not been detected. Hopefully, this is only a matter of sensitivity: future CMB experiments might see these primordial gravitational waves. If they do, there would be one more very convincing evidence in favor of inflation (primordial gravitational waves would be the "smoking gun" of inflation), and the previous issue would become much more interesting and promising, since we would finally know the energy scale of inflation, as well as the details of the inflaton potential $V(\varphi)$ within some interval $\Delta\varphi$. As long as we don't see these primordial gravitational waves, we can only constrain the function V^3/V'^2 (see Eq. (7.41)). There exist one-parameter families of potentials all giving the same combination V^3/V'^2, and hence the same primordial scalar spectrum. Hence, the scale of inflation will remain unknown until primordial gravitational waves are observed. Of course, this might never

occur, in which case the theory of inflation would not be excluded, but its existence would not be proved or disproved as convincingly as one would like to.

Conclusions

Modern cosmology offers a detailed and self-consistent scenario, able to explains most (if not all) observations of the global properties of the universe. The most impressive success of the past years is the fact that cosmological perturbation theory (with initial conditions motivated by inflation) allowed to predict the non-trivial spectrum of CMB temperature fluctuations much before it was actually observed; the good agreement between the CMB data obtained by WMAP and the predictions of the minimal ΛCDM scenario is one of the greatest successes of modern science. The fact that the value of ω_b deduced from WMAP agrees so well with that infered from nucleosynthesis is also particularly impressive.

However, the observations of the last decade reveal that the universe contains nearly 25% of dark matter and 70% of cosmological constant (or of another fluid leading to accelerated expansion today, generically called dark energy), which are both of completely unknown nature and origin. This two issues are now the main challenges in cosmology (the third challenge being to understand the nature and origin of the inflaton). Next semester, the course of Pierre Salati will provide possible clues about these two fundamental issues.

Bibliography

[1] *An Introduction to Modern Cosmology*, by Andrew Liddle, John Wiley & Sons, Chichester, 2003.
[2] *Modern cosmology*, by Scott Dodelson, New York, NY: Academic Press, 2003.
[3] *The Early universe*, by Edward W. Kolb and Michael S. Turner, Redwood City, CA: Addison-Wesley, 1990.
[4] *Cosmologie primordiale*, par Patrick Peter et Jean-Philippe Uzan, Editions Belin, Paris, 2005.
[5] *Cosmologie: des fondements théoriques aux observations*, par Francis Bernardeau, CNRS Editions, Paris, 2007.
[6] *Physical Foundations of Cosmology*, by Viatcheslav Mukhanov, Cambridge : Cambridge Univ. Press, 2005.
[7] *Cosmology*, by Steven Weinberg, Oxford : Oxford Univ. Press, 2008.