RWTH Aachen University, Master of Physics,
SS25

The Ingredients of the Universe

Julien Lesgourgues

July 17, 2025

Contents

Lecture schedule

Lecture 1	Monday 7.04	Chapter 1 Revisions (FLRW metric and ΛCDM model)
Lecture 2	Thursday 10.04	Chapter 2: 2.1.1 - 2.1.2 (Thermal equilibrium)
Lecture 3	Friday 11.04	Chapter 2: 2.1.3 - 2.1.5 (Intro. thermodynamics)
Lecture 4	Monday 14.04	Chapter 2: 2.2.1 - 2.2.2 (Early Universe)
Lecture 5	Monday 14.04	Chapter 2: 2.2.3 - 2.2.4 (Neutrinos and electrons)
Lecture 6	Thursday 17.04	Chapter 2: 2.2.5 (Nucleosynthesis)
Lecture 7	Thursday 24.04	Chapter 2: 2.2.6 - 2.3.3 (Photon and Dark Matter decoupling)
Lecture 8	Thursday 28.04	Chapter 3: 3.1 - 3.3 (Intro. perturbations)
Lecture 9	Monday 5.05	Chapter 3: 3.4 - 3.6 (Linearised equations)
Lecture 10	Thursday 8.05	Chapter 4: 4.1 (Motivations for inflation)
Lecture 11	Monday 12.05	Chapter 4: 4.2 - 4.3.1 (Slow-roll and quantum perturbations)
Lecture 12	Thursday 15.05	Chapter 4: 4.3.2 - 4.3.3 (Primordial perturbations)
Lecture 13	Monday 19.05	Chapter 4: 4.3.4 - 4.3.7 (Primordial power spectra)
Lecture 14	Thursday 22.05	Chapter 4: 4.4 (Inflationary models)
Lecture 15	Monday 26.05	Chapter 5: 5.1 (Thomson scattering)
Lecture 16	Monday 2.06	Chapter 5: 5.2 (Linearised Boltzmann equation)
Lecture 17	Thursday 5.06	Chapter 5: 5.2 (Anisotropy in given direction)
Lecture 18	Monday 16.06	Chapter 5: 5.3.1-5.3.2 (Temperature spectrum)
Lecture 19	Monday 23.06	Chapter 5: 5.3.3 (Line-of-sight integral in Fourier space)
Lecture 20	Thursday 26.06	Chapter 5: 5.4.1 (Perturbation evolution)
Lecture 21	Monday 30.06	Chapter 5: 5.4.2 (Contributions to CMB spectrum)
Lecture 22	Thursday 3.07	Chapter 5: 5.4.2 (Parameter effects)
Lecture 23	Monday 7.07	Chapter 5: 5.4.3 - 5.4.4 (Polarisation, tensors, observations)
Lecture 24	Thursday 10.07	Chapter 6: 6.1.1 - 6.1.2 (Linear matter power spectrum)
Lecture 25	Monday 14.07	Chapter 6: 6.1.3 BAO
Lecture 26	Thursday 17.07	(Summary, open questions, conclusions)

Exercise schedule

This exercise schedule is still preliminary and subject to variations.

	Assigned (midnight)	Discussed in tutorium	Uploaded (< midnight)	Corrected in tutorium	Graded and given back	
Exercises 1	Wed 9.04	–	Wed 23.04	Fri 25.04	Fri 2.05	Basic thermodynamics Relation time-temperature
Exercises 2	Wed 23.04	Fri 25.04	Wed 30.04	Fri 2.05	Fri 9.05	Chemical potentials, Entropy conservation
Exercises 3	Wed 30.04	Fri 2.05	Wed 7.05	Fri 9.05	Fri 16.05	Recombination, extra relics and BBN
Exercises 4	Wed 7.05	Fri 9.05	Wed 14.05	Fri 16.05	Fri 23.05	Recombination with He, Horizon problem
Exercises 5	Wed 14.05	Fri 16.05	Wed 21.05	Fri 23.05	Fri 30.05	Solving horizon problem, Analytical perturbations
Exercises 6	Wed 21.05	Fri 23.05	Wed 28.05	Fri 30.05	Fri 6.06	Klein-Gordon, slow-roll, adiabatic multi-fluid
Exercises 7	Wed 28.05	Fri 30.05	Wed 4.06	Fri 6.06	Fri 20.06	Power spectra, amplitude of primordial spectra
Exercises 8	Wed 4.06	Fri 6.06	Wed 18.06	Fri 20.06	Fri 27.06	Inflationary models
Exercises 9	Wed 18.06	Fri 20.06	Wed 25.06	Fri 27.06	Fri 4.07	Linearised Boltzmann equation
Exercises 10	Wed 25.06	Fri 27.06	Wed 2.07	Fri 4.07	Fri 11.07	Boltzmann hierarchy, tight coupling
Exercises 11	Wed 2.06	Fri 4.07	Wed 9.07	Fri 11.07	Fri 18.07	Perturbation evolution, peak scale
Exercises 12	Wed 9.07	Fri 11.07	Wed 16.07	Fri 18.07	Fri 25.07	Dynamical Dark Energy

Chapter 1

Recalls on homogeneous cosmology

Our Universe can be described either on average or taking fluctuations into account.

Thanks to observations of the Cosmic Microwave background (CMB), which show temperature fluctuations of the order of $\mathcal{O}(10^{-5})$, we have a direct visual proof that our Universe was nearly homogeneous and isotropic at early times.[1] Thus the homogeneity and isotropy of the early universe is a fact established by observations up to a high degree of confidence. This means that the early universe can be described mathematically using perturbation theory, with a homogeneous background plus small perturbations of order $\mathcal{O}(10^{-5})$.

We can model the perturbations and simulate their evolution. As we will see in chapter VI, these simulations allow us to understand how the universe became very inhomogeneous today, with clusters and halos of matter separated by voids, and compact objects like stars and planets. However, even today, the universe still appears as homogeneous after smoothing the matter distribution over a scale larger than a few tens of Megaparsecs[2], that is, about 10^{24} meters or 10^8 light years. This scale is still about a thousand times smaller than the diameter of the observable universe, so there is still a wide range of scales over which the universe can be described as homogeneous today.

Thus, cosmologists can split the study of the observable universe in two regimes: the regime of linear perturbation theory, which describes all scales at early times and only large scales at late times, and the non-linear regime, which describes small scales at late times. The first regime requires a good knowledge of general relativity and particle physics: it is usually best known by theoretical physicists. The second is closer to astrophysics, but it can also be studied at a very theoretical level, using concepts close to quantum field theory and statistical field theory. However, in this course, due to time constraints, we will only study linear perturbation theory.

[1] In reality, CMB observations only prove homogeneity on our photon last scattering surface, rather than in the whole universe at the time of decoupling. We further need to assume that all observers also see a homogeneous CMB map, and that we are not in a very special position, in order to conclude that the Universe was close to homogeneous and isotropic at least within a very wide region that encompassed our last scattering surface. If we relax the assumption that we are not in a special position, we can build inhomogeneous universe models. Such models lead to testable predictions which, so far, are not favoured by observations.

[2] 1 Mpc $\sim 3 \times 10^{22}$ m $\sim 3 \times 10^6$ ly.

One of the great advantages of linear perturbation theory is that the equations can be split in two sectors: the homogeneous background and the linear perturbations. In chapter 1 and 2, we will deal with the homogeneous background, while in the other chapters we will deal with linear perturbations.

1.1 The FLRW model

1.1.1 FLRW metric and cosmological distances

In the course *Relativity & Cosmology*, you have seen that the most general homogeneous and isotropic solution of the Einstein equations is given by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric,

$$ds^2 = -c^2dt^2 + a(t)^2\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\ d\phi^2)\right] \ . \tag{1.1}$$

We are using the metric signature $(-1,1,1,1)$ which is slightly more common in the fields of relativity and cosmology, while the metric signature $(1,-1,-1,-1)$ is slightly more common in particle physics.

The FLRW metric is appropriate to describe our universe at the background level. Here t is called the cosmological time while (r,θ,ϕ) are the spatial comoving coordinates. The function $a(t)$ is called the scale factor. Physical distances are given by the distances defined in coordinate space multiplied by the scale factor. Thus, the increase of $a(t)$ accounts for the expansion of the universe. Observers at fixed coordinates are called comoving observers. These observers have no motion, and still the distance between them increases together with $a(t)$, as a pure effect of general relativity and of the deformation of spacetime. Observers with non-static coordinates have a so-called peculiar velocity. The cosmological time t plays the role of proper time for any comoving observers. Indeed, along the worldline of comoving observers, one has $ds^2 = -c^2dt^2$ and thus $dt = \sqrt{-ds^2}/c$, which is the general definition of proper time.

We will sometimes use alternative time coordinates, such as the conformal time η defined through $a\,d\eta = dt$ and thus through

$$\eta = \int_{t_*}^{t} \frac{d\tilde{t}}{a(\tilde{t})} \ , \tag{1.2}$$

with an arbitrary reference time t_*. Conformal time often leads to simpler equations because with such a coordinate system, the metric is very simple: it has an overall scaling factor (often called a conformal factor) a:

$$ds^2 = a(\eta)^2\left\{-c^2d\eta^2 + \left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\ d\phi^2)\right]\right\} \ . \tag{1.3}$$

The constant k is related to the homogeneous spatial curvature of the universe. The comoving radius of curvature of the universe is $1/\sqrt{|k|}$, with positive or elliptic curvature for $k > 0$ and negative or hyperbolic curvature for $k < 0$. The universe may then have the geometry of a 3D-sphere or a homogeneous 3D horse saddle respectively. The physical radius of curvature is $R_c(t) = a(t)/\sqrt{|k|}$.

There is another way to construct a physical distance from the FLRW metric. First, one can define the expansion rate or Hubble rate as $H(t) = \dot{a}(t)/a(t)$, which has the dimension of inverse time (a dot will always denote a derivative with respect to cosmological time t). Then one can construct the distance

$R_H(t) = c/H(t) = c\,a/\dot{a}$ called the Hubble radius. The Hubble radius is a second radius of curvature of the FLRW spacetime, independent of the spatial curvature radius. R_c gives the curvature of spatial slices at fixed time. Intuitively, R_H further contributes to the curvature of the 2D manifold of coordinates (t, r) for fixed (θ, ϕ).

When you want to describe a phenomenon in the universe that takes place on a scale $\lambda \ll R_c$ and $\lambda \ll R_H$, like, for instance, some every-day life phenomenon on the earth, or even some phenomenon on stellar system scales or sub-galactic scales, you can usually neglect general relativistic effects and employ Newtonian mechanics in Minkowski spacetime.

When the spatial curvature vanishes, $k = 0$, it is equally convenient to use spherical comoving coordinates (r, θ, ϕ) or cartesian comoving coordinates (x, y, z) such that

$$\begin{aligned} ds^2 &= -c^2dt^2 + a(t)^2\left[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right] && (1.4)\\ &= -c^2dt^2 + a(t)^2\left[dx^2 + dy^2 + dz^2\right]\ . && (1.5)\end{aligned}$$

In *Relativity & Cosmology*, you have also seen that the wavelength λ of the light emitted by objects at time t and seen by observers at time t_0 gets redshifted by a factor $1 + z = \lambda(t_0)/\lambda(t) = a(t_0)/a(t) = a_0/a$, as a pure effect of the expansion of the universe. The traditional Doppler effect would bring an extra contribution to $1 + z$ when the emitter or the observer have a peculiar velocity. The expansion of the universe always leads to $a(t) < a(t_0)$ for any $t < t_0$. Thus, in absence of peculiar velocities, the light can only be redshifted, with $z > 0$, rather than blueshifted, with $z < 0$. The portion of spacetime that we can see is called our past-light cone. The images of objects that we can see have been emitted in the past. For instance, when we see an image that was emitted at a time when the scale factor was half of its current value, the wavelength gets stretched by a factor 2 and the image is seen with a redshift $z = 2 - 1 = 1$. The redshift z can be used as a measure of both distance and look-back time[3] to objects on our past light-cone. Thus, cosmologists often say that "a galaxy is located at a distance $z = 3$" or that "an event took place at redshift "z=10", even if these are slight abuses of language.

In lots of forthcoming calculations, we will need to perform changes of variable between t, η, a and z. These changes of variables rely on the differential relations:

$$a\,d\eta = dt, \qquad da = a\,H\,dt, \qquad dz = -\frac{a(t_0)\,da}{a^2}\ . \tag{1.6}$$

These relations can be combined in different ways, for instance:

$$dz = -\frac{a(t_0)}{a^2}\frac{da}{dt}dt = -\frac{a(t_0)}{a}Hdt = -(1+z)Hdt\ . \tag{1.7}$$

In order to do any exercise of cosmology, you should be very fluent with these frequently used relations.

Equal-time distance and comoving distance. We can define equal-time distances or instantaneous distances as proper distances between two events with the same time coordinate. We will usually assume that we[4] are located at

[3] The "look-back time" means "the time with respect to today", so the cosmological look-back time to an object emitting its image at t reads $t_0 - t$, and the conformal look-back time reads $\eta_0 - \eta$.

[4] here "we" means "observers" or "mankind" or "some instrument located on earth or onboard a a satellite".

the origin of the spherical comoving coordinate system. This does not mean that we are located at a special place form a physical point of view, since the FLRW metric is homogeneous: the origin plays a special role at the mathematical level, but not at the physical level. Then, let us measure the proper distance between two events: us at time t and an object of coordinates (r, θ, ϕ) at time t. Imagine that you simply draw a line relating these two events. Along this line, the line element is

$$ds = a(t) \left[\frac{dr^2}{1 - kr^2} \right]^{1/2} . \tag{1.8}$$

Thus the proper distance d is

$$d = a(t) \int_0^r \left[\frac{d\tilde{r}^2}{1 - k\tilde{r}^2} \right]^{1/2} . \tag{1.9}$$

We usually decompose this distance as $d = a(t)\, \chi(r)$ with

$$\chi(r) = \int_0^r \left[\frac{d\tilde{r}^2}{1 - k\tilde{r}^2} \right]^{1/2} . \tag{1.10}$$

The quantity $\chi(r)$ is not a physical distance, it is a coordinate distance. The coordinate distance $\chi(r)$ between two comoving objects is constant in time: thus $\chi(r)$ is called the comoving distance. However, the physical distance between two comoving objects increases proportionally to the scale factor. You have seen in the course *Relativity & Cosmology* that there is a simple analytical relation between χ and r:

$$\chi(r) = \begin{cases} \frac{1}{\sqrt{k}} \sin^{-1}(\sqrt{k}\, r) & \text{if} \quad k > 0, \\ r & \text{if} \quad k = 0, \\ \frac{1}{\sqrt{-k}} \sinh^{-1}(\sqrt{-k}\, r) & \text{if} \quad k < 0. \end{cases} \tag{1.11}$$

We can invert this relation as

$$r = f_k(\chi) \equiv \begin{cases} \frac{1}{\sqrt{k}} \sin(\sqrt{k}\, \chi) & \text{if} \quad k > 0, \\ \chi & \text{if} \quad k = 0, \\ \frac{1}{\sqrt{-k}} \sinh(\sqrt{-k}\, \chi) & \text{if} \quad k < 0. \end{cases} \tag{1.12}$$

Equal-time distances and comoving distances are a purely abstract notion, like coordinates. We cannot infer them directly from observations, since we can only see objects in our past-light cone, rather than in a slice of spacetime at time t_0. Suppose that an object located at comoving radius r emitted an image at time t that we see today at t_0. The equal-time distance and the comoving distance between us and the object refers to the distance between us and the object today, assuming that it has no peculiar velocity and that it remained at the same spatial coordinates. Thus this is really purely abstract.

We keep assuming that light was emitted at time t (and conformal time η) by an object located at comoving radius r, and reaches the origin at time t_0 (and conformal time η_0). Along the radial worldline of photons, $ds = 0 = d\theta = d\phi$ and

$$c\, dt = a \left[\frac{d\tilde{r}^2}{1 - k\tilde{r}^2} \right]^{1/2} = c\, a\, d\eta . \tag{1.13}$$

Thus

$$\left[\frac{d\tilde{r}^2}{1 - k\tilde{r}^2} \right]^{1/2} = c\, d\eta , \tag{1.14}$$

and

$$\chi(r) = \int_0^r \left[\frac{d\tilde{r}^2}{1-k\tilde{r}^2}\right]^{1/2} = c\int_\eta^{\eta_0} d\tilde{\eta} = c(\eta_0 - \eta) \ . \tag{1.15}$$

The quantity $\eta_0 - \eta$ is the look-back time to the object. Thus, in general, we can say that "the comoving distance to an object is equal to is conformal look-back time multiplied the speed of light". Using simultaneously conformal distances and comoving time is a way to make all relations look artificially like in Minkowski space, even if the universe expands and has spatial curvature.

Since equal-time distances and comoving distances are a purely abstract, they have a limited interest. You have seen in the course *Relativity & Cosmology* that we can also define some directly measurable distances in at least two ways: the luminosity distance and the angular diameter distance.

Luminosity distance. The luminosity distance d_L is the distance inferred at t_0 from the comparison between the absolute luminosity L and apparent luminosity l of an object located at radial coordinate r and redshift z, emitting is image at t. You have already proved that

$$\begin{aligned}
d_L &\equiv \sqrt{\frac{L}{4\pi l}} = a(t)\ r\ (1+z)^2 = a(t_0)\ (1+z)\ f_k\left(\chi\right) && (1.16)\\
&= a(t_0)\ (1+z)\ f_k\left(\int_0^r \left[\frac{d\tilde{r}^2}{1-k\tilde{r}^2}\right]^{1/2}\right) && (1.17)\\
&= a(t_0)\ (1+z)\ f_k\left(\int_0^t \frac{c\,d\tilde{t}}{a(\tilde{t})}\right) && (1.18)\\
&= a(t_0)\ (1+z)\ f_k\left(c(\eta_0-\eta)\right) && (1.19)\\
&= a(t_0)\ (1+z)\ f_k\left(\int_0^z \frac{c\,d\tilde{z}}{a(t_0)H(\tilde{z})}\right) \ . && (1.20)
\end{aligned}$$

This notion is only useful for objects whose absolute luminosity can be determined in some indirect way, not knowing the distance to them. Such objects exist and are called standard candles. For standard candles, we can measure independently L, l and z. Thus we can measure independently d_L and z. After considering several such objects, we can determine experimentally the curve $d_L(z)$. This curve depends both on the spatial curvature parameter k and on the expansion history $H(z)$. Thus, measuring the luminosity distance - redshift relation is one way to constrain the expansion history and the curvature of our universe.

Angular diameter distance. The angular diameter distance d_A is the distance inferred from the comparison between the proper length λ and the angular size θ of a stick located at radial coordinate r, emitting its image at t. You have already proved that

$$d_A \equiv \frac{\lambda}{\theta} = a(t)\ r = \frac{a(t_0)}{1+z}\ f_k(\chi) = \frac{d_L}{(1+z)^2} \ . \tag{1.21}$$

This notion is only useful for objects whose proper length can be determined in some indirect way, not knowing the distance to them. Such objects exist and are called standard rulers. For standard rulers, we can measure independently λ, θ and z. Thus we can measure independently d_A and z. After considering several such objects, we can determine experimentally the curve $d_A(z)$. This

curve depends both on the spatial curvature parameter k and on the expansion history $H(z)$. Thus, measuring the angular distance - redshift relation is another way to constrain the expansion history and the curvature of our universe. However, the luminosity distance and the angular distance provide the same type of information, since they are trivially related by $d_L(z) = (1+z)^2 d_A$. We will see important applications of these concepts in the next sections.

If we compare these distances with the equal-time distance evaluated today, given by equation (1.9) with $t = t_0$, that is, $d = a(t_0)\chi$, we see that d_A, d_L and d are all different. They differ, first, by three different powers of the factor $(1+z)$, and second, by the fact that d involves the comoving distance χ, while d_A and d_L involve the comoving radius $r = f_k(\chi)$:

$$d = a(t_0)\,\chi\ , \quad d_L = a(t_0)\ (1+z)\ f_k\,(\chi)\ , \quad d_A = \frac{a(t_0)}{1+z}\ f_k(\chi)\ . \tag{1.22}$$

1.1.2 Momentum of particles in the FLRW universe

In chapter 2, when we will study the evolution of particles in the universe, we will need to understand how the momentum of individual particles evolves when the universe expands, both for massless and massive particles.

From now on, we will adopt natural units in which $c = 1$. Additionally, in the following proof, we will assume for simplicity that the spatial curvature is zero and that we can write the FLRW metric in cartesian comoving coordinates as

$$ds^2 = -(dx^0)^2 + a(x^0)^2\left[(dx^1)^2 + (dx^2)^2 + (dx^3)^2\right] \tag{1.23}$$

where $x^0 = t$ is proper time, while x^i are cartesian comoving coordinates. With a non-zero spatial curvature, the intermediate steps of the calculation would be more cumbersome, but the final results would be identical.

In the course *Relativity & Cosmology*, you studied the difference between "valid coordinates" and "physical coordinates":

- valid coordinates are such that the metric is well behaved and that all calculations can be performed without any problem. The FLRW coordinates of equation (2.1) are a valid coordinate system.
- physical coordinates represent the actual time and actual distances measured by the clocks and rulers of observers at rest in the coordinate system. In special relativity, the coordinates were always physical. In GR, one can recovers locally some physical coordinates by doing a change of variable that brings the metric locally to the Minkowski form, up to order two corrections.

Of course, the spatial comoving coordinates of the flat FLRW universe are not physical coordinates, since $g_{ii} = a(x^0)^2 \neq 1$. Let us assume that we want to define physical coordinates $\bar{x}^\alpha$ close to a particular event E of spacetime. Obviously, the change of coordinates that leads to physical coordinates $\bar{x}^\alpha$ around this event is such that

$$\left.\frac{\partial x^\mu}{\partial \bar{x}^\alpha}\right|_E = \mathrm{diag}\left(1, a(\bar{x}^0)^{-1}, a(\bar{x}^0)^{-1}, a(\bar{x}^0)^{-1}\right).$$

Only then can the FLRW metric

$$g_{\mu\nu} = \mathrm{diag}(1, a(x^0)^2, a(x^0)^2, a(x^0)^2)$$

be transformed in E into the Minkowski metric[5]

$$\eta_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, 1).$$

Thus, in the new coordinate system and in E, infinitesimal intervals of time are unchanged ($t = x^0 = \bar{x}^0$ represents proper time in the two coordinate systems), and infinitesimal distances are multiplied by a, showing as usual that physical distances in FLRW are given by coordinate distances times a.

Under this change of coordinate, the energy-momentum 4-vector P^μ transforms like any contravariant vector (e.g. like x^μ): $\bar{P}^0 = P^0$, $\bar{P}^i = a(\bar{t})P^i$. The coordinates which are (locally) physical are the ones with bars. Hence, like in SR, we can say that $\bar{P}^0$ (resp. $\bar{P}^i$) represents the energy (resp. 3-momentum) of particles measured by comoving observers, which are by definition at rest with respect to x^i, and hence also at rest locally with respect to $\bar{x}^i$.

We conclude that the energy-momentum 4-vector P^μ defined in the FLRW coordinate system has the following interpretation: P^0 is the energy of particles measured by comoving observers; and the product aP^i is the 3-momentum of particles measured by comoving observers. Since the vector aP^i plays an important role, we will denote it p^i (this notation is lighter than the previous $\bar{P}^i$, although it is the same thing: $p^i = \bar{P}^i = aP^i$).

Next, we want to know how the physical energy and momentum of a free-falling particle evolves in the FLRW universe, as a consequence of cosmological expansion. This will be given by the geodesic equation. We will distinguish two cases: massive and massless (ultra-relativistic) particles.

Momentum of massive particles

Consider a massive particle following a geodesic $x^\mu(\lambda)$, where λ is a parameter monotonically increasing along the trajectory. By definition, the energy-momentum 4-vector P^μ is the tangent vector $P^\mu = m\frac{dx^\mu}{d\lambda}$, provided that λ is the proper time of the particle and is normalised in such a way that:

$$\vec{P} \cdot \vec{P} = P_\mu P^\mu = g_{\mu\nu} P^\mu P^\nu = m^2 \vec{U} \cdot \vec{U} = -m^2 \ . \tag{1.24}$$

(This normalisation condition is equivalent to requiring that λ is the proper time of the particle, which is different from the proper time $t = x^0$ of all comoving observers, since the particle is not comoving.) The geodesics equation

$$\frac{d^2 x^\alpha}{d\lambda^2} + \Gamma^\alpha_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} = 0 \tag{1.25}$$

gives after multiplication by m^2

$$m\frac{dP^\alpha}{d\lambda} + \Gamma^\alpha_{\mu\nu} P^\mu P^\nu = 0 \ . \tag{1.26}$$

[5] You could naively think that the transformation from comoving coordinates to coordinates that are locally physical in E is given by $\bar{x}^0 = x^0$, $\bar{x}^i = a(x^0_E)x^i$. Things are not that simple, because we would then get a mteric $\bar{g}_{\mu\nu}(x^\alpha)$ such that $\bar{g}_{ii}(\bar{x}^\alpha) = (a(x^0)/a(x^0_E))^2 = 1 + 2[\dot{a}(x^0_E)/a(x^0_R)](x^0 - x^0_E) + \mathcal{O}((x^0 - x^0_E)^2)$. In the vicinity of E, the new metric would differ from the Minkowski one by terms linear in $(x^0 - x^0_E)$ and would not describe the tangent space in E to the spacetime manifold. We could fix this by adding to the transformation law a higher-order term $x^i = a(x^0_E)^{-1}\bar{x}^i - [\dot{a}(x^0_E)/a(x^0_E)^2](\bar{x}^i - \bar{x}^i_E)(x^0 - x^0_E)$. With such a correction, everything we wrote in the main text still holds, but additionally, there are some cancellations leading to $\bar{g}_{ii} = 1 + \mathcal{O}((x^0 - x^0_E)^2)$.

We can write the first term as

$$m\frac{dP^\alpha}{d\lambda} = m\frac{dt}{d\lambda}\frac{dP^\alpha}{dt} = P^0\frac{dP^\alpha}{dt} \ , \tag{1.27}$$

so

$$P^0\frac{dP^\alpha}{dt} + \Gamma^\alpha_{\mu\nu}P^\mu P^\nu = 0 \ . \tag{1.28}$$

This is true in general. Now we want to write the evolution of the spatial part P^i in the flat FLRW universe. After computing the Christofell symnbols, one finds that $\Gamma^i_{00} = \Gamma^i_{jk} = 0$, and $\Gamma^i_{0j} = \Gamma^i_{j0} = \delta^i_j\,(\frac{\dot a}{a})$, where $\dot a \equiv \frac{da}{dt}$. So

$$P^0\frac{dP^i}{dt} + \Gamma^i_{j0}P^jP^0 + \Gamma^i_{0j}P^0P^j = 0 = P^0\frac{dP^i}{dt} + 2\frac{\dot a}{a}P^0P^i \ . \tag{1.29}$$

So P^0 can be simplified, and we are left with

$$\frac{dP^i}{dt} + 2\frac{\dot a}{a}P^i = 0 \ . \tag{1.30}$$

Finally, this can be written as

$$\frac{dP^i}{P^i} = -2\frac{da}{a} \ , \tag{1.31}$$

which shows that P^i scales like $a(t)^{-2}$ when the universe expands. Hence the physical momentum defined at each event by $p^i \equiv a(t)P^i$ scales like $a(t)^{-1}$ due to the expansion. We can infer the behavior of the energy from the condition $g_{\mu\nu}P^\mu P^\nu = -m^2$. It gives

$$-(P^0)^2 + a(t)^2\delta_{ij}P^iP^j = -m^2 \ . \tag{1.32}$$

We write this replacing P^0 by the physical energy E and aP^i by the physical momentum p^i:

$$-E^2 + \delta_{ij}p^ip^j = -m^2 \ . \tag{1.33}$$

Since these are physical quantities, the previous relation is like in SR. The modulus $|\delta_{ij}p^ip^j|$ can be denoted p, and we get

$$E = \sqrt{m^2 + p^2} \ . \tag{1.34}$$

We have seen that p^i scales like a^{-1} with the expansion of the universe, so p^2 scales like a^{-2}. We conclude that the energy of massive particles is nearly constant in the non-relativistic limit ($p \ll m$) and scales like a^{-1} in the relativistic limit ($p \gg m$) .

Momentum of massless particles

The behavior of the physical energy and momentum of massless particles is exactly what would guess from the relativistic limit of massive particles: so one could skip this subsection. However, momentum is defined differently for massless particles, so for clarity, we give here the full proof in the exact massless limit.

Consider a massless particle following a geodesic $x^\mu(\lambda)$, where λ is a parameter monotonically increasing along the trajectory. We have seen that the energy-momentum 4-vector P^μ is the vector tangent to the trajectory, $P^\mu = \frac{dx^\mu}{d\lambda}$, where λ is normalised in such a way that $P^0 = \tilde P^0$ is the physical energy of the photon

measured by an observer at rest in the coordinate system, $P^0 = h\nu$. We also get by construction

$$\vec{P} \cdot \vec{P} = P_\mu P^\mu = g_{\mu\nu} P^\mu P^\nu = 0 \ . \tag{1.35}$$

The geodesics equation

$$\frac{d^2 x^\alpha}{d\lambda^2} + \Gamma^\alpha_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} = 0 \tag{1.36}$$

gives

$$\frac{dP^\alpha}{d\lambda} + \Gamma^\alpha_{\mu\nu} P^\mu P^\nu = 0 \ . \tag{1.37}$$

We can write the first term as

$$\frac{dP^\alpha}{d\lambda} = \frac{dt}{d\lambda} \frac{dP^\alpha}{dt} = P^0 \frac{dP^\alpha}{dt} \ , \tag{1.38}$$

so

$$P^0 \frac{dP^\alpha}{dt} + \Gamma^\alpha_{\mu\nu} P^\mu P^\nu = 0 \ . \tag{1.39}$$

This relation is identical for massive and massless particles. Hence the rest of the proof is identical: P^i scales like $a(t)^{-2}$ and $p^i \equiv a(t) P^i$ scales like $a(t)^{-1}$. Finally the condition (1.35) gives $E = \sqrt{a^2 \delta_{ij} P^i P^j} = p$. We find that E scales like $a(t)^{-1}$. Since E is the inverse of a wavelength, this is consistent with the fact that wavelengths $\lambda(t)$ scale with $a(t)$.

1.1.3 Friedmann and energy conservation equation

In the course *Relativity & Cosmology*, you have seen that when you substitute the FLRW metric in the Einstein equation $G^0_0 = 8\pi G\, T^0_0$, you obtain the Friedmann equation:

$$H^2 = \left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho - \frac{k}{a^2} \ , \tag{1.40}$$

where ρ is the total background energy density in the universe. It is one of the components of the background stress-energy tensor of the universe,

$$T^\mu_{\ \nu} = (\rho + p) U^\mu U_\nu + p \delta^\mu_\nu = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \ , \tag{1.41}$$

where p is the background pressure. This is the most general form of the stress-energy tensor that is compatible with the assumption of homogeneity and isotropy.

The Einstein equation is always relating the curvature of spacetime to the matter content. The Friedmann equation is a particular example of such a relation. This is particularly obvious when we rewrite the Friedmann equation in the equivalent form

$$G_{00} = 3 \left[\frac{k}{a^2} + \left(\frac{\dot{a}}{a} \right)^2 \right] = 3 \left[R_c^{-2} + R_H^{-2} \right] = 8\pi G \rho \ . \tag{1.42}$$

You have already seen that the Einstein equation implies Bianchi identities of the form $G^{\mu}{}_{\nu;\mu} = T^{\mu}{}_{\nu;\mu} = 0$. The first Bianchi identity $T^{\mu}{}_{0;\mu} = 0$ is nothing but the energy conservation equation. In the FLRW case this relation reads

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p) \ . \tag{1.43}$$

Here, ρ and p stand for the total energy and pressure, summed over each individual component of the universe, for instance, photons (i.e. electromagnetic radiation) and non-relativistic matter. However, when the species are all decoupled form each other, with no transfer of energy and momentum between them, the relation must be satisfied by each individual species i: $\dot{\rho}_i = -3\frac{\dot{a}}{a}(\rho_i + p_i)$.

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 cases in cosmology are:

- *non-relativistic matter,* for which $p \simeq 0$ and

$$\dot{\rho} = -3\frac{\dot{a}}{a}\rho \qquad \Rightarrow \qquad \rho \propto a^{-3}. \tag{1.44}$$

  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[6] is by definition constant. Since a comoving volume V increases like $V \propto a^3$ in physical units, ρ decreases like a^{-3}. In the jargon of cosmology, the non-relativistic component of the cosmological fluid is just called "matter".

- *ultra-relativistic matter,* for which the equation of state reads $p = \rho/3$, and

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(1 + \frac{1}{3}\right)\rho = -4\frac{\dot{a}}{a}\rho \qquad \Rightarrow \qquad \rho \propto a^{-4}. \tag{1.45}$$

  Thus the energy density of an ultra-relativistic fluid gets diluted *faster* than that of non-relativistic matter. The reason is that the energy of each individual photon or massless particle is given by $E = h\nu = h/\lambda$. λ gets redshifted by the expansion, thus $E \propto a^{-1}$, in order to respect energy conservation (having $E =$ constant for an ultra-relativistic particle in an expanding universe would be a violation of energy conservation, with energy creation out of nothing. This effect, together with that of dilution, gives $\rho \propto a^{-4}$. In the jargon of cosmology, the ultra-relativistic component of the cosmological fluid is just called "radiation".

- *a cosmological constant,* for which the equation of state reads $p = -\rho$, and

$$\dot{\rho} = 0 \qquad \Rightarrow \qquad \rho = \text{constant}. \tag{1.46}$$

[6] In cosmology, we use very often the notion of comoving volume. This means "a three-dimensional volume whose boundary is defined by fixed comoving coordinates". Thus, by definition, comoving particles can be either inside or outside a comoving volume, but cannot cross its boundaries. The physical volume of a comoving volume $\mathcal{V}$ must then be proportional to a^3. Indeed, if we call V_0 the "coordinate volume" of $\mathcal{V}$, its physical volume must be $V = V_0 a^3$.

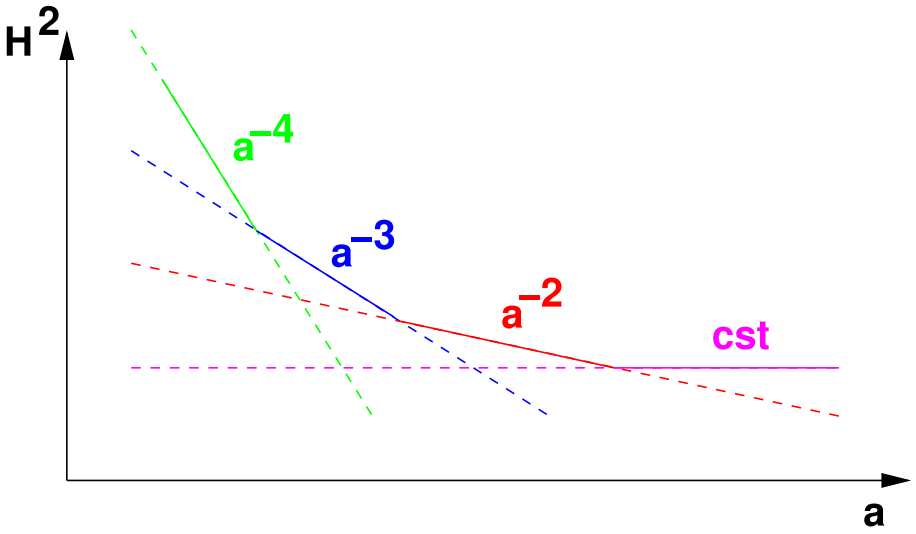

Figure 1.1: Evolution of the square of the Hubble parameter, in a scenario in which all typical contributions to the universe expansion (radiation, matter, curvature, cosmological constant) dominate one after each other.

Such an energy does not get diluted by the expansion! A priori, a cosmological constant could be present in the universe, either as a purely geometrical term (as Einstein once proposed) 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.

1.1.4 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:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho_{\rm R} + \frac{8\pi G}{3}\rho_{\rm M} - \frac{k}{a^2} + \frac{\Lambda}{3} \tag{1.47}$$

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 1.1). 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 → matter domination today.
- radiation dom. → matter dom. → curvature dom. today
- radiation dom. → matter dom. → cosmological constant dom. today

But all the cases that do not respect the order (like for instance: curvature domination → 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:

$$\frac{\dot{a}^2}{a^2} \propto a^{-4}, \qquad a(t) \propto t^{1/2}, \qquad H(t) = \frac{1}{2t}, \qquad R_H(t) = 2t. \tag{1.48}$$

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

2. Matter domination:

$$\frac{\dot{a}^2}{a^2} \propto a^{-3}, \qquad a(t) \propto t^{2/3}, \qquad H(t) = \frac{2}{3t}, \qquad R_H(t) = \frac{3}{2}t. \tag{1.49}$$

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

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

$$\frac{\dot{a}^2}{a^2} \propto a^{-2}, \qquad a(t) \propto t, \qquad H(t) = \frac{1}{t}, \qquad R_H(t) = t. \tag{1.50}$$

   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:

$$\frac{\dot{a}^2}{a^2} \to \text{constant}, \qquad a(t) \propto \exp(H\,t), \qquad H = 1/R_H = \sqrt{\Lambda/3}. \tag{1.51}$$

   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.

1.1.5 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[7]. Let us express this idea mathematically. We take the Friedmann equation, evaluated today, and divide it by H_0^2:

$$1 = \frac{8\pi G}{3H_0^2}\left(\rho_{\mathrm{r}0} + \rho_{\mathrm{m}0}\right) - \frac{k}{a_0^2 H_0^2} + \frac{\Lambda}{3H_0^2}. \tag{1.52}$$

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

$$\Omega_{\mathrm{r}} = \frac{8\pi G}{3H_0^2}\rho_{\mathrm{r}0}, \tag{1.53}$$

$$\Omega_{\mathrm{m}} = \frac{8\pi G}{3H_0^2}\rho_{\mathrm{m}0}, \tag{1.54}$$

$$\Omega_k = -\frac{k}{a_0^2 H_0^2}, \tag{1.55}$$

$$\Omega_\Lambda = \frac{\Lambda}{3H_0^2}, \tag{1.56}$$

$$\tag{1.57}$$

and the "matter budget" equation is

$$\Omega_{\mathrm{r}} + \Omega_{\mathrm{m}} + \Omega_k + \Omega_\Lambda = 1. \tag{1.58}$$

The universe is flat provided that

$$\Omega_0 \equiv \Omega_{\mathrm{r}} + \Omega_{\mathrm{m}} + \Omega_\Lambda \tag{1.59}$$

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

$$\rho_{\mathrm{cr}}(t) = \frac{3H^2(t)}{8\pi G}. \tag{1.60}$$

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

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:

$$\Omega_x = \frac{\rho_{x0}}{\rho_{\rm cr0}}. \tag{1.61}$$

The physical density today ρ_{x0} of a component x can be expressed in standard units, e.g. g.cm^{-3}. Another alternative is to decompose it using the "reduced Hubble parameter" h:

$$H = 100\ h\ \text{km s}^{-1}\text{Mpc}^{-1}. \tag{1.62}$$

Then the physical density today ρ_{x0} of a component x reads:

$$\rho_{x0} = \Omega_x \frac{3H_0^2}{8\pi G} = \Omega_x h^2 \frac{3(H_0/h)^2}{8\pi G} \tag{1.63}$$

$$= \Omega_x h^2 \times 1.8788 \times 10^{-29} \text{g.cm}^{-3}\ . \tag{1.64}$$

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

The parameters Ω_x, ω_x, H_0, h are all called "cosmological parameters", bit of course they are not all insependent of each other. So far, we can say that the evolution of the Friedmann universe can be described entirely in terms of four independent parameters, for instance:

$$\Omega_{\rm r}, \Omega_{\rm m}, \Omega_\Lambda, H_0. \tag{1.65}$$

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

1.2 Preliminary overview of the ΛCDM model

In this section, we will very briefly summarise the main assumptions of the standard cosmological model, without trying to prove them yet. It is only in the later sections that you will understand precisely how observations (of primordial element abundances, of CMB anisotropies, etc.) allowed to establish this model on a firm basis.

1.2.1 The Hot Big Bang model

Obviously our universe contains at least non-relativistic matter and has a non-zero matter density parameter $\omega_{\rm m}$. Indeed, what we see (planets, stars, intergalactic gas) is made up of non-relativistic matter.

The pioneers of the understanding of Big bang Nucleosynthesis, in particular, George Gamow in 1940, understood that we also need to have a lof of photons around, in order to increase the expansion rate at the time of nucleosynthesis and find that the most abundant element in the universe after nucleosynthesis is hydrogen. We will come back to this in details in chapter II. This prediction of Gamow has been proved by the discovery of the CMB by Penzias and Wilson in 1964, with a temperature $T_{\rm CMB} = T_{\gamma\,0} = 2.7255$ K. As we shall see later, this photon temperature today implies a photon density parameter $\omega_\gamma \sim 2.49 \times 10^{-5}$. Given the measured value of H_0 and h, which corresponds to $h \sim 0.7$, this gives $\Omega_\gamma \sim 5 \times 10^{-5}$. We will see later that the radiation component of the universe is made at least of photons and neutrinos, but the most abundant

ultra-relativistic particles are photons. For the time being, we can neglect the neutrino contribution and identify the radiation density with the photon density, such that $\omega_{\rm r} \sim 2.5 \times 10^{-5}$.

If the universe contains only matter and radiation, there should have been two stages: an early stage of Radiation Domination (RD), with $\rho \sim \rho_{\rm r} \propto a^{-4}$, and a late stage of Matter Domination (MD), with $\rho \sim \rho_{\rm m} \propto a^{-3}$. The two energy densities can be parametrised as

$$\begin{aligned} \rho_{\rm r} &= \rho_{\rm r0} \left(\frac{a_0}{a}\right)^4 = \Omega_{\rm r} \rho_{\rm c0} (1+z)^4 \\ &= \Omega_{\rm r} \frac{3H_0^2}{8\pi G}(1+z)^4 = \omega_{\rm r} \frac{3(H_0/h)^2}{8\pi G}(1+z)^4 \ , \qquad (1.66) \\ \rho_{\rm m} &= \rho_{\rm m0} \left(\frac{a_0}{a}\right)^3 = \Omega_{\rm m} \rho_{\rm c0} (1+z)^3 \\ &= \Omega_{\rm m} \frac{3H_0^2}{8\pi G}(1+z)^3 = \omega_{\rm m} \frac{3(H_0/h)^2}{8\pi G}(1+z)^3 \ . \qquad (1.67) \end{aligned}$$

Thus the redshift of equality, that is the redshift $z_{\rm eq}$ at which $\rho_{\rm m} = \rho_{\rm r}$ and the universe goes from RD to MD, can be inferred from

$$1 + z_{\rm eq} = \frac{\Omega_{\rm m}}{\Omega_{\rm r}} = \frac{\omega_{\rm m}}{\omega_{\rm r}} \ . \qquad (1.68)$$

We will see in the next subsection how this ratio can be estimated.

In the XXth century, people called "Cold Big Bang" the scenario with MD only, and "Hot Big Bang" the scenario with RD followed by MD, because in that scenario, just after the singularity, the universe is dominated by ultra-relativistic particles with a large pressure $p \sim \rho/3$ and thus large kinetic energy, which can be called "hot".

The two models competed for many years, but the observation fo the CMB ruled out the Cold Big Bang model.

1.2.2 The need for a cosmological constant

Once the Hot Big Bang scenario was established in the 1960's, two big uncertainties remained: is there a curvature dominated stage and/or a cosmological constant dominates stage at late times? Or, instead, is our universe still dominated by matter today?

It became clear that our universe could not be matter dominated today when people found that the theoretical age of the universe in a matter-dominated universe was smaller than the age of the oldest known objects in the universe, estimated, e.g. from isotope ratios. This was called the age problem. The solution was to have either $\Omega_\Lambda > 0$ or $\Omega_k < 0$.

The evidence for 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 as a probe of the luminosity distance – redshift relation, and thus, of the scale factor evolution.

The top panel of figure 1.2 shows the luminosity distance – redshift diagram for the Pantheon data set, released in 2018. The vertical axis called "magnitude" is essentially equivalent to $\ln(d_L(z))$. The corresponding constraints on

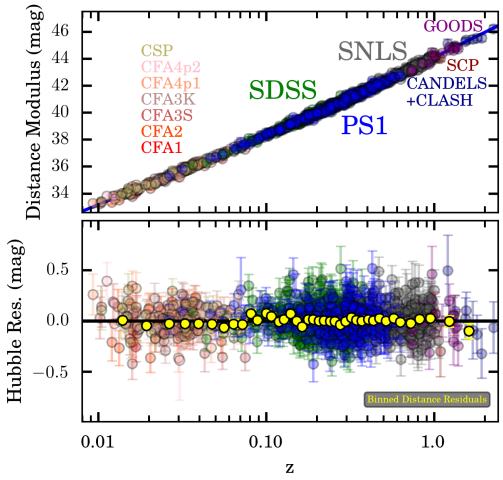

Figure 1.2: (Top panel) Luminosity distance – redshift diagram with the supernovae of the Pantheon sample. (Lower panel) Same data points and errors, divided by the theoretical prediction for the best fit ΛCDM model. *Plot taken from the preprint [arXiv:1710.00845] by D. M. Scolnic et al.*

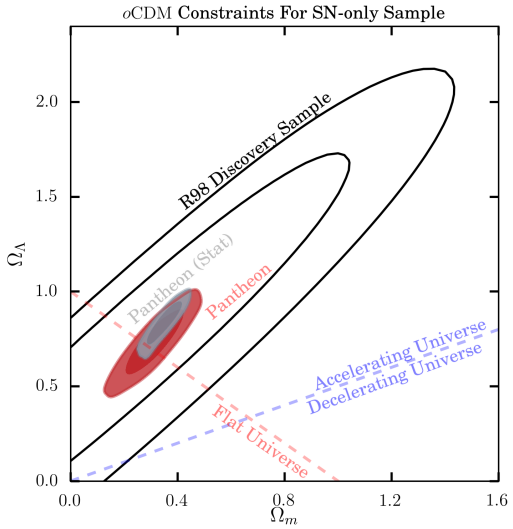

Figure 1.3: Contours at the 68.3% and 95.5% confidence levels in the ($\Omega_{\rm m}$, Ω_Λ) plane from historical supernovae data from 1998 (black lines) and from the 2018 Pantheon sample (red contours) These plots are all assuming a ΛCDM cosmology, as we are doing in this chapter. *Plot taken from the preprint [arXiv:1710.00845] by D. M. Scolnic et al.*

$\Omega_{\rm m}$ and Ω_Λ are displayed in Figure 1.3. They strongly suggest the existence of a cosmological constant today ($\Omega_\Lambda > 0$) and the fact that the universe is currently in accelerated expansion. The supernovae data does not say whether the parameter Ω_k is zero, small and positive, or small and negative.

The possibility to measure independently Ω_Λ and Ω_k raised from putting together two types of experiments: the measurement of the luminosity distance-redshift relation, $d_L(z)$, using Type Ia supernovae at redshifts $0 \leq z \leq 1$; and the measurement of the angular diameter distance, $d_A(z)$, using the CMB at redshift $z \sim 10^3$. You will understand the second measurement after the study of Chapter V on CMB physics.

These two measurements provide two constrains in the plane Ω_Λ, Ω_k, that are two orthogonal ellipses. They show that $\Omega_\Lambda \sim 0.7$ and $\Omega_k \sim 0$. Thus our universe is flat, but today, it is dominated by the density of the cosmological constant. The flatness implies that $\Omega_{\rm r}+\Omega_{\rm m}+\Omega_\Lambda = 1$. Since $\Omega_{\rm r}$ is of the order of 10^{-4} to 10^{-5}, we can neglect the radiation contribution to the budget equation, and assume simply $\Omega_{\rm m} + \Omega_\Lambda = 1$. With $\Omega_\Lambda \sim 0.7$ we get $\Omega_{\rm m} \sim 0.3$. Thus

$$\omega_{\rm m} = \Omega_{\rm m} h^2 \sim 0.15 \ . \tag{1.69}$$

If the universe is flat, its total density today is given by the critical density. Finally, the redshift of equality between matter and radiation can be estimated from

$$z_{\rm eq} = \frac{\omega_{\rm m}}{\omega_{\rm r}} \sim \frac{0.15}{2.5 \times 10^{-5}} \sim 6000 \ . \tag{1.70}$$

Later on, we will be able to make a more precise estimate taking into account the neutrino contribution to radiation. We will find that in reality, $z_{\rm eq} \simeq 3400$. We can also estimate the redshift of equality between matter and Λ, z_Λ, by equating the two densities

$$\rho_{\rm m} = \Omega_{\rm m} \frac{3H_0^2}{8\pi G}(1+z)^3 \ , \tag{1.71}$$

$$\rho_\Lambda = \Omega_\Lambda \frac{3H_0^2}{8\pi G} \ . \tag{1.72}$$

This gives

$$1 + z_\Lambda = \left(\frac{\Omega_\Lambda}{\Omega_{\rm m}}\right)^{1/3} = \left(\frac{\Omega_\Lambda}{1-\Omega_\Lambda}\right)^{1/3} \ . \tag{1.73}$$

With $\Omega_\Lambda \sim 0.7$ this gives $z_\Lambda \sim 0.3$, showing that the onset of Lambda Domination (ΛD) is very recent compared to the onset of MD.

1.2.3 The need for Cold Dark Matter

Cold Dark Matter was first introduced to solve astrophysical problems: the rotation curves of galaxies inside galaxy clusters and that of stars inside galaxies. These observations required the existence of a form of matter contributing to gravity but emitting no light.

The nature of this component is still not known. Dark Matter has not been detected directly, it has been detected only through its gravitational effect. Among various Dark Matter candidates, you find:

- non-relativistic particles that do not couple with electromagnetism and thus cannot emit, absorb or scatter light, called Cold Dark Matter particles;

- ultra-relativistic particles that also do not couple to electromagnetism, called Hot Dark Matter;

- modifications of the laws of gravity that would mimic Dark Matter and solve the galaxy rotation curve problem;
- a very light and inhomogeneous scalar field filling up the universe and whose kinetic and potential energy would mimic the role Dark Matter.

Hot Dark Matter was excluded in the 1990's because it would contradict observations concerning the formation history of galaxies and galaxy clusters (in particular, the fact that smaller objects form before larger objects, which is called hierarchical structure formation). Scenarios based on modifications of gravity were excluded by several observations between 2000 and 2010, including the map of CMB anisotropies and the observation of collisions between clusters (but some people do not acknowledge these observations and still believe in these modified gravity scenarios). Finally, highly precise observations of CMB anisotropies bring an excellent confirmation of the CDM assumption, but the scalar field assumption is also doing well, because current observations are not accurate enough to distinguish between CDM and ultra-light scalar dark matter.

We conclude that the minimal model of the universe is based on four species:

- Cold Dark Matter, with a density parametrized with $\omega_{\rm c}$.
- the rest of matter, that is, ordinary massive particles which belong to the Standard Model of particle physics. We will see that the most abundant particles today are ordinary atoms in their neutral or ionised form. Most of their energy density comes from the rest energy of the nuclei, that is, from the rest energy of protons and neutrons, which are called baryons. Thus, by tradition, cosmologists call ordinary matter "baryons". Their density is parametrised with $\omega_{\rm b}$. The total non-relativistic matter density today is parametrised by $\omega_{\rm m} = \omega_{\rm b} + \omega_{\rm c}$.
- radiation, which consist mainly of photons (we will see later about neutrinos), with a density parametrised by $\omega_{\rm r}$.
- the cosmological constant, usually parametrised by its fractional density Ω_Λ.

This model is called ΛCDM. It has at least four free parameters: $\omega_{\rm b}$, $\omega_{\rm c}$, $\omega_{\rm r}$ and Ω_Λ. However, $\omega_{\rm r}$ is usually considered as fully determined by the accurate measurement of the CMB temperature, and withdrawn from the list of free parameters. We are left with three free parameters to measure, ($\omega_{\rm b}$, $\omega_{\rm c}$, Ω_Λ). Flatness imposes a relation

$$\Omega_{\rm m} + \Omega_\Lambda = 1 = \frac{\omega_{\rm b} + \omega_{\rm c}}{h^2} + \Omega_\Lambda \ . \tag{1.74}$$

Thus the parameter h can be inferred from the others,

$$h = \left(\frac{\omega_{\rm b} + \omega_{\rm c}}{1 - \Omega_\Lambda} \right)^{1/2} \ . \tag{1.75}$$

As a consequence, it is equivalent to use the parameter basis ($\omega_{\rm b}$, $\omega_{\rm c}$, Ω_Λ) or ($\omega_{\rm b}$, $\omega_{\rm c}$, h).

In the next chapters, we will show how we have been able to measure these parameters with percent precision.

Chapter 2

Thermal history of the Universe

2.1 Relativistic quantum thermodynamics in the FLRW universe

We have seen that it is possible to use natural units such that $c = \hbar = 1$. It is also possible to choose the units of temperature in such a way that the Boltzmann constant k_B is one.

Also, in the rest of this chapter, we will assume for simplicity that there is no spatial curvature: $k = 0$. We do this for two reasons:

- a pedagogical reason: assuming spatial curvature renders the formulas more complicated (with several extra factors like $\sqrt{1-kr^2}$ appearing here and there in the equations), without affecting qualitatively the physical results of this chapter. To understand the physics, it is enough to work in a spatially flat universe.
- a physical reason: as argued in the Winter Semester, curvature tends to dominate only in the late universe (after radiation and matter domination), while this chapter deals mainly with the early universe (radiation and matter domination): hence the curvature can be safely neglected anyway.

When $k = 0$ it is simpler to use the FLRW metric defined with respect to cartesian comoving coordinates $(x^1, x^2, x^3) = (x, y, z)$:

$$ds^2 = -dt^2 + a(t)^2 \left[(dx^1)^2 + (dx^2)^2 + (dx^3)^2\right] \ . \tag{2.1}$$

We recall that $x^0 = t$ coincides with the proper time of all comoving observers, since $-g_{00} = 1$. This time is also often called *cosmological time*.

2.1.1 Phase-space distribution

Let us assume that the cosmological fluid is formed of many different species X (which can be either interacting with other species or free-streaming), each described by a phase-space distribution function $f_X(x^\mu, p^\nu)$. We use here the physical momentum defined in section 1.1.2, such that $p^0 = P^0 = E$ (resp. $p^i = aP^i$) is the physical energy (resp. physical 3-momentum) of particles

as measured by comoving observers.[1] By definition, at time x^0, the product $[f_X(x^\mu, p^\nu)\, d^3x^i\, d^3p^j]$ gives the number of particles in a small volume with $x^i \pm dx^i/2$, and with a momentum in the range $p^j \pm dp^j/2$.

The number of arguments can be reduced in the FLRW universe by noticing that:

- Homogeneity implies that f_X should be the same everywhere and should not depend on x^i.
- Isotropy implies that f_X should not depend on the direction of the 3-momentum p^i. However it could depend on the modulus $p = (\sum_i p_i^2)^{1/2}$.
- p^0 is not an additional independent argument if the particle mass m_X is known, since $p^0 = E = \sqrt{m_X^2 + p^2}$.

Hence the phase-space distribution can be written as a function of time and p only[2]: $f_X(t, p)$. The number density, energy density and pressure of each species read:

$$n_X(t) = \int d^3p^i\ f_X(t,p)\ , \tag{2.2}$$

$$\rho_X(t) = \int d^3p^i\ E\ f_X(t,p) = \int d^3p^i\ \sqrt{m_X^2 + p^2}\ f_X(t,p)\ , \tag{2.3}$$

$$P_X(t) = \int d^3p^i\ \frac{p^2}{3E}\ f_X(t,p)\ , \tag{2.4}$$

and isotropy implies that we can replace d^3p^i with $4\pi p^2 dp$.

The phase-space evolution f_X has a non-trivial evolution given by the expansion of the universe and by interactions between species. Interactions can be represented by a set of reactions $1 + 2 \leftrightarrow 3 + 4$ (for elastic scattering, $1 = 3$ and $2 = 4$). The Boltzmann equation states that along each geodesic, the evolution of f_X is given by the total derivative

$$\frac{df_X}{dt} = F[f_1, f_2, f_3, f_4] \tag{2.5}$$

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. Being a total derivative, the left-hand side contains the effect of expansion; the right-hand side contains the effect of interactions.

2.1.2 Kinetic (or thermal) equilibrium

If two species X and Y have frequent interactions (like elastic scattering $X + Y \longrightarrow X + Y$), 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,

[1] In different books you will find different arguments for the phase-space distribution function: 4-momentum in Friedmann coordinates, physical momentum, conjugate momentum, comoving momentum... These choices all have different meanings and properties. At the simple level of this course, you may just think that all these choices are equivalent and related to each other by a change of variable. We work with the physical momentum for simplicity, in order to have a straightforward interpretation of the results.

[2] Equivalently, we could choose to write it as a function of time and E only: this is just a matter of convention, and in some books you would often find $f_X(t, E)$ instead of $f_X(t, p)$.

the temperature T. However the distributions f_X are not all equal to each other. They depend on:

- the mass m_X of each species (the mass appears in the energy of each particle, $E_X = \sqrt{m_X^2 + p^2}$);
- an additional parameter μ_X, 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).
- the number of internal degrees of freedom (spin or helicity states) of the considered species (e.g. $g_X = 2$ for photons γ, electrons e^-, positrons e^+, protons p, anti-protons $\bar{p}$, neutrons n, anti-neutrons $\bar{n}$, or $g_X = 1$ for neutrinos ν_i and anti-neutrons $\bar{\nu}_X$ where X is one of e, μ or η).

Hence, a plasma of N particles in thermal equilibrium with known masses m_X and known quantum properties (fermion or boson, g_X) 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

$$f_X(t,p) = \begin{cases} \frac{g_X}{(2\pi)^3} \frac{1}{\exp\left[\frac{E_X - \mu_X}{T}\right]+1} = \frac{g_X}{(2\pi)^3} \frac{1}{\exp\left[\frac{\sqrt{m_X^2+p^2}-\mu_X(t)}{T(t)}\right]+1} & \text{(Fermi-Dirac)} , \\ \\ \frac{g_X}{(2\pi)^3} \frac{1}{\exp\left[\frac{E_X - \mu_X}{T}\right]-1} = \frac{g_X}{(2\pi)^3} \frac{1}{\exp\left[\frac{\sqrt{m_X^2+p^2}-\mu_X(t)}{T(t)}\right]-1} & \text{(Bose-Einstein)} . \end{cases} \tag{2.6}$$

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 X and Y is characterized by a "thermally averaged cross-section – velocity product" $\langle \sigma v \rangle$. The interaction rate (or scattering rate) of X is given by $\Gamma_X = n_Y \langle \sigma v \rangle$, that of Y by $\Gamma_Y = n_X \langle \sigma v \rangle$. A detailed study would show that the scattering is efficient enough for maintaining X in thermal equilibrium with Y provided that the scattering rate Γ_X is larger than the inverse of the characteristic time set by the universe expansion: namely, $\Gamma_X > H$. Intuitively, when $\Gamma_X < H$, the cross-section is so low or the species Y is so diluted that the chance for X to scatter over Y within a time comparable to the age of the universe becomes negligible. When all possible scattering reactions which could maintain X in thermal equilibrium have $\Gamma_X < H$, the species X 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: $p \propto a^{-1}$).

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

- *Non-degenerate gas of relativistic particles.* Let us assume for simplicity that $|\mu_X| \ll T$. A gas with a with negligible chemical potential is called

"non-degenerate". In this case,

$$f_X = \frac{g_X}{(2\pi)^3} \frac{1}{\exp\left[\frac{\sqrt{m_X^2+p^2}}{T}\right] \pm 1} \ . \tag{2.7}$$

From eq. (2.2), we see that in general the particles contributing mostly to the number density are those for which $p^2 f_X(p)$ is maximum. Let us show that this happens near $p \sim T$.

For $p \ll T$ and fermions, $f_X \simeq (...)\frac{1}{e^0+1} \simeq (...)\frac{1}{2}$ where $(...)$ means "some constant number". For $p \ll T$ and bosons, $f_X \simeq (...)\frac{1}{e^{m_X/T}-1} \simeq (...)\frac{1}{m_X/T} = (...)\frac{T}{m_X}$. In both cases, $p^2 f_X(p)$ increases proportionally to p^2 for $p \ll T$. In the opposite limit $p \gg T$, for both fermions and bosons, $p^2 f_X(p)$ decreases proportionally to $p^2 \exp[-p/T]$. Hence the maximum is reached between these two limits, for a value of the order of T.

We conclude that when $T \gg m_X$, a huge majority of particles has $p \gg m_X$. 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 $m_X \longrightarrow 0$. The result is found to be:

$$n_X = \frac{\zeta(3)}{\pi^2} g_X T^3 \quad \left(\times\frac{3}{4} \quad \text{for fermions}\right) \ , \tag{2.8}$$

$$\rho_X = \frac{\pi^2}{30} g_X T^4 \quad \left(\times\frac{7}{8} \quad \text{for fermions}\right) \ , \tag{2.9}$$

$$P_X = \frac{1}{3}\rho_X \ , \tag{2.10}$$

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_X. Note that the usual equation of state of a relativistic gas, $p = \sum_X p_X = \sum_X \rho_X/3 = \rho/3$, is recovered here. We conclude that boson and fermions in thermal equilibrium with each other and such that $m_X \ll T$ and $|\mu_X| \ll T$ share roughly the same number/energy density, apart from possible factors of order one.

- *Gas of non-relativistic particles.* In the limit $m_X \gg T$, the function $p^2 f_X(p)$ peaks in between T and m_X. To prove it, noticing that $\frac{m_X}{T} \gg 1$ and that $\exp[\frac{m_X}{T}] \gg 1$, we Taylor expand the phase-space distribution as:

$$f_X \simeq (...)\frac{1}{\exp\left[\frac{\sqrt{m_X^2+p^2}}{T} - \frac{\mu_X}{T}\right]} \tag{2.11}$$

$$\simeq (...)\exp\left[-\frac{m_X}{T}\sqrt{1+(p/m_X)^2} + \frac{\mu_X}{T}\right] \tag{2.12}$$

$$\simeq (...)\exp\left[-\frac{m_X}{T}\left(1+\frac{1}{2}(p/m_X)^2\right) + \frac{\mu_X}{T}\right] \tag{2.13}$$

$$\simeq (...)\exp\left[-\frac{m_X}{T} + \frac{\mu_X}{T} - \frac{p^2}{2Tm_X}\right] \tag{2.14}$$

So $p^2 f_X(p)$ is proportional to $p^2 \exp\left[-\frac{p^2}{2Tm_X}\right]$, i.e. increases like p^2 for $p \ll \sqrt{2Tm_X}$ and decreases exponentially for $p \gg \sqrt{2Tm_X}$. Hence the

maximum is reached between these two limits, for a value of the order of $\sqrt{Tm_X}$.

We conclude that when $T \ll m_X$, most particles have $p \ll m_X$: hence this limits describes a gas of non-relativistic particles. Then, a detailed integration shows that for both fermions and bosons

$$n_X = g_X \left(\frac{m_X T}{2\pi}\right)^{3/2} \exp\left[-\frac{(m_X - \mu_X)}{T}\right] , \tag{2.15}$$

$$\rho_X = m_X n_X , \tag{2.16}$$

$$P_X = T n_X \ll \rho_X . \tag{2.17}$$

Let us compare the number density of these particles with that of relativistic particles belonging to the same thermal bath, i.e. sharing the same temperature:

$$\frac{n_X^{\rm NR}}{n_Y^{\rm R}} = e^{\frac{\mu_X}{T}} \left[\frac{g_X}{g_Y}\frac{\sqrt{\pi}}{2\sqrt{2}\zeta(3)}\right] \left(\frac{m_X}{T}\right)^{3/2} e^{-\frac{m_X}{T}} . \tag{2.18}$$

The factor between brackets is of order one. The part after the brackets is much smaller than one since we assumed $m_X \gg T$. Hence, unless the chemical potential is huge ($\mu_X \gg m_X \gg T$, 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.

2.1.3 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:

$$\mu_1 + \mu_2 = \mu_3 + \mu_4 . \tag{2.19}$$

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:

$$\dot{n}_1 + 3Hn_1 = n_1 n_2 \langle\sigma v\rangle \left[\exp\left(\frac{-\mu_1 - \mu_2 + \mu_3 + \mu_4}{T}\right) - 1\right] . \tag{2.20}$$

Here, we made 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.

In thermal equilibrium, n_1 has only two unknown (and a priori time-dependent) parameters: the temperature (common to all species) and the chemical potential μ_1. In a few section, we will study entropy conservation, which gives the evolution of temperature with the scale factor and with time. Hence the only free function of time in $n_1(t)$ is $\mu_1(t)$. This means that the above Boltzmann equation can be seen as an equation of evolution for the chemical potential of each species, here, species one. It is a first-order equation, describing the way in which the system tracks an equilibrium solution. By solving this equation one should see that the system adjusts to a particular value of μ_1 (and hence to a particular n_1), which depends on the other variables H, n_2, $\langle\sigma v\rangle$, and $\mu_{2,3,4}/T$.

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.

2.1.4 Conservation of quantum numbers

First, let us clarify a point of vocabulary. We call "comoving volume" a closed region whose boundaries keep fixed comoving coordinates. We call "coordinate volume" the volume of a comoving volume measured in coordinate space, and "physical volume" the physical volume of a comoving volume, obtained by multiplying its coordinate volume by a^3. Thus, "comoving volume" refers to a geometric notion, not to a number; while "coordinate volume" and "physical volume" refer to numbers. For instance, let us call $\mathcal{V}$ a little square whose edges have coordinates $x_* \pm \frac{1}{2}\Delta x$, $y_* \pm \frac{1}{2}\Delta y$, $z_* \pm \frac{1}{2}\Delta z$. The coordinate volume of $\mathcal{V}$ is $\Delta x\,\Delta y\,\Delta z$, while the physical volume of $\mathcal{V}$ is $V = a^3\Delta x\,\Delta y\,\Delta z$.

If the number of particles of a given type X was conserved in any comoving volume, we would have $n_X V$ =constant, and thus $n_X a^3$ =constant. This is usually *not* the case since in general, the particles X 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_- (weighted by the absolute value of the charge so that $n_- > 0$). The total density of electric charge in the universe is then simply $n_Q \equiv n_+ - n_-$. Electric charge is a conserved number, so it cannot vary in a comoving volume due to annihilation/creation processes. Also, in homogeneous cosmology, it cannot vary because of more charges leaving the comoving volume than entering the comoving volume (or vice-versa): this would lead to an accumulation of charges in some regions and a depletion in other regions, which would contradict our assumption of a homogeneous universe. 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. (except 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 carries no conserved charge, nothing prevents reactions of the type $nX \to mX$ with $n \neq m$. This is the case for photons. For instance, as long as the universe contains electrons and positrons, the two reactions

$$3\gamma \longleftrightarrow e^+ + e^- \longleftrightarrow 2\gamma \tag{2.21}$$

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.

2.1.5 Entropy conservation

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 that the total entropy (i.e. the number of possible states) in any comoving volume is conserved.

In a universe containing a mixture of relativistic and non-relativistic particles, each obeying a Fermi-Dirac or Bose-Einstein distribution with a temperature T and chemical potential μ_X, it is possible to infer from the first law of thermodynamics applied to each fluid X inside a given comoving volume,

$$T\, dS_X = dE_X + P_X\, dV - \mu_X\, dN_X \ , \tag{2.22}$$

that the total entropy reads

$$s = \frac{\rho + P - \mu\, n}{T} \ , \tag{2.23}$$

with $\rho = \sum_X \rho_X$, $P = \sum_X P_X$ and $\mu\, n = \sum_X \mu_X\, n_X$. During radiation domination, ρ and P receive a dominant contribution from relativistic species like photons (and, as we shall see later, neutrinos) with a vanishing chemical potential, such that we can write

$$s \simeq \frac{\rho_{\rm r} + P_r}{T} = \frac{\rho_{\rm r} + \frac{1}{3}\rho_{\rm r}}{T} = \frac{4}{3}\frac{\rho_{\rm r}}{T} \ , \tag{2.24}$$

where $\rho_{\rm r}$ is the total energy density of relativistic species only. In this course, we will rarely use the expression of the entropy density during matter domination and Λ domination; however we could show in a small exercise that the result is unchanged: the entropy density is still given by the above formula involving

only relativistic densities.[3] Using equation (2.9), we get

$$\rho_{\rm r} = \frac{\pi^2}{30} g_* T^4 \ , \tag{2.25}$$

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

$$g_* = \sum_{\rm rel.bosons} g_X + \frac{7}{8} \sum_{\rm rel.fermions} g_X \ . \tag{2.26}$$

The entropy density is then

$$s = \frac{4}{3} \frac{\pi^2}{30} g_* T^3 \ , \tag{2.27}$$

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

$$n_{\rm tot} = \frac{\zeta(3)}{\pi^2} \left[\sum_{\rm rel.bosons} g_X + \frac{3}{4} \sum_{\rm rel.fermions} g_X \right] T^3 \ . \tag{2.28}$$

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!

2.2 Thermal history of the visible sector

2.2.1 Early stages

We keep using natural units in which $c = \hbar = (h/2\pi) = k_B = 1$. It is then possible to express everything (energy, mass, inverse length, inverse time, momentum, temperature) in units, for instance, of energy: the most conventional choice is then to use electron-Volts (eV) as a "universal" unit. Thus an energy density can be expressed in eV^4. Its quartic root $\rho^{1/4}$ can be expressed in eV and the total $\rho_{\rm tot}^{1/4}$ is called the "energy scale of the universe" at a given time. During radiation domination and in thermal equilibrium, it has the same order of magnitude as the temperature in eV, since $\rho_{\rm tot}^{1/4} \sim T$.

The earliest stages in the evolution of the universe are still partially unknown and subject to investigation, while the latest stages are very well modelled and

[3] In a nutshell: during matter domination, we have $\rho_{\rm m} > \rho_{\rm r}$, where $\rho_{\rm m}$ is the total density of non-relativistic species. However these species are mainly baryons and cold dark matter particles, which have a non-zero chemical potential that adjusts itself in order to preserve the conservation of the baryon number and of an equivalent number carried by cold dark matter particles. As will become clearer in the next section, this forces the chemical potential of each of these species to remain close to their mass, $m_X \simeq \mu_X$. These species contribute to the entropy density with a term $\rho_X + P_X - \mu_X n_X \simeq (m_X - \mu_X) n_X$. It is possible to quantify the difference $m_X - \mu_X$ using the conservation of B (or of the quantum number carried by the cold dark matter particle), and to prove that it is small enough to fulfil the inequality $(m_X - \mu_X) n_X \ll \rho_r$ by many orders of magnitude. Thus the contribution of non-relativistic particles to the entropy density can be neglected even during matter domination.

constrained 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 t_P, called the Planck time, which is such that $\rho \sim M_P^4 \sim (10^{19}\text{GeV})^4$, where the Planck mass is defined by $M_P = G^{-1/2}$. Since the Friedmann equation can also be written as $3M_P^2H^2 = 8\pi\rho$, the Planck time corresponds to $H = M_P$, i.e. to a Hubble radius equal to the Planck length, $R_H = 1/M_P = \lambda_P$.[4]

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, when $\rho^{1/4} \sim 10^{15}$ to 10^{16}GeV. Assuming radiation domination between the Planck time and the time of GUT symmetry breaking, this would correspond to the cosmological time $t \sim 10^{-32}$s. However, inflation could also take place at a much lower energy scale. Besides, we are not even sure that Grand Unification ever occurred.

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, although recent LHC results bring no evidence for such an extension, at least until now. The study of reheating shows that all these particles reached thermal equilibrium after some time. At such high energy, most (if not all) particles were ultra-relativistic ($T > m_X$), and the total energy and pressure were given by eq. (2.25).

The end of reheating marks the beginning of the most recent radiation dominated era, the one assumed by Gamow, Peebles and others in order to explain Nucleosynthesis, which we will refer to as just "Radiation Domination" (RD). Note that during this era, $T \propto a^{-1}$, $\rho \propto T^4 \propto a^{-4}$ and $t \propto t^{1/2}$ in good approximation, although these scaling laws 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 or "quark-hadron 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 address them). Let us mention that a particularly fascinating and important issue is the evolution of the baryon and lepton number.

[4] If we assume the same expansion law as during matter domination, that is, $a(t) = kt^{1/2}$ where k is a constant, then we find that $t_P = 10^{-44}$s. Note that this calculation makes no real sense, since extrapolating a law like $a(t) = kt^{1/2}$ before the Planck time is very hasardeous. However, the origin of time, $t = 0$, is often chosen in such a way, in order to have the simple relation $a(0) = 0$. It is nonsensical but it does not hurt, because choosing the origin of time to be at $t_P = 0$ or at $t_P - 10^{-44}$s=0 makes no difference in practise. It can just been see as a convention.

Let us focus first on the baryon number. Before reheating, there is no baryon number (there is only the scalar field responsible for inflation, which carries no quantum 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 if there was only one type of particle with a baryon number, say b with baryon number $B = 1$ and its antiparticle $\bar{b}$ with $B = -1$. After inflation, we should have $n_b = n_{\bar{b}}$. These particles could in principle annihilate through e.g.

$$b + \bar{b} \leftrightarrow n\gamma \tag{2.29}$$

(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 $n_b = n_{\bar{b}}$ just before the annihilation starts, then this 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, n_b should be slightly larger than $n_{\bar{b}}$ before annihilation. If we stick to the idea that baryons and anti-baryons are created by pairs after inflation, then the only possibility is that baryon number conservation is weakly 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, a mechanism that simple extensions of the standard model of particle physics can accommodate very well. When $T \sim m_b$, all baryons annihilate with antibaryons, excepted the few ones in excess, which remain around until 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

$$b + \bar{b} \leftrightarrow n\gamma \tag{2.30}$$

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 obtained from

$$n_B = n_b - n_{\bar{b}} = \frac{g_b}{(2\pi)^3} \int d^3p \left[\frac{1}{\exp(\frac{E-\mu_b}{T}) + 1} - \frac{1}{\exp(\frac{E+\mu_b}{T}) + 1} \right] . \tag{2.31}$$

In the relativistic limit $E = p$ this gives

$$n_B = \frac{g_b T^3}{6\pi^2} \left[\pi^2 \left(\frac{\mu_b}{T} \right) + \left(\frac{\mu_b}{T} \right)^3 \right] , \tag{2.32}$$

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

$$\frac{n_B}{n_b + n_{\bar{b}}} \simeq n_B / \left[2 \times \frac{3}{4} \frac{\zeta(3)}{\pi^2} g_b T^3 \right] \sim \left[\pi^2 \left(\frac{\mu_b}{T} \right) + \left(\frac{\mu_b}{T} \right)^3 \right] \tag{2.33}$$

but this is not a conserved number. Usually, the asymmetry is parameterized by n_B/s, which is really a conserved number since $n_B a^3$ is conserved at any time between baryogenesis and today, and sa^3 is conserved at any time. We will see later that in order to obtain the correct baryon density today,[5] we must assume that n_B/s is of the order of 10^{-10}.

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

$$n_b = g_b \left(\frac{m_b T}{2\pi}\right)^{3/2} e^{-\frac{(m_b-\mu_b)}{T}} , \tag{2.35}$$

so the quantity μ_b/T now varies with time, in order to maintain a constant $n_b a^3$. The role of the chemical potential is then to compensate the Bolztmann suppression factor, which implies that $\mu_b \simeq m_b$, up to small time-dependent logarithmic corrections[6] to μ_b, that ensure that n_b has the right order of magnitude and the right scaling $n_b \propto a^{-3}$.

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.

2.2.2 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:

[5] 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 is 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

$$\eta_b \equiv \frac{n_B}{n_\gamma} \sim 5 \times 10^{-10} . \tag{2.34}$$

[6] This brings us back to the footnote of section 2.1.5 and to the fact that the contribution of baryons to the entropy density is negligible with respect to that of relativistic particles, even during matter domination. Indeed, $\mu_b \simeq m_b$ implies that $\rho_b + P_b - \mu_b n_b = (m_b + T - \mu_b) n_B \simeq (m_b - \mu_b) n_B$ is small. Using the order of magnitude of n_b and n_γ, one can prove (as a small exercise) that $(m_b - \mu_b) n_B \ll \rho_\gamma$, thus the contribution of baryons to the entropy density is always negligible with respect to that of photons, even during matter domination when $\rho_b > \rho_\gamma$.

- 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 non-relativistic: $m \gg T$ (meaning that it should normally have decayed into lighter particles), but it is stabilised by the fact that it is the lightest particle carrying a given conserved quantum number. For instance, the proton is the lightest baryon, and we will study later dark matter particles, which are usually assumed to be the lightest particles of some new sector of particle physics carrying some new charge.

Hadrons. 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. We could think that neutrons have decayed because they are non-relativistic and heavier than protons. However the mass difference is only $m_n - m_p = 1.293$ MeV. When the thermal bath has a temperature of about 10 MeV, this difference is irrelevant, and neutrons can coexist with protons. Indeed, electrons and neutrons carry enough energy for converting a proton into a neutron through inverse beta decay: this only requires $m_n - m_p = 1.293$ MeV. We will actually prove that $n_n \simeq n_p$ at this temperature at the end of this subsection.

Charged leptons. In the charged lepton sector, μ, $\bar{\mu}$, τ and $\bar{\tau}$ 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, and both present. Electric neutrality implies $n_{e^-} - n_{e^+} = n_p$, so there must be a small excess of electrons over 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,

$$\frac{n_{e^-} - n_{e^+}}{n_{e^-} + n_{e^+}} \sim \frac{n_p}{s} \sim \frac{n_B}{s} \sim 10^{-10} \, . \tag{2.36}$$

We see that electric neutrality implies that the electron-positron asymmetry is as tiny as the initial baryon asymmetry. Both electrons and positrons are in thermal equilibrium and their annihilation reaction $e^- + e^+ \leftrightarrow n\gamma$ is in chemical equilibrium, so they must have opposite chemical potentials. Then, using equations (2.31-2.33) for electrons/positrons instead of baryons/antibaryons, we get

$$\frac{n_{e^-} - n_{e^+}}{n_{e^-} + n_{e^+}} \sim \frac{\mu_e}{T} \sim 10^{-10} \tag{2.37}$$

where μ_e is the (positive) chemical potential of electrons.

Neutrinos. 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 they 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 $\nu_e + \bar{\nu}_e \longleftrightarrow e^- + e^+ \longleftrightarrow \gamma$,

and the same in the muonic and tau sectors). The asymmetry could be large if $\mu \geq T$ or negligible if $\mu \ll T$. Some cosmological observations that we will not discuss here prove that the asymmetry is actually very small. We will even assume $\mu_e = \mu_\mu = \mu_\tau = 0$ for simplicity. Thus all six neutrino species share exactly the same number density.

Bosons. Finally, the universe should contain ultra-relativistic photons. Other gauge bosons are non-relativistic (Z^0, $W^\pm$) or must remain confined inside baryons (gluons).

Beyond the standard model. There are lots of motivations for the existence of new symmetries and particles beyond the standard model. It is usually assumed that these particles have masses bigger than 10 MeV and have decayed by that time, excepted the lightest one carrying a new conserved number, which plays the role of a "cold dark matter particle" (see section 1.3). We will also see in that chapter that cold dark matter particles probably decoupled early from the thermal bath, so for $T \leq 10$ MeV we can view them as completely isolated/decoupled species. It is also possible that the new sector brings extra relativistic particles called "dark radiation" that we shall briefly discuss later (but there is no evidence for such particles from current data).

In summary, around $T \sim 10$ MeV, the universe should contain: p, n, e^-, e^+, six neutrino species, γ and non-relativistic Dark Matter particles. The number of relativistic effective degrees of freedom (relevant for computing the total energy density ρ and the entropy density s) is given by photons, electrons, positrons and six neutrinos:

$$g_*(\sim 10\text{MeV}) = 2 + \frac{7}{8}(2+2+6) = 10.75 \ . \tag{2.38}$$

This assumes no Dark Radiation. If there is also a Dark Radiation component, this number gets increased. Very often, Dark Radiation is parametrised through an "effective neutrino number" $N_{\rm eff}$. If $N_{\rm eff} = 3$ it means that we assume no extra Dark Radiation beyond the ordinary three neutrino families. $N_{\rm eff} = 4$ would mean that the Dark radiation would have the same energy density as one neutrino family. Then the number of relativistic degrees of freedom reads

$$g_*(\sim 10\text{MeV}) = 2 + \frac{7}{8}(4 + 2N_{\rm eff}) \ . \tag{2.39}$$

Neutron to proton ratio at high energy. We can actually prove that $n_n \simeq n_p$ at this temperature by using the fact that the reaction of proton-to-neutron conversion:

$$p + e^- \longleftrightarrow n + \nu_e \tag{2.40}$$

is in chemical equilibrium at high energy ($T >$ MeV), implying $\mu_p + \mu_e = \mu_n + \mu_{\nu_e}$. Using eq. (2.15) for the number density of non-relativistic species, we get:

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{3/2} \exp\left(\frac{m_p - m_n}{T} + \frac{\mu_n - \mu_p}{T}\right) \tag{2.41}$$

$$= \left(\frac{m_n}{m_p}\right)^{3/2} \exp\left(\frac{m_p - m_n}{T} + \frac{\mu_e - \mu_{\nu_e}}{T}\right) \tag{2.42}$$

Let us discuss the size of the terms in the argument of the exponential. The difference between the neutron and proton mass is $Q \equiv m_n - m_p = 1.293$ MeV. Hence, for $T \sim 10$ MeV, the ratio Q/T is close to 0.1 and $e^{-Q/T} \sim \mathcal{O}(1)$. Then we have seen that $\frac{\mu_e}{T} \sim 10^{-10}$, so clearly $e^{\mu_e/T} \simeq 1$. Also, we have just explained that the neutrino asymmetry has been proved to be negligible in the standard cosmological model, so $e^{\mu_{\nu_e}/T} \simeq 1$. Finally, the neutron to proton ratio at $T \geq Q$ is given by:

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{3/2} \exp\left(-\frac{Q}{T}\right) \sim \mathcal{O}(1), \tag{2.43}$$

i.e. the density of neutrons and protons is essentially the same.

2.2.3 Neutrino decoupling

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

$$\begin{aligned} \nu_e + e^- &\longleftrightarrow \nu_e + e^- \qquad (2.44)\\ \nu_e + \nu_i &\longleftrightarrow \nu_e + \nu_i \\ \nu_e + \bar{\nu}_i &\longleftrightarrow \nu_e + \bar{\nu}_i \\ &\text{etc.} \qquad (2.45) \end{aligned}$$

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

$$\frac{\Gamma}{H} \sim M_P G_F^2 T^3 \sim \left(\frac{T}{1\text{ MeV}}\right)^3 . \tag{2.46}$$

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

$$f_i(p) = \frac{1}{(2\pi)^3} \frac{1}{\exp[p/T_\nu] + 1} . \tag{2.47}$$

By "frozen", one means that f_i varies only due to the universe expansion, which imposes a very trivial evolution. Each decoupled particle is 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:

$$n_{\nu_i} = \frac{3}{4}\frac{\zeta(3)}{\pi^2}T_\nu^3 \propto a^{-3} , \quad (2.48)$$

$$\rho_{\nu_i} = \frac{7}{8}\frac{\pi^2}{30}T_\nu^4 \propto a^{-4} , \quad (2.49)$$

$$p_{\nu_i} = \frac{1}{3}\rho_{\nu_i} , \quad (2.50)$$

$$s_{\nu i} = \frac{4}{3}\frac{\rho_{\nu_i}}{T_\nu} = \frac{4}{3}\frac{7}{8}\frac{\pi^2}{30}T_\nu^3 \propto a^{-3} . \quad (2.51)$$

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:

$$s = \left.\frac{\rho+p}{T}\right|_{\rm plasma} = \frac{4}{3}\frac{\pi^2}{30}g_* T^3 \qquad \text{with} \quad g_* = 2+\frac{7}{8}(2+2+6) = 10.75 . \quad (2.52)$$

Note that in this section and in the next one, for simplicity, we assume the standard cosmological model, in which there is no Dark Radiation (so $N_{\rm eff} = 3$). 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 like the entropy of relativistic species in equilibrium, with the appropriate value of the temperature:

$$s = \left.\frac{\rho+p}{T}\right|_{\rm plasma} + \left.\frac{\rho_\nu+p_\nu}{T_\nu}\right|_{\rm neutrinos} \quad (2.53)$$

$$= \frac{4}{3}\frac{\pi^2}{30}\left(2+\frac{7}{8}(2+2)\right)T^3 + \frac{4}{3}\frac{\pi^2}{30}\left(\frac{7}{8}\times 6\right)T_\nu^3 \quad (2.54)$$

$$= \frac{4}{3}\frac{\pi^2}{30}\left(\frac{11}{2}\right)T^3 + \frac{4}{3}\frac{\pi^2}{30}\left(\frac{7}{8}\times 6\right)T_\nu^3 . \quad (2.55)$$

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.

2.2.4 Electron-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 regenerated, 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 go into other

species, namely: photons, which are the only remaining relativistic species in the plasma. In other words, the reaction $e^- + e^+ \longrightarrow n\gamma$ 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. (2.55). After annihilation, it reads:

$$s = \left.\frac{\rho+p}{T}\right|_{\rm plasma} + \left.\frac{\rho_\nu+p_\nu}{T_\nu}\right|_{\rm neutrinos} \tag{2.56}$$

$$= \frac{4}{3}\frac{\pi^2}{30}(2)\,T^3 + \frac{4}{3}\frac{\pi^2}{30}\left(\frac{7}{8}\times 6\right)T_\nu^3\ . \tag{2.57}$$

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 $s_{\rm plasma}a^3$ is also conserved separately. Hence:

$$\frac{11}{2}(Ta)^3_{\rm before} = 2(Ta)^3_{\rm after}\ . \tag{2.58}$$

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:

$$\frac{(T_\nu a)_{\rm after}}{(Ta)_{\rm after}} = \frac{(T_\nu a)_{\rm before}}{(11/4)^{1/3}(Ta)_{\rm before}} = \left(\frac{4}{11}\right)^{1/3}\ . \tag{2.59}$$

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 densities:

$$n_\gamma^0 = \frac{\zeta(3)}{\pi^2}\times 2\ T_0^3 = 411\ {\rm cm}^{-3}\ , \tag{2.60}$$

$$n_\nu^0 = \frac{\zeta(3)}{\pi^2}\times\frac{3}{4}\times 6\times\frac{4}{11}\,T_0^3 = 336\ {\rm cm}^{-3}\ , \tag{2.61}$$

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

We now have a complete understanding of what is called the "radiation density" $\rho_{\rm r}$ in the standard cosmological model, at least for temperatures below $T \sim 10$ MeV. It is the total density of photons, neutrinos, electrons and positrons until $T \sim 0.5$ MeV, and later, it is the total density of just photons

and neutrinos. After positron annihilation, we get exactly:

$$\begin{aligned} \rho_{\rm r} &= \frac{\pi^2}{30}(2)\,T^4 + \frac{\pi^2}{30}\left(\frac{7}{8}\times 6\right) T_\nu^4 \\ &= \frac{\pi^2}{30}\left(2 + \frac{7}{8}\times 6\left(\frac{4}{11}\right)^{4/3}\right) T^4 \ . \qquad (2.62) \\ \Rightarrow \omega_{\rm r} &= \frac{\rho_{\rm r0}h^2}{\rho_{\rm cr0}} = 4.183\times 10^{-5} \ . \qquad (2.63) \end{aligned}$$

Since Ta is exactly constant after positron annihilation, the above relation is compatible with the fact that $\rho_{\rm r}$ scales like a^{-4} for relativistic particles. As explained previously, in a non-minimal cosmological model with some Dark Radiation, the relativistic density could be enhaced, and this is usually parametrised with an effective neutrino number $N_{\rm eff}$ such that:

$$\rho_{\rm r} = \frac{\pi^2}{30}\left(2 + \frac{7}{8}\times 2\,N_{\rm eff}\left(\frac{4}{11}\right)^{4/3}\right) T^4 \ . \qquad (2.64)$$

In the minimal model $N_{\rm eff}$ is equal to three[7]. $N_{\rm eff} = 4$ would mean that there is one extra species of relativistic fermions in the universe, with the same density as a pair $\nu_e + \bar{\nu}_e$.

2.2.5 Nucleosynthesis

A nucleus X containing Z protons can have various isotopes ${}^A X$ 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):

$$\begin{aligned} p + n &\longrightarrow D + \gamma \qquad (2.65) \\ D + D &\longrightarrow {}^3He + n \qquad (2.66) \\ {}^3He + D &\longrightarrow {}^4He + p \qquad (2.67) \\ &\ldots \qquad (2.68) \end{aligned}$$

In order to know whether these reactions are favoured 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:

$$E_0(X) = m_X = Zm_p + (A-Z)m_n - B \ . \qquad (2.69)$$

[7] *More advanced considerations (optional): our calculation of (T_ν/T) after positron annihilation assumed that all electron-positron pairs annihilated into photons, and none into neutrinos. In fact, at positron annihilation, we are not far from neutrino decoupling. A minority of neutrinos still has a momentum larger than 1 MeV, and is still partially in thermal equilibrium. Hence the energy released by electron-positron annihilation, after going first into photons, will quickly be shared by a small fraction of the neutrinos, the most energetic ones. This means that there are small effects that we did not take into account. If we had, we would have found also small corrections to equation (2.62). These corrections can be absorbed by the factor $N_{\rm eff}$. As a result, although the number of neutrino species is exactly three, the density of radiation in the standard model is in fact given by equation (2.64) with $N_{\rm eff} = 3.045$ instead of 3.*

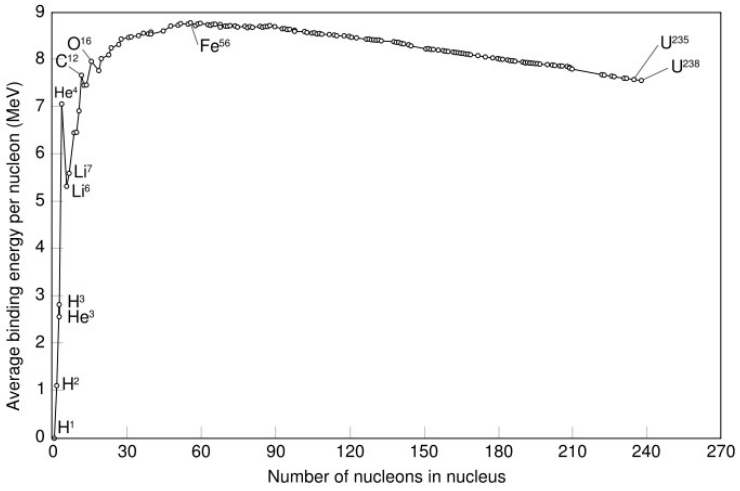

Figure 2.1: Average binding energy per nucleon B/A as a function of A.

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 2.1 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

$$D + \gamma \longrightarrow p + n \tag{2.70}$$

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 $p + n$ (photo-dissociation of Deuterium). This means that for $T > B_D$ the reaction can take place frequently in both directions:

$$D + \gamma \longleftrightarrow p + n \tag{2.71}$$

and the relative abundance of the non-relativistic species D, p, n can be inferred from the chemical equilibrium relation. We will study this and find that chemical equilibrium imposes $n_D \ll n_n \sim n_p$, at least between the quark-hadron transition and the energy scale $T \sim B_D$.

But when the temperature drops below B_D, we could expect that since the average photons are not energetic enough to produce photo-dissociation, the reaction only goes in the direction $p + n \longrightarrow D + \gamma$, and protons and neutrons get progressively converted into deuterium. Then, once there is a significant abundance of deuterium, it is energetically more favorable to convert D into

^3He, and so on and so forth, 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 (2.70) 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 then suppressed by consideration on number densities. Once deuterium is formed, all number densities are small and all two-body weak interactions have a rate Γ that is not much bigger than H. They are still relatively efficient, and most of Deuterium it is converted into ^4He as could be expected from energetics. But three-body weak interactions have a rate that is not smaller than H. This observation is crucial, because 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 out in two steps.

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

$$n + p \longleftrightarrow D + \gamma \ . \tag{2.72}$$

The cross-section of this reaction is large enough for ensuring chemical equilibrium in the temperature range considered here. Hence $\mu_D = \mu_n + \mu_p$. We are only considering times after the quark-hardon transition, such that $T \ll$ GeV, which implies that neutrons, protons and deuterium are all non-relativistic with densities given by eq. (2.15). Hence

$$\frac{n_D}{n_p n_n} = \exp\left(\frac{\mu_D - \mu_p - \mu_n}{T}\right) \ \frac{3}{4}\left(\frac{2\pi m_D}{m_p m_n T}\right)^{3/2} \exp\left(\frac{m_p + m_n - m_D}{T}\right) \ , \tag{2.73}$$

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:

$$\frac{n_D}{n_p n_n} = \frac{3}{4}\left(\frac{2\pi m_D}{m_p m_n T}\right)^{3/2} \exp\left(\frac{B_D}{T}\right) \ . \tag{2.74}$$

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

$$\begin{aligned} \frac{n_D}{n_p} &\sim \frac{n_n}{T^3}\left(\frac{T}{m_p}\right)^{3/2} \exp\left(\frac{B_D}{T}\right) \\ &\sim \frac{n_B}{T^3}\left(\frac{T}{m_p}\right)^{3/2} \exp\left(\frac{B_D}{T}\right) \\ &\sim 10^{-10}\left(\frac{T}{1\ \mathrm{GeV}}\right)^{3/2} \exp\left(\frac{2.22\ \mathrm{MeV}}{T}\right) . \end{aligned} \tag{2.75}$$

So, between the QCD transition ($T \sim 100$ MeV) and the time at which $T \sim B_D$, it is clear that the ratio remains tiny ($n_D/m_P \sim 10^{-10}$). Thus, as expected, there is no significant deuterium abundance above this 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 the deuterium abundance becomes comparable to $n_B \sim n_p$ 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 neutrons. 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:

$$p + e^- \longleftrightarrow n + \nu_e \ . \tag{2.76}$$

At high energy ($T >$ MeV), this reaction is in chemical equilibrium, with $\mu_p + \mu_e = \mu_n + \mu_{\nu_e}$. Using eq. (2.15) for the number density of non-relativistic species, we get:

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{3/2} \exp\left(\frac{m_p - m_n}{T} + \frac{\mu_n - \mu_p}{T}\right) \tag{2.77}$$

$$= \left(\frac{m_n}{m_p}\right)^{3/2} \exp\left(\frac{m_p - m_n}{T} + \frac{\mu_e - \mu_{\nu_e}}{T}\right) \tag{2.78}$$

Let us see which of the terms in the argument of the exponential are really important. The difference between the neutron and proton mass is $Q \equiv m_n - m_p = 1.293$ MeV. Hence, for $T \gg 1$ MeV, the ratio Q/T becomes larger than one and will lead to an exponential drop of the ratio n_n/n_p. The other two terms are instead unimportant because:

- in the simplest cosmological model favoured by observations, the neutrino-anti-neutrino asymmetry is either zero or negligible. Since the asymmetry $(n_{\nu_e} - n_{\bar{\nu}_e})/(n_{\nu_e} + n_{\bar{\nu}_e})$ is of the order of μ_{ν_e}/T, the ratio μ_{ν_e}/T is negligible with respect to one, and can be dropped from the equation.

- we have already seen that electric neutrality implies a very small electron-positron asymmetry before electron-positron annihilation,

$$(n_{e^-} - n_{e^+})/(n_{e^-} + n_{e^+}) \sim \frac{n_B}{s} \sim 10^{-10},$$

so μ_e/T is also of the order of 10^{-10} and can be dropped from the equation[8].

Finally, the neutron to proton ratio is given by:

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{3/2} \exp\left(-\frac{Q}{T}\right) = 1.002 \exp\left(-\frac{1.293\,\text{MeV}}{T}\right) \tag{2.79}$$

For $T \gg 1$ MeV, the neutron to proton ratio is given by:

$$\left.\frac{n_n}{n_p}\right|_{T\gg 1\ \text{MeV}} = \left(\frac{m_n}{m_p}\right)^{3/2} = 1.002\ , \tag{2.80}$$

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:

$$\dot{n}_n + 3Hn_n = n_n[n_{\nu_e}\langle\sigma v\rangle]\left\{\exp\left(\frac{\mu_e + \mu_p - \mu_n - \mu_{\nu_e}}{T}\right) - 1\right\}\ . \tag{2.81}$$

The term between square brackets is the scattering rate Γ_{np} for neutron to proton conversion. In the exponential, we can again neglect the tiny μ_e and μ_{ν_e} chemical potentials, and replace $\exp\left((\mu_p - \mu_n)/T\right)$ using eq. (2.77) (but not assuming chemical equilibrium, and therefore, not assuming that eq. (2.78) is also true). Then

$$\dot{n}_n + 3Hn_n = n_n\Gamma_{np}\left\{\frac{n_p}{n_n}\left(\frac{m_n}{m_p}\right)^{3/2} e^{-Q/T} - 1\right\}\ . \tag{2.82}$$

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

$$n_n = X_n(n_n + n_p) = X_n n_B, \qquad n_p = (1 - X_n)n_B\ . \tag{2.83}$$

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

$$\dot{n}_n = \dot{X}_n n_B - 3HX_n n_B\ . \tag{2.84}$$

[8]To be precise, we have explained here that $\mu_e/T \sim 10^{-10}$ before electron-positron annihilation. After the annihilation, at $T \leq 0.5$MeV, μ_e/T starts to grow as a consequence of charge and electron number conservation, but it starts from such small values that for the range of temperatures interesting here, it remains much smaller than one

Replacing n_n and n_p in eq. (2.82) and dividing by n_B, we get

$$\dot{X}_n = \Gamma_{np} \left[(1 - X_n)\, e^{-Q/T} - X_n \right] \ . \tag{2.85}$$

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,

$$\frac{dT}{dt} = -T\,\frac{da}{a\,dt} = -TH \tag{2.86}$$

$$= -\sqrt{\frac{8\pi G}{3}\rho T^2} \tag{2.87}$$

$$= -\sqrt{\frac{8\pi^3 G}{90} g_* T^6} \tag{2.88}$$

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

$$\frac{dX_n}{dT} = -\sqrt{\frac{90}{8\pi^3 g_*}}\frac{M_P}{T^3}\Gamma_{np}(T) \left[(1 - X_n) e^{-Q/T} - X_n \right] \ . \tag{2.89}$$

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 (2.89) 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-to-proton conversion ($n \longrightarrow p + e^- + \bar{\nu}_e$) 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, $X_n(T < 0.1 \text{ MeV}) \sim 0.11$, while at the time of Deuterium creation, $n_n = 0.124\, n_B$ and $n_p = 0.876\, n_B$. Then, all available neutrons will combine into deuterium, ^{3}He and finally ^{4}He nuclei, together with the same number of protons. The final ^{4}He density should be $n_{^4\text{He}} = 0.062\, n_B$, with a leftover of $n_H = 0.752\, n_B$ protons. The helium fraction, usually defined as:

$$Y_P \equiv \frac{4 n_{^4He}}{n_B} \ , \tag{2.90}$$

is predicted to be 0.248 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 simulation 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). Table 2.1 shows, for instance, the first 40 reactions used in the public code PARTHENOPE[9]. In section 2.2.5, we only studied the reactions called 1 and 12 in this table.

The main differences between the outcome of a full simulation and the results of the above section are:

[9] `http://parthenope.na.infn.it/`

No.	Reaction	Type	No.	Reaction	Type
1	$n \longrightarrow p$	*weak*	22	$^6Li + p \longrightarrow \gamma + {}^7Be$	(p,γ)
2	$^3H \rightarrow \bar{\nu}_e + e^- + {}^3He$	*weak*	23	$^6Li + p \longrightarrow {}^3He + {}^4He$	^{3}He Pickup
3	$^8Li \rightarrow \bar{\nu}_e + e^- + 2\,{}^4He$	*weak*	24	$^7Li + p \longrightarrow {}^4He + {}^4He$	^{4}He Pickup
4	$^{12}B \rightarrow \bar{\nu}_e + e^- + {}^{12}C$	*weak*	24 bis	$^7Li + p \longrightarrow \gamma + {}^4He + {}^4He$	(p,γ)
5	$^{14}C \rightarrow \bar{\nu}_e + e^- + {}^{14}N$	*weak*	25	$^4He + {}^2H \longrightarrow \gamma + {}^6Li$	(d,γ)
6	$^8B \rightarrow \nu_e + e^+ + 2\,{}^4He$	*weak*	26	$^4He + {}^3H \longrightarrow \gamma + {}^7Li$	(t,γ)
7	$^{11}C \rightarrow \nu_e + e^+ + {}^{11}B$	*weak*	27	$^4He + {}^3He \longrightarrow \gamma + {}^7Be$	(^{3}He,γ)
8	$^{12}N \rightarrow \nu_e + e^+ + {}^{12}C$	*weak*	28	$^2H + {}^2H \longrightarrow n + {}^3He$	^{2}H Strip.
9	$^{13}N \rightarrow \nu_e + e^+ + {}^{13}C$	*weak*	29	$^2H + {}^2H \longrightarrow p + {}^3H$	^{2}H Strip.
10	$^{14}O \rightarrow \nu_e + e^+ + {}^{14}N$	*weak*	30	$^3H + {}^2H \longrightarrow n + {}^4He$	^{2}H Strip.
11	$^{15}O \rightarrow \nu_e + e^+ + {}^{15}N$	*weak*	31	$^3He + {}^2H \longrightarrow p + {}^4He$	^{2}H Strip.
12	$p + n \longrightarrow \gamma + {}^2H$	(n,γ)	32	$^3He + {}^3He \longrightarrow p + p + {}^4He$	(^{3}He,2p)
13	$^2H + n \longrightarrow \gamma + {}^3H$	(n,γ)	33	$^7Li + {}^2H \longrightarrow n + {}^4He + {}^4He$	(d,n α)
14	$^3He + n \longrightarrow \gamma + {}^4He$	(n,γ)	34	$^7Be + {}^2H \longrightarrow p + {}^4He + {}^4He$	(d,p α)
15	$^6Li + n \longrightarrow \gamma + {}^7Li$	(n,γ)	35	$^3He + {}^3H \longrightarrow \gamma + {}^6Li$	(t,γ)
16	$^3He + n \longrightarrow p + {}^3H$	charge ex.	36	$^6Li + {}^2H \longrightarrow n + {}^7Be$	^{2}H Strip.
17	$^7Be + n \longrightarrow p + {}^7Li$	charge ex.	37	$^6Li + {}^2H \longrightarrow p + {}^7Li$	^{2}H Strip.
18	$^6Li + n \longrightarrow {}^3H + {}^4He$	^{3}H Pickup	38	$^3He + {}^3H \longrightarrow {}^2H + {}^4He$	(^{3}H,d)
19	$^7Be + n \longrightarrow {}^4He + {}^4He$	^{4}He Pickup	39	$^3H + {}^3H \longrightarrow n + n + {}^4He$	(t,n n)
20	$^2H + p \longrightarrow \gamma + {}^3He$	(p,γ)	40	$^3He + {}^3H \longrightarrow p + n + {}^4He$	(t,n p)
21	$^3H + p \longrightarrow \gamma + {}^4He$	(p,γ)			

Table 2.1: The first forty reactions used in the Nucleosynthesis code PARTENOPE. *Table taken from [arXiv:0705.0290] by Ofelia Pisanti et al.*

- when reactions freeze-out, the density n_i of other elements than ^{4}He is nonzero – but still very small: the number density of D and ^{3}He is smaller than that of ^{4}He 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 predictions for Y_P and other light element abundances are not fixed numbers, they depend slightly on the two free parameters of this problem, namely n_B/s and $N_{\rm eff}$:

  1. the value of $n_B/s \sim 10^{-10}$ controls mainly the temperature at which deuterium starts forming (see eq. (2.75)). Hence the neutron-to-proton ratio at the beginning of deuterium formation depends on n_B/s, as well as the final helium abundance. The ratio n_B/s is easy to relate today to $(n_p+n_n)/n_\gamma$, and for fixed CMB temperature, this ratio can finally be expressed as a function of ω_b.

  2. the value of g_*, which enters the change of variable from time to temperature, $dT/dt \propto g_*^{1/2}$. In models with Dark Radiation,

$$g_*(\sim 10\text{MeV}) = 2 + \frac{7}{8}\left(4 + 2N_{\rm eff}\right) \ , \tag{2.91}$$

  as defined in equation (2.39). Thus the speed of neutron-to-proton conversion depends on the effective neutrino number $N_{\rm eff}$, and the ratio n_n/n_p at $T = 0.07$ MeV is slightly different, leading to different abundances.

Observed abundance of primordial elements In sections 2.2.5, 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 $\rho \sim (0.07\,\text{MeV})^4$. 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.

Numerical simulation of Nucleosynthesis accurately predict all relative abundances as a function of the two free parameters in the theory, $n_B/s \propto \omega_b$ and $N_{\rm eff}$. The primordial abundances that are easiest to measure are the primordial helium abundance $Y_P \equiv 4n_{^4\rm He}/h_{\rm H}$ and the primordial deuterium abundance $y_D = n_D/n_H$ (according to our previous study, the latter should vanish, but in reality, when all weak reactions freeze out, there is a tiny left-over of Deuterium such that $y_D \sim 10^{-5}$). So it is possible to measure two observables and determine the value of the two free parameters of the model.

The current observations give $Y_P = 0.246 \pm 0.010$ and $y_D = (2.53 \pm 0.04)10^{-5}$ (68% CL). By fitting this data with the BBN model, one gets a measurement of both ω_b and $N_{\rm eff}$:

$$\omega_b = 0.023 \pm 0.002 \ , \tag{2.92}$$

$$N_{\rm eff} = 3 \pm 1 \ . \tag{2.93}$$

This confirms that we have $\frac{n_B}{s} \sim 10^{-10}$, as assumed throughout this chapter, and that radiation could consist only of photons and ordinary neutrinos.

We will see later that CMB observations give independent measurements of (ω_b, $N_{\rm eff}$) which are in very good agreement with the above results (but more precise). The latest CMB observations from the Planck satellite indicate (always at the 68%CL, i.e. 1-σ confidence level)

$$\omega_b = 0.02226 \pm 0.00016, \tag{2.94}$$

$$N_{\rm eff} = 2.99 \pm 0.20. \tag{2.95}$$

It is considered as a huge success for cosmology that two very different techniques (Nucleosynthesis, which relies on nuclear physics when $T \sim (0.01-1)$ MeV, and CMB, which relies on relativistic hydrodynamics and QED when $T \sim (0.1-100)$ eV) give compatible results for ω_b and $N_{\rm eff}$.

Hence, for $h = 0.67$ (the current best-fit value), the baryon fraction is of the order of $\Omega_b \sim 0.05$: approximately five percent of the universe density comes from ordinary matter. This is already more than the sum of all luminous matter, which represents only one per cent: so, 80% of ordinary matter is not even visible.

2.2.6 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 (and probably decoupled dark matter particles studied separately in section). At $T \ll$ MeV, weak interactions are inefficient, but electromagnetic interactions ensure thermal 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 $e^- + {}^4\mathrm{He}^{++} \longrightarrow e^- + {}^4\mathrm{He}^{++}$). These interactions are efficient at least as long as hydrogen and helium remain ionized.

The formation of neutral hydrogen depends on the reaction:

$$e^- + p \longleftrightarrow H + \gamma \ . \tag{2.96}$$

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. Here we will neglect this issue and do as if hydrogen could only be in its fundamental state.

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

$$m_H = m_p + m_e - \epsilon_0 \ , \tag{2.97}$$

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} times smaller than that of photons. So, much below $T \sim 13.6$ eV, there should still be enough energetic photons for preventing recombination.

In one of the exercise sheets, you will find that this expectation is confirmed by the actual equations. You will define the hydrogen ionization fraction:

$$X_e \equiv \frac{n_e}{n_e + n_H} = \frac{n_p}{n_p + n_H} \ , \tag{2.98}$$

where n_e stands for the density of free electrons only, and n_p of free protons only. Assuming thermal equilibrium, you will derive the Saha equation

$$\frac{X_e^2}{1 - X_e} = \frac{1}{n_e + n_H} \left(\frac{m_e T}{2\pi} \right)^{3/2} e^{-\epsilon_0 / T} \ , \tag{2.99}$$

which gives an approximation of the temperature of recombination, found to be close to $T_{\rm rec} \sim 0.254$ eV, which happens at $z_{\rm rec} \sim 1080$ (see exercises). This redshift is called the redshift of recombination. A very important remark is that it is smaller than the redshift of equality, $z_{\rm dec} \sim 3000$. Thus, recombination takes place during Matter Domination (MD).

Below this temperature, the reaction leaves thermal equilibrium, implying that the ionisation fraction freezes out. In the exercises, will write the Boltzmann equation governing the evolution of X_e. By integrating this equation, one would get a confirmation that the ionization fraction X_e becomes significantly smaller than one around $z_{\rm rec} \sim 1080$, and tends to an asymptotic freeze-out value of order $X_e \to 5 \times 10^{-4}$ for $z < 100$.

The same kind of analysis can be done for the formation of neutral Helium, which takes place in two stages,

$$e^- + {}^4He^{++} \quad \longleftrightarrow \quad {}^4He^+ + \gamma \ , \tag{2.100}$$

$$e^- + {}^4He^{+} \quad \longleftrightarrow \quad {}^4He + \gamma \ . \tag{2.101}$$

On can show that the first and second recombination of Helium both took place before Hydrogen recombination.

2.2.7 Photon decoupling

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

$$\gamma + e^- \longrightarrow \gamma + e^- \ . \tag{2.102}$$

The cross section of the above reaction is the Thomson cross section, equal to $\sigma_T = 0.665 \times 10^{-24}\text{cm}^2$. From there, one can compute a thermally averaged cross-section $\langle \sigma_T v \rangle$. Compton scattering of photons off electrons becomes inefficient roughly when the scattering rate $\Gamma = n_e \langle \sigma_T v \rangle$ drops below the Hubble parameter. Note that n_e denotes the number of free electrons, not the total number of electrons (that includes also the electrons inside neutral Hydrogen or Helium atoms). In order to evaluate this characteristic time, we can write

$$n_e = X_e n_e^{\text{total}} = X_e n_p^{\text{total}} \simeq X_e n_B \tag{2.103}$$

and

$$n_B \sim \frac{\rho_b}{m_p} \sim \left(\frac{a_0}{a}\right)^3 \frac{\rho_{b0}}{m_p} \sim \left(\frac{a_0}{a}\right)^3 \frac{10^{-29}\omega_b \text{ g.cm}^{-3}}{1 \text{ GeV}} \ . \tag{2.104}$$

Given the value of the Thomson scattering cross-section $\langle \sigma_T v \rangle$, one finds

$$\Gamma \sim 2 \times 10^{-21} \ X_e \ \left(\frac{a_0}{a}\right)^3 \omega_b \ \text{s}^{-1} \ , \tag{2.105}$$

and then

$$\frac{\Gamma}{H} \sim \frac{2 \times 10^{-21} \ X_e \ \left(\frac{a_0}{a}\right)^3 \omega_b \ \text{s}^{-1}}{H_0} \ \frac{H_0}{H} \ . \tag{2.106}$$

Next, one case use the following relation during matter and Λ domination:

$$\left(\frac{H}{H_0}\right)^2 = \Omega_{\text{m}}(1+z)^3 + \Omega_\Lambda \ . \tag{2.107}$$

Thus, during MD, one has $H/H_0 \simeq \Omega_{\text{m}}^{1/2}(a_0/a)^{3/2}$, and

$$\begin{aligned}
\frac{\Gamma}{H} &\sim \frac{2 \times 10^{-21} \ X_e \ \left(\frac{a_0}{a}\right)^3 \omega_b \ \text{s}^{-1}}{H_0 \, \Omega_{\text{m}}^{1/2} \left(\frac{a_0}{a}\right)^{3/2}} \\
&\sim \frac{2 \times 10^{-21} \ X_e \ \omega_b \ \text{s}^{-1}}{[100\, h \, \text{km}\, \text{s}^{-1}\text{Mpc}^{-1}] \ \Omega_{\text{m}}^{1/2}} \ \left(\frac{a_0}{a}\right)^{3/2} \\
&\sim \frac{2 \times 10^{-21} \ X_e \ \omega_b \ \text{s}^{-1}}{[3 \times 10^{-18}\text{s}^{-1}] \ \omega_{\text{m}}^{1/2}} \ \left(\frac{a_0}{a}\right)^{3/2} \\
&\sim 0.007 \ \frac{\omega_b}{\omega_{\text{m}}^{1/2}} \ X_e \ \left(\frac{a_0}{a}\right)^{3/2}
\end{aligned} \tag{2.108}$$

The evolution of this ratio comes from the product $X_e \left(\frac{a_0}{a}\right)^{3/2}$, where we recall that X_e is equal to one before recombination, then drops abruptly (exponentially) around z_{dec}, and then stabilises around 5×10^{-4}. A numerical application shows that $\Gamma \sim H$ occurs precisely at recombination, during the exponential drop of X_e. Thus the redshift of photon decoupling is $z_{\text{dec}} \sim z_{\text{rec}} \sim 1080$.

Hence, recombination directly triggers photon decoupling. This is in fact the main reason for which recombination is important to study: it controls the time at which the universe becomes transparent, that is, the decoupling of the CMB photons that we observe today.

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 $p \propto a^{-1}$ for individual photons, and hence $T \propto a^{-1}$ in the photon phase-space distribution.

Translating the redshift $z_{\text{dec}} \sim z_{\text{rec}} \sim 1080$ in terms of proper time, one finds photons decouple approximately 380 000 years after the Planck time.

2.2.8 Very recent stages

After photon decoupling, there are no more events in the thermal history of the universe that play a significant role in this course. Let us mention however a couple of interesting phenomena.

Neutrinos have small masses, all in the range $0 < m_i < 0.1\,\mathrm{eV}$. 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 energy density and pressure of neutrinos are affected by the non-relativistic transition. The consequences of this transition on the formation of structures (clusters, galaxies) are interesting, but not discussed in this course.

There is another remarkable phenomenon occurring at low redshift. When the first stars form around $z \sim 6$, they emit a new population of photons that 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. This can be understood from eq. (2.108): when X_e goes back to one at small redshifts $z \sim 10$, the ratio $(a/a_0)^{3/2}$ is much smaller than at the time of recombination (by a factor $(1000/6)^{3/2}$), so Γ/H remains smaller than one.

In figure 2.2, we summarise qualitatively the main results of this section.

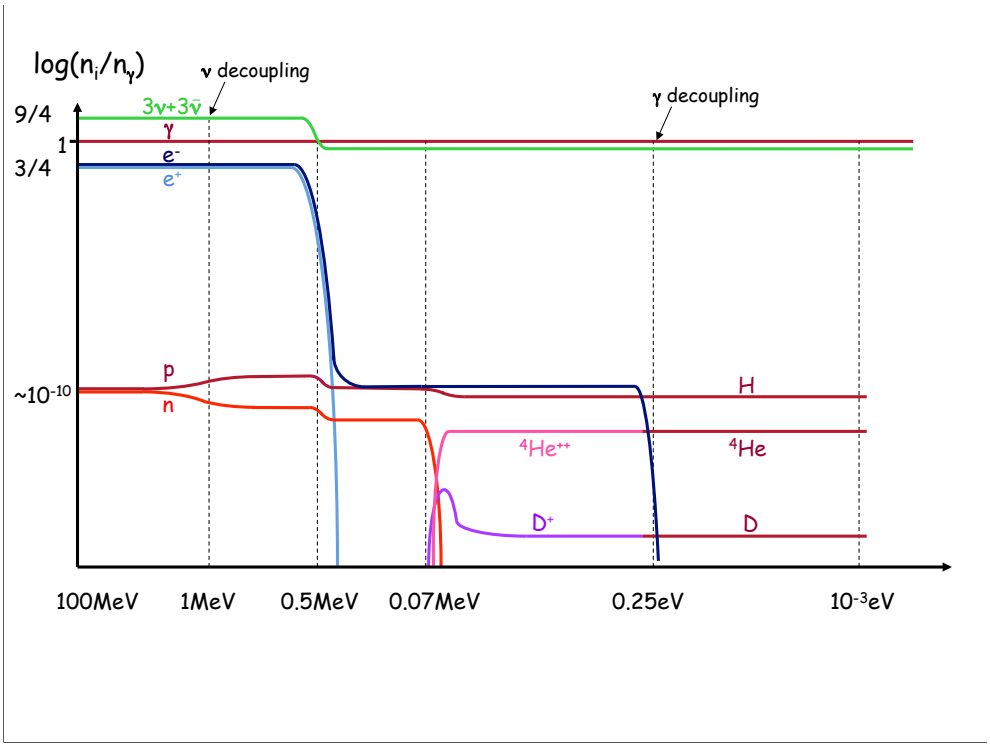

Figure 2.2: As a summary of this chapter, we show the qualitative evolution of n_i for each species, normalized in terms of n_γ.

2.3 Dark Matter

2.3.1 Historical arguments

There are many strong 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: $\rho_{\rm lum}(r) = b\ I(r)$, where b is an unknown coefficient of proportionality called the bias. From this, we can compute the gravitational potential $\Phi_{\rm lum}$ generated by the luminous matter, and the corresponding orbital velocity, given by ordinary Newtonian mechanics:

$$\rho_{\rm lum}(r) = b\ I(r), \tag{2.109}$$

$$\Delta\Phi_{\rm lum}(r) = 4\pi G\ \rho_{\rm lum}(r), \tag{2.110}$$

$$v^2_{\rm lum}(r) = r\frac{\partial}{\partial r}\Phi_{\rm lum}(r). \tag{2.111}$$

So, $v_{\rm lum}(r)$ is known up to an arbitrary normalisation factor $\sqrt{b}$. However, for many galaxies, even by varying b, it is impossible to obtain a rough agreement between $v(r)$ and $v_{\rm lum}(r)$ (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.

We can explain the same result in slightly different words. Assuming that stars have a circular orbit (this is just an approximation), the relation between force and accelerations gives us

$$\frac{v^2(r)}{r} = \frac{\partial}{\partial r}\Phi(r) \tag{2.112}$$

while the Poisson equation of newtonian mechanics gives us

$$\Delta\Phi(r) = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\Phi\right) = 4\pi G\rho(r) \tag{2.113}$$

Finally, the mass of objects enclosed in a radius r is just

$$M(r) = 4\pi\int_0^r dr'(r')^2\rho(r')\ . \tag{2.114}$$

These relations and a simple integration give the exact relation

$$v^2(r) = \frac{GM(r)}{r}\ . \tag{2.115}$$

If we assume that all the mass is in the form of visible matter, there is a mismatch between measurements of $v(r)$ and estimates of $M(r)$. In particular, when we see that most of the mass is located within a radius r_v (where v stands for visible), we expect that above r_v, $M(r)$ reaches a constant asymptote. Then $v(r > r_v)$ should decrease nearly like $1/r$ (this is obvious from the last relation, and such

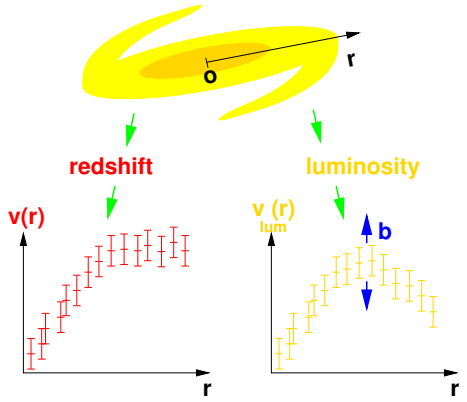

Figure 2.3: A sketchy view of the galaxy rotation curve issue. The genuine orbital velocity of the stars is measured directly from the redshift. From the luminosity distribution, we can reconstruct the orbital velocity under the assumption that all the mass in the galaxy arises form of the observed luminous matter. Even by varying the unknown normalisation 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.

a decrease is called a Keplerian decrease). Instead, in many galaxies, beyond such a radius r_v, the few remaining starts tend to orbit much too fast. An obvious solution is to assume that there is another type of non-visible matter contributing to $M(r)$, and even dominating it. If the non-visible matter is spread over a larger radius than visible matter, then the most distant observable galaxies are still orbiting in the gravitational potential created by dark matter. This supports the notion of a dark matter halo.

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.

2.3.2 Other evidences for dark matter

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.

The observation of *CMB anisotropies* is the strongest one. It requires the presence of a component *not interacting with ordinary electromagnetic forces*. This component had to be present in the universe at least at early times, between $z \sim 10^6$ and $z \sim 10^3$. Assuming that this dark matter is stable and that its number density is conserved until today, we obtain from the CMB an estimate of the dark matter density today corresponding to 25% of the total energy density of the universe (and 85% of the total non-relativistic matter density). Before CMB observations, one could have thought that dark matter is just ordinary matter that we cannot see, because it is not luminous, and it does not absorb the light of other objects. If this was the case, from the point of view of CMB

physics, non-luminous ordinary matter would count as baryonic matter, not as dark matter. Thanks to CMB data, we are now sure that dark matter is truly different from ordinary atomic matter.

Strong lensing. Gravitational lensing is by definition sensitive to the total gravitational potential created by both ordinary matter and dark matter. The study of arclets and of strong lensing patterns allows to reconstruct the shape of dark matter halos around some particular galaxy clusters, and to bring further proofs of the existence of such halos.

Weak lensing. By looking at the statistics of the apparent orientation of galaxies in different regions of the sky, one can average out the random distribution of intrinsic shapes, and estimate the effect of the weak lensing of galaxy images by dark matter (since this effect is coherent over many galaxies in a given region). Weak lensing works surprisingly well and gives us a very good map of the gravitational potential projected along the line-of-sight in each direction around us. By studying the weak lensing of galaxies located at a given redshift, one can even do *tomography* and reconstruct the 3D distribution of dark matter (always with a poorer resolution along the 3rd dimension, i.e. along the line-of-sight). This technique brings further evidence for dark matter, and allows to estimate its abundance. It gives consistent results with other techniques, and in particular with CMB observations.

The analysis of the *bullet cluster* shows very well the presence of two halos that have crossed each other in the recent past without being deformed, unlike the two associated clouds of gas, now displaced from the center of halos, and shaped like a shock wave. This is another way to find that dark matter is very weakly interacting - possibly only gravitationally.

There are other ways to prove the existence and the properties of dark matter that we do not have time to summarise here. However, we should stress one important fact: dark matter must be *cold* rather than *hot*. What do we mean by this?

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. Today it would still possess large velocities, just a few orders of magnitude smaller than the speed of light (this hypothesis is called Hot Dark Matter or HDM). Because of these large velocities, neutrinos could not remain confined in small gravitational potential wells. Even in presence of gravitational clustering, they would form very large and not-so-dense halos (while particles with negligible velocities, called Cold Dark Matter or CDM, can cluster much better: they can form much smaller and much denser halos).

If halos were huge and not very dense like in the HDM case, the number and the distribution of galaxies (which depends on the gravitational effects of dark matter) would be very different from what we observe. For that reason, HDM is strongly excluded by several types of observations. Dark matter particles have to be strongly non-relativistic, otherwise galaxy could not form with a sufficient abundance during matter domination.

2.3.3 Thermal WIMP model

Here we will review only one out of many possible models for explaining the dark matter problem: the case of so-called WIMPs. WIMP means *Weakly Interacting Massive Particle*.

Here *Interacting* refers to the fact that these particles are assumed to have

interactions with standard model particles (otherwise no interesting calculations or predictions could be made...). More specifically, WIMP interactions are supposed to be sufficiently efficient in the early universe for bringing WIMPs in thermal equilibrium with other particles.

Weakly refers to the fact that we assume specifically that these interactions are of the weak type (i.e. mediated by Z and $W^{\pm}$ bosons). Hence we expect that these particles decouple at the latest when the temperature is roughly of the order of the MeV, like for neutrinos; They may decouple even before if they become very diluted and their interaction rate is suppressed. After decoupling, WIMPs interact only gravitationally with other species.

Massive refers to the fact that we impose a sufficient mass for WIMPs to decouple when they are non-relativistic: i.e. we assume a mass $m_\chi \gg O(\text{MeV})$. If we did not make such an assumption, i.e. if the WIMPs decoupled when they are still relativistic, they would share the same number density as neutrinos until today. Then the correct value of the relic density ρ^0_{dm} (or equivalently ω_{dm}) could only be obtained for $m_\chi \sim O(10)$ eV, and the typical velocity of WIMPs today would be of the order of $v \sim \langle p \rangle / m_\chi \sim T/m_\chi \sim 10^{-4}c$. Such velocities are still very large, and this dark matter candidate would fall in the category of Hot Dark Matter. To avoid this, we impose non-relativistic decoupling: we will see that this leads to a number density n^0_χ very suppressed with respect to that of neutrinos, and because the mass is large, to very small dark matter velocities: in that case WIMPs fall in the category of Cold Dark Matter.

For simplicity, WIMPs are usually assumed to be neutral and to be Majorana fermions instead of Dirac fermions, which means that they are their own anti-particle. We will denote them as χ. Moreover, one usually imposes a Z_2 symmetry on these particles. This means that the Lagrangian can only feature even powers of χ. For instance, an interaction term χAB is forbidden by the Z_2 symmetry, while a term $\chi^2 AB$ respects the symmetry. As a consequence, WIMPs cannot decay ($\chi \longrightarrow A + B + ...$) but they can annihilate ($\chi + \chi \longrightarrow A + B + ...$). In the latter reaction, the total charge on the left-hand side is zero. Hence the total charge must be zero also on the right-hand side. It means that pairs of WIMPs can only annihilate into neutral particles, or into pairs of particles and anti-particles. For instance, they can annihilate into higgs bosons, Z^0 bosons, quark-antiquark pairs, lepton-anti-lepton pairs, etc. This is model-dependent. "Popular" annihilation channels are, for instance, electron-positron, or muon-antimuon, but there are many other possibilities. For the calculations in the rest of this section, it is not necessary to specify explicitly what the dominant annihilation channel is.

Our goal is to compute the relic density of WIMPs and see whether it can be related to the WIMP mass and/or annihilation cross-section. First, we will assume that the annihilation reaction

$$\chi + \chi \longrightarrow A(+B) \tag{2.116}$$

remains in chemical equilibrium at every time. In that case, we can write $2\mu_\chi = \mu_A(+\mu_B)$. But since the right-hand side must be a particle carrying no conserved charge (so $\mu_A = 0$) or a particle-antiparticle pair (so $\mu_A = -\mu_B$), this simplifies to $\mu_\chi = 0$. In that case, we know that for $T \gg m_\chi$

$$n_\chi = \frac{\zeta(3)}{\pi^2} \frac{3}{4} g_\chi T^3 \ , \tag{2.117}$$

while for $T \ll m_\chi$

$$n_\chi = g_\chi \left(\frac{m_\chi T}{2\pi}\right)^{3/2} e^{-\frac{m_\chi}{T}} \ . \tag{2.118}$$

In fact, one could integrate the Fermi-Dirac phase-space distribution in order to get n_χ for any value of temperature. The exact relation has the two asymptotes written above. This means that $n_\chi a^3$ evolves with T like: $(aT)^3$ for $T \gg m_\chi$, i.e. like a constant when the effective number of effective degrees of freedom g_* is constant; and like an exponentially decaying function for $T \ll m_\chi$. In the rest of this section, we will call this particular solution $n_\chi^{\text{ch.eq.}}(T)$ (the "density assuming chemical equilibrium").

In reality, we expect that WIMPs will not remain in chemical equilibrium for a long time after the non-relativistic transition. Indeed, the mass of WIMPs is usually assumed to be slightly above the order of the MeV (a typical range for the most popular models is 100 MeV $< m_\chi <$ 10^3 GeV). But we know that weak interactions become inefficient around $T \sim O(\text{MeV})$ for neutrinos, because $\Gamma = n_\nu \langle \sigma_\nu v \rangle$ becomes smaller than H. For WIMPs, we expect a cross section of the same order of magnitude (because it still depends on the Fermi constant), and an annihilation rate $\Gamma = n_\chi \langle \sigma_A v \rangle$ even smaller than for neutrinos, because n_χ gets exponentially suppressed after the non-relativistic transition (here $\langle \sigma_A v \rangle$ is the WIMP thermally averaged annihilation cross section). Hence we expect WIMPs to leave thermal equilibrium even before neutrinos, and very soon after their non-relativistic transition, when n_χ becomes very small.

We know that "leaving chemical equilibrium" means, concretely, that the WIMP number density will freeze out, and that $n_\chi a^3$ will be conserved after freeze-out. The evolution of n_χ is given by the Boltzmann equation

$$\dot{n}_\chi + 3Hn_\chi = n_\chi^2 \langle \sigma_A v \rangle \left[e^{-\frac{2\mu_\chi}{T}} - 1 \right] \ . \tag{2.119}$$

Using our definition of $n_\chi^{\text{ch.eq.}}(T)$, and the fact that for $T \ll m_\chi$ it is given by eq. (2.118), we can write the Boltzmann equation in the following form for any time *after* the WIMP non-relativistic transition:

$$\dot{n}_\chi + 3Hn_\chi = n_\chi^2 \langle \sigma_A v \rangle \left[\left(\frac{n_\chi^{\text{ch.eq.}}}{n_\chi} \right)^2 - 1 \right] = \langle \sigma_A v \rangle \left[\left(n_\chi^{\text{ch.eq.}} \right)^2 - \left(n_\chi \right)^2 \right] \ . \tag{2.120}$$

Since $n_\chi^{\text{ch.eq.}}(T)$ is a known function of temperature , and since the derivative with respect to time can be transformed into a derivative with respect to temperature (like in the BBN and recombination sections), we have all the ingredients for integrating this equation numerically. This is what we would do in order to get a precise prediction for the relic number density of WIMPs after freeze-out (i.e. after leaving chemical equilibrium and reaching the asymptote with a constant value of $n_\chi a^3$).

But for a rough order-of-magnitude estimate, there is a simpler way to estimate the relic density of WIMPs at freeze-out. Let us use a superscript f to denote quantities at that time (for instance, the freeze-out temperature will be T^f). We know that freeze-our occurs roughly when $\Gamma \sim H$. This gives directly:

$$n_\chi^{\text{f}} \sim \frac{H^f}{\langle \sigma_A v \rangle^f} \ . \tag{2.121}$$

(we have put a superscript also on the thermally averaged cross-section, in case this quantity would evolve with time, but this is not the case for typical

annihilation channels). We can estimate H^f as a function of temperature using the Friedmann equation and the density of the thermal bath as a function of the effective number of bosonic relativistic degrees of freedom g_*:

$$(H^f)^2 = \frac{8\pi G}{3}\rho^f = \frac{8\pi G}{3}\frac{\pi^2}{30}g_*^f(T^f)^4 \sim \frac{g_*^f(T^f)^4}{M_P^2} \tag{2.122}$$

Hence

$$n_\chi^{\mathrm{f}} \sim \frac{(g_*^f)^{1/2}(T^f)^2}{M_P\langle\sigma_A v\rangle^f} \tag{2.123}$$

Moreover, we know that after freeze-out, $n_\chi a^3$ is exactly conserved. At the same time the entropy conservation law says that $g_*(Ta)^3$ is constant, so n_χ evolves proportionally to g_*T^3. Finally we can evaluate the WIMP relic density today:

$$n_\chi^0 = n_\chi^{\mathrm{f}}\ \frac{g_*^0}{g_*^f}\ \left(\frac{T^0}{T^f}\right)^3 \ . \tag{2.124}$$

Putting everything together, this gives

$$n_\chi^0 \sim \frac{g_*^0}{(g_*^f)^{1/2}}\frac{(T^0)^3}{T^f}\frac{1}{M_P\langle\sigma_A v\rangle^f} \ . \tag{2.125}$$

Since WIMPs are strongly non-relativistic today, we can infer the energy density by multiplying the number density by the WIMP mass. This gives:

$$\rho_\chi^0 \sim \frac{m_\chi}{T^f}\frac{g_*^0}{(g_*^f)^{1/2}}\frac{(T^0)^3}{M_P\langle\sigma_A v\rangle} \ . \tag{2.126}$$

In a given model motivated by particle physics, we would know g_*^0, g_*^f and $\langle\sigma_A v\rangle$. Then, in order to get a definite prediction, the only remaining task is to evaluate the ratio $\frac{m_\chi}{T^f}$.

We know that $\frac{m_\chi}{T}$ crosses one at the time of the non-relativistic transition. A crude approximation would consist in saying that freeze-out takes place very soon after the non-relativistic transition, because it is triggered by the fast exponential decay of n_χ as a function of $\frac{m_\chi}{T}$ (as indicated by eq. (2.118)). In this approximation we could use $\frac{m_\chi}{T^f} \sim 1$.

This approximation is actually not so bad, because the true value of the freeze-out temperature would be given by solving the equation

$$\frac{\Gamma}{H} = \frac{n_\chi\langle\sigma_A v\rangle}{H} \sim 1 \ , \tag{2.127}$$

with n_χ given by eq. (2.118). Due to the factor $e^{-\frac{m_\chi}{T}}$, the solution for $\frac{m_\chi}{T^f}$ depends *logarithmically* on the parameters of the problem (mass, cross-section): hence it is true that $\frac{m_\chi}{T^f}$ is of order one, and can be considered as nearly independent of the mass and cross-section in first approximation.

We are led to our final result: the relic density of WIMPs is governed by the inverse annihilation cross-section, with

$$\rho_\chi^0 \sim \frac{g_*^0}{(g_*^f)^{1/2}}\frac{(T^0)^3}{M_P\langle\sigma_A v\rangle^f} \ . \tag{2.128}$$

This is a very well-known result. Intuitively, WIMPs with a larger cross-section remain in thermal equilibrium for a longer time after their non-relativistic transition. Hence they are more Boltzmann-suppressed at freeze-out, and a smaller

fraction of them survives until today. By matching ρ_χ^0 to observations, one gets a prediction for $\langle\sigma_A v\rangle^f \sim 10^{-26}\mathrm{cm}^3\mathrm{s}^{-1}$. This is precisely the order of magnitude that one would expect for a particle interacting only with weak forces. In other words, by assuming that $\langle\sigma_A v\rangle$ is a weak annihilation cross-section, we automatically get the right order of magnitude for the DM relic density. This nice coincidence is often called the *WIMP miracle*.

Many experiments of so-called *dark matter direct detection* have been built for probing such particles. They are usually located in underground laboratories, to filter out as many cosmic rays, ordinary electromagnetic radiation and terrestrial radio-activity as possible. Most of them try to measure the small heating of the detector caused by elastic interactions between detector particles and the WIMPs crossing them. This is of course very difficult, due to the very small interaction rate. But at least, people know what to search for: we have seen that we can estimate the WIMP annihilation cross-section, and hence, also the typical WIMP interaction cross sections. These detectors have not found any significant signal so far.

There are many dark matter candidates falling in the category of WIMPs: it is not so difficult to build a reasonable extension of the standard model of particle physics featuring new particles, some of them having the basic properties that we mentioned in this section. A famous examples is the *neutralino* of supersymmetric models. However, not only WIMPs have not been discovered, but the LHC is currently bringing no evidence in favour of supersymmetry.

Search for WIMPs will continue in the next years. In parallel, people are thinking about other types of dark matter candidates (axions, sterile neutrinos, ...), and are working on other types of experiments to probe them. We do not have time to describe these alternative scenarios in this chapter.

Chapter 3

Linearised gravity

The cosmological model offers a coherent and very well experimentally tested framework in which the primordial universe is very homogeneous, with tiny fluctuations in density. Later on, these fluctuations grow by gravitational collapse, until they form the strong inhomogeneities observed today (filaments, vacuums, halos, galaxies, etc.). The best experimental evidence for this paradigm is precisely the observation of CMB anisotropies, which directly shows that the relative density fluctuations at the time of photon decoupling were only of the order of 10^{-5}.

Consequently, the primordial universe is very well described by a theory of linear perturbations. Subsequently, non-linear perturbations appear first at small scales and then gain larger and larger scales. Currently, the scale separating the linear regime from the non-linear regime is of the order of a few tens of Megaparsecs. However, the main CMB observables (such as the temperature spectrum) depend essentially on the evolution of perturbations at very early times: before, during and shortly after recombination. Then, all currently observable scales were in the linear regime. Therefore, the study of linear cosmological fluctuations allows to study the early universe and CMB physics on all scales, as well as structure formation in the recent universe on large scales.

Throughout this section, we will assume that we live in a flat universe. Indeed, all current observations suggest that the spatial curvature of the universe is negligible. Besides, the mathematical description of perturbations is considerably more complicated in a non-flat universe. We recall that the flat FLRW line element can be written in natural units and cartesian coordinates as:

$$\begin{aligned} ds^2 &= -dt^2 + a(t)^2 \left[(dx^1)^2 + dx_2^2 + dx_3^2\right] \\ &= a(\eta)^2 \left[-d\eta^2 + dx_1^2 + dx_2^2 + dx_3^2\right] \ . \end{aligned} \tag{3.1}$$

3.1 Comoving Fourier space

Through a Fourier transformation of each perturbation with respect to the comoving coordinates, the linear evolution equations are decomposed into independent systems for each Fourier mode. For example, for a fluctuation of density

δ_X, we will write the transformation as

$$\begin{aligned} \delta_X(t,\vec{k}) &= \int \frac{d^3\vec{x}}{(2\pi)^{3/2}}\delta_X(t,\vec{x})e^{-i\vec{k}\cdot\vec{x}} \ , \\ \delta_X(t,\vec{x}) &= \int \frac{d^3\vec{k}}{(2\pi)^{3/2}}\delta_X(t,\vec{k})e^{i\vec{k}\cdot\vec{x}} \ , \qquad (3.2) \\ \text{with} \qquad & \int \frac{d^3\vec{x}}{(2\pi)^3}e^{i(\vec{k}'-\vec{k})\cdot\vec{x}} = \delta_D(\vec{k}'-\vec{k}) \ . \qquad (3.3) \end{aligned}$$

The quantity $2\pi/k$, where $k \equiv |\vec{k}|$, stands for the comoving wavelength. The physical wavelength is given by

$$\lambda(t) = a(t)\frac{2\pi}{k} \ . \qquad (3.4)$$

3.2 Observable wavelengths and causality

We want to estimate the scale of the largest observable comoving wavelength $2\pi/k$, that is, the minimal observable wavelength $k_{\rm min}$. We notice that a ray of light emitted at photon decoupling, when $t = t_{\rm dec}$, and travelling radially towards us in a flat FLRW universe must travel over a comoving distance given by

$$\chi_{\rm obs} = \int_{t_{\rm dec}}^{t_0} \frac{d\tilde{t}}{a(\tilde{t})} \ . \qquad (3.5)$$

This quantity can thus be called the comoving radius of the observable universe. Then the minimal observable wavelength $k_{\rm min}$ is defined from

$$\frac{2\pi}{k_{\rm min}} = \chi_{\rm obs} \ . \qquad (3.6)$$

Today, this corresponds to a physical wavelength given by

$$\lambda_{\rm max} = a(t_0)\frac{2\pi}{k_{\rm min}} = a(t_0)\chi_{\rm obs} = a(t_0)\int_{t_{\rm dec}}^{t_0} \frac{d\tilde{t}}{a(\tilde{t})} \equiv R_{\rm obs} \ . \qquad (3.7)$$

We defined a physical length $R_{\rm obs}$ usually called the "radius of the observable universe". In the minimal ΛCDM model this is close to 46 Gyr.

The radius of the observable universe is a particular case of what people call the "causal horizon". The causal horizon $d_H(t_1,t_2)$ is simply the distance travelled by a photon between two items t_1 and t_2, expressed in physical units at time t_2. Indeed, let us study a physical mechanism that starts at time t_1: for instance the free propagation of light in a transparent universe. At time t_2, two points separated by a distance greater than $2d_H(t_1,t_2)$ cannot be in "causal contact", in the sense that they cannot receive information from any comment event located at any time $t \geq t_1$. The causal horizon reads

$$d_H(t_1,t_2) = a(t_2)\int_{t_1}^{t_2} \frac{d\tilde{t}}{a(\tilde{t})} \qquad (3.8)$$

and we see that $R_{\rm obs} = d_H(t_{\rm dec},t_0)$.

During RD and MD, when $a(t) \propto t^n$ with $n < 1$, the causal horizon $d_H(t_1,t_2)$ is of the same order of magnitude as the Hubble radius $R_H(t_2) = \frac{a(t_2)}{\dot{a}(t_2)}$ for any

time $t_2 \gg t_1$. Indeed,

$$\begin{aligned} d_H(t_1,t_2) &= t_2^n \int_{t_1}^{t_2} \frac{d\tilde{t}}{t^n} = t_2^n \frac{(t_2^{1-n} - t_1^{1-n})}{1-n} \simeq \frac{t_2}{1-n} \ , \\ R_H(t_2) &= \frac{t_2}{n} \ , \\ d_H(t_1,t_2) &= \frac{n}{1-n} R_H(t_2) \ . \end{aligned} \tag{3.9}$$

In particular, if we neglect the impact of Λ domination in the recent universe, we see that $R_{\rm obs} \sim R_H(t_0)$: the current value of the Hubble radius gives the order of magnitude of the radius of the observable universe. This result also tells us that in a universe with first RD and then MD, wavelengths are in the "causal regime" or "acausal regime" depending on the fact that they are either "sub-Hubble", $\lambda < R_H$, or "super-Hubble", $\lambda > R_H$. This relation between causal/acausal and sub-Hubble/super-Hubble is not rue exactly, but is true in terms of orders of magnitude. The condition for being, for instance, sub-Hubble at time t, reads

$$\lambda(t) \ll R_H(t) \ \Leftrightarrow \ \frac{2\pi}{k} a(t) \ll \frac{1}{H(t)} \ \Leftrightarrow \ k \gg a(t)\, H(t) = \dot{a}(t) = \frac{a'(\eta)}{a(\eta)} \ . \tag{3.10}$$

The Friedmann equation shows that $a(\eta) \propto \eta$ during radiation domination, while $a(\eta) \propto \eta^2$ during matter domination. In both cases, within a factor of two, $\frac{a'}{a}$ is of the order of η^{-1}. Therefore, the condition for a mode to be in the sub-Hubble regime can also be written in terms of conformal time as:

$$\lambda \gg R_H \quad \Leftrightarrow \quad k \ll aH \quad \Leftrightarrow \quad k\eta \ll 1 \ . \tag{3.11}$$

During the domination of radiation and matter, the expansion is decelerated, which implies that $\lambda(t) \propto a(t)$ grows slower than $R_H(t) \propto \frac{a}{\dot{a}}$. Indeed, $\lambda(t)/R_H(t) \propto \dot{a}(t)$ has a negative derivative. Since the radius of the observable universe is given by the current Hubble radius (within a numerical factor), the observable Fourier modes are currently of the sub–Hubble type, i.e. obeying $\lambda < R_H$. But in the past, these modes must have been super–Hubble ($\lambda > R_H$), as shown in figure 3.1. Inflation explains how to produce primordial fluctuations with properties consistent with observations at wavelengths initially much longer than the Hubble radius (see Chapter IV). Later on, these modes then enter the Hubble radius one by one, starting with the shortest wavelengths. [1] The Fourier modes observable in the CMB anisotropy spectrum correspond very roughly to four decades below the current Hubble radius, $10^{-4} R_H(t_0) < \lambda(t_0) < R_H(t_0)$. Among these wavelengths, the longest ones entered the Hubble radius during matter domination, and the shortest ones during radiation domination. These two intervals are separated by a characteristic value of the wavenumber $k_{\rm eq} \equiv a(t_{\rm eq})\, H(t_{\rm eq})$.

3.3 The gauge ambiguity

Let us think how a perturbation (that is, a spatial fluctuation) is defined. We first focus on the perturbation of a Lorentz scalar, that is, a simple function

[1] The experimental fact that the expansion is currently accelerated means that the longest observable wavelengths are about to fall outside the Hubble radius again, but for us this will be a minor detail.

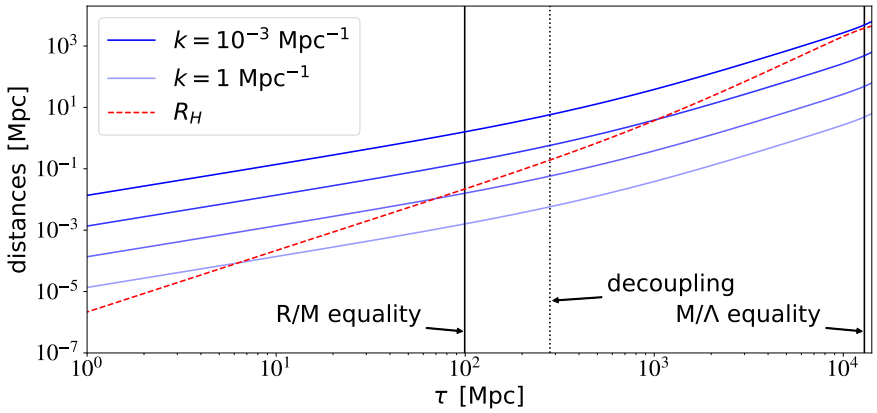

Figure 3.1: Evolution of different physical distances with respect to conformal time η (called τ in the label of this plot) in the standard cosmological model: the wavelength $\lambda = a\frac{2\pi}{k}$ of the modes typically observable in the CMB spectrum, and the Hubble radius R_H. The vertical lines show the time of equality between radiation and matter (R/M) and between matter and Λ (M/Λ), as well as the time of photon decoupling. Obtained with CLASS (`class-code.net`).

$S(x^\mu)$ of time and space, left invariant by a change of coordinate. Its perturbation at an event E of coordinate x^μ reads

$$\delta S(x^\mu_E) = S(x^\mu_E) - \bar{S}_H(x^0_E) \ , \tag{3.12}$$

where $\bar{S}_H(x^0_E)$ denotes the spatial average of $S(x^\mu_E)$ at time x^0_E, that is, the average of $S(x^\mu)$ over the 3D hypersurface H formed by all events of time x^0_E.

The mapping between the events E and the perturbations δS cannot be left invariant by a change of coordinates. Indeed, a general change of coordinates will change the times-slicing of the spacetime manifold, that is, the ensemble of equal-time hypersurfaces. In a new system of coordinates $\tilde{x}^\mu$, the perturbation in E reads

$$\delta S(\tilde{x}^\mu_E) = S(\tilde{x}^\mu_E) - \bar{S}_{\tilde{H}}(\tilde{x}^0_E) \ , \tag{3.13}$$

where $\tilde{H}$ is on of the equal-time hypersurfaces of the new coordinate system.

When we say that the universe can be described by small, linear perturbations, we implicitly assume that there exists at least one coordinate system in which $\delta S(x^\mu) \ll \bar{S}(x^0)$ in every event. Intuitively, in such a system, the equal-time hypersurfaces remain close to the hypersurfaces on which each quantity is homogeneous. In other words, in such a system, the equal-time hypersurfaces are nearly orthogonal to the gradient of all other quantities in every event.

Let us start from such a system of coordinates and let us perform a coordinate transformation $x^\mu \mapsto \tilde{x}^\mu$, defined by a displacement vector field $\epsilon^\mu(x^\alpha)$ such that

$$\tilde{x}^\mu = \delta^\mu_\alpha \, x^\alpha + \epsilon^\mu(x^\alpha) = x^\mu + \epsilon^\mu(x^\alpha) \ . \tag{3.14}$$

If ϵ^μ is small enough in every event, then the property $\delta S \ll \bar{S}$ should be preserved in the new system. Thus the ensemble of all coordinate systems in which the property $\delta S \ll \bar{S}$ holds are related to each other by infinitesimal coordinate transformations. After such a transformation, the value of δS at a

given event E changes. We call "gauge transformation" the mapping from δS in one coordinate system to $\delta\tilde{S}$ in another coordinate system but at the same event. A given gauge transformation is induced by a given change of coordinate. Thus there is a one-to-one correspondence between coordinate transformations and gauge transformations, but they are not the same thing:

- A coordinate transformation acts on Lorentz scalars like the identity; It acts and on Lorentz vectors / tensors by contracting contravariant (resp. covariant) indices with $\frac{\partial\tilde{x}^\mu}{\partial x^\nu}$ (resp. $\frac{\partial x^\nu}{\partial\tilde{x}^\mu}$).
- A gauge transformation acts on the *perturbations* of Lorentz scalars / vectors / tensors, and its mathematical expression is different.

In this course, we will not have time to derive the mathematical expression of a general gauge transformation.[2] We just admit that at order one in perturbations, it is possible to express the new perturbation as a function of: *(i)* the old perturbation, *(ii)* the spatial average, *(iii)* the displacement field. For instance:

- For a Lorentz scalar S, $\delta\tilde{S}$ can be expressed as a function of δS, $\bar{S}$ and ϵ^μ.
- For a Lorentz vector V^μ, $\delta\tilde{V}^\mu$ can be expressed as a function of δV^μ, $\bar{V}^\mu$ and ϵ^μ.
- For a Lorentz tensor $T^{\mu\nu}$, $\delta\tilde{T}^{\mu\nu}$ can be expressed as a function of $\delta T^{\mu\nu}$, $\bar{T}^{\mu\nu}$ and ϵ^μ.

Since a gauge transformation is induced by a coordinate transformation, and a coordinate transformation has four free functions $\epsilon^\mu(x^\alpha)$, gauge transformations also have four degrees of freedom, that is, they are a family of transformations defined by four free functions.

In a typical problem of cosmology, there are many perturbations: at least ten for the perturbations of the metric $\delta g_{\mu\nu}(x^\alpha)$, and even much more if the universe contains several fluids X, whose fluctuations are encoded in the perturbations $\delta T^{\mu\nu}_X$ of the stress-energy tensor, or more generally of the phase-space density function $f_X(x^\alpha, p^\beta)$.

With an appropriate coordinate transformation, it is possible to adjust the gauge in such a way that some of the perturbations vanish, that is, that some quantities appear as homogeneous in the time-slicing defined by the new coordinate system. If one starts form an arbitrary coordinate system and searches for a coordinate transformation leading to the cancellation of one perturbation, one finds that $\epsilon^\mu(x^\alpha)$ must satisfy one constraint equation. Since $\epsilon^\mu(x^\alpha)$ has four degrees of freedom, it is possible to cancel four perturbations at the same time. This leads to a simplification of the problem, but it does not render the problem completely trivial either, since many perturbations always remain. In the next section, we will present explicitly the most popular gauge choice in cosmology courses and textbooks.

3.4 Classification of perturbations

We can decompose the metric $g_{\mu\nu}$ and the total stress-energy tensor $T^{\mu\nu}$ of the universe into homogeneous quantities and perturbations of order one,

$$g_{\mu\nu}(t,\vec{x}) = \bar{g}_{\mu\nu}(t) + \delta g_{\mu\nu}(t,\vec{x}) \ , \quad T^{\mu\nu}(t,\vec{x}) = \bar{T}^{\mu\nu}(t) + \delta T^{\mu\nu}(t,\vec{x}) \ . \tag{3.15}$$

[2] The mathematics of gauge transformations can be found for instance in section 3 of Ma & Bertschinger, Astrophys. J. 455 (1995) 7-25 [astro-ph/9506072].

The background metric $\bar{g}_{\mu\nu}$ is that of FLRW, while the background stress-energy tensor is of the form $\bar{T}^{\mu\nu} = (\bar{\rho}+\bar{p})\bar{U}^\mu\bar{U}^\nu + p\bar{g}^{\mu\nu}$ with $\bar{U}^i = 0$. We will often use the stress-energy tensor with one covariant and one contravariant index, whose background value is the same in proper or conformal time, $\bar{T}^\mu{}_\nu = \bar{T}^{\mu\alpha}\bar{g}_{\alpha\nu} = \mathrm{diag}(-\bar{\rho}, \bar{p}, \bar{p}, \bar{p})$. For the time coordinate, we an use either the cosmological time t or the conformal time η.

The symmetric tensors $g_{\mu\nu}$ and $T^\mu{}_\nu$ each contain ten degrees of freedom (ten functions of $(\eta, \vec{x})$). A seminal paper by James Bardeen in 1980 shows that the perturbations can be decomposed into three sectors which remain decoupled from each other at first order in perturbations. These sectors are irreducible representations of the group of spatial rotations, which leaves the unperturbed Friedmann–Lemaître metric invariant. Bardeen scalars consist of rotation-invariant functions. Bardeen vectors have a spatial index and transform with a rotation matrix. Bardeen tensors have two spatial indices and transform with two rotation matrices.

Bardeen scalars describe the gravitational forces generated by density and pressure fluctuations in the universe, generalising Newtonian gravity. They will be the most important in this chapter. Bardeen vectors describe the gravitational effects of a vorticity component in the stress-energy tensor, which are negligible in the standard cosmological model (because any vorticity effect scales like $1/a$ and is rapidly diluted by the expansion). Bardeen tensors describe the two degrees of polarisation of gravitational waves and the matter fluctuations that can generate them or damp them[3]. CMB observations show that the role of such tensor modes is either negligible or very small with respect to that of scalar perturbations. They are however potentially detectable by future experiments (see Chapter V).

In the metric perturbation $\delta g_{\mu\nu}$ and stress-energy perturbation $\delta T^{\mu\nu}$, the Bardeen scalars are associated with four degrees of freedom (out of a total of ten):

1. the term of index (00),
2. the trace of the 3×3 spatial tensor (= stress tensor) of indices (ij).,
3. the curl-free part of the 3-vector of index $(0i)$,
4. the longitudinal[4] and trace-free component of the spatial tensor.

In the perturbed metric $\delta g_{\mu\nu}$, these four components have the following physical meaning (in the same order as above):

1. The local relative perturbation ψ of the mean proper time of comoving observers. When travelling over density fluctuations, different comoving observers perceive time slightly differently, as $d\tilde{t}(\vec{x}) = (1 + \psi(t, \vec{x}))\, dt$.
2. The local relative perturbation ϕ of the average scale factor $a(t)$, the "perturbed scale factor" being given by $\tilde{a}(t, \vec{x}) = (1 - \phi(t, \vec{x}))\, a(t)$.

[3] In the vacuum, scalar and vector perturbations can only be zero, as they only describe the response of the metric to matter, without having their own propagation equation. On the other hand, tensor perturbations have their own propagation equation, and can therefore be generated by a few mechanisms in the primordial universe (in particular, inflation, as we will see in chapter V). Then they propagate freely even in an empty universe as waves of the metric.

[4] A symmetric spatial tensor A^l_k is called longitudinal when its divergence $\partial_l A^l_k$ forms an irrotational vector B_k: $\forall i$, $\epsilon^{ijk}\partial_j B_k = 0$, and thus $\epsilon^{ijk}\partial_j\partial_l A^l_k = 0$ (ϵ^{ijk} being the maximally antisymmetric tensor).

3. A preferred spatial direction in the perturbed metric generating a "coordinate flow".

4. Another preferred spatial direction resulting in a "coordinate shear effect".

The last two effects would deserve more explanation, but this will not be necessary to understand this chapter. Indeed, the most popular gauge choice in cosmology courses and textbooks is the one in which the last two degrees of freedom vanish. One can show explicitely that this is always possible (indeed, gauge transformations have the ability to cancel two scalar degrees of freedom and two vector degrees of freedom; tensor perturbations are gauge-invariant by construction).

The advantage of this gauge is mainly pedagogical: it gets rid of the two degrees of freedom that are the less intuitive. In this gauge, one gets a relatively simple expression for the perturbed Friedmann-Lemaître metric (in a flat universe),

$$\begin{aligned} ds^2 &= -(1+2\psi)dt^2 + (1-2\phi)a^2(t)d\vec{x}^{\,2} \\ &= a^2(\eta)\left[-(1+2\psi)d\eta^2 + (1-2\phi)d\vec{x}^{\,2}\right] , \end{aligned} \tag{3.16}$$

which only involves local distortions of time and of the scale factor. This gauge is called indifferently longitudinal (because the non-diagonal perturbations are zero) or Newtonian (because in the limit of small scales with respect to the radii of curvature of space-time, the two potentials ϕ and ψ are equal and play the role of the Newtonian gravitational potential, as we shall see later).[5]

In the perturbed stress-energy tensor $\delta T^{\mu\nu}$, the scalar perturbations have the following physical meaning (still in the same order as above, and with definitions valid both in conformal time and in proper time):[6]

1. the energy density perturbation $\delta\rho = -\delta T^0_{\ 0}$;

2. the pressure perturbation $\delta p = \delta T^i_{\ i} = \bar{g}_{ii}\delta T^{ii}$ where i is fixed (to 1, 2, or 3);

3. the potentiel v of the curl-free component of the fluid's bulk velocity v^i, normalised such that $\sqrt{-\bar{g}^{00}}v^i = \bar{g}^{ij}\partial_j v = a^{-2}\partial_i v$, that contributes to the curl-free component $(\delta T^0_{\ i})_{\rm cf}$ of the stress-energy tensor as $(\delta T^0_{\ i})_{\rm cf} \equiv (\bar{\rho}+\bar{p})\partial_i v$;

4. the potential s of the shear forces in the fluid, which contributes to a longitudinal and trace-free component $\delta T^{\ ij}_{\rm lon.}$ of the stress-energy tensor as $(\delta T^i_{\ j})_{\rm lon.} \equiv (\bar{\rho}+\bar{p})\bar{g}^{ik}(\partial_k\partial_j - \frac{1}{3}\delta_{kj}\Delta)\,s$. Such shear forces are equivalent to an anisotropic pressure with the geometry of a quadrupole (i.e. with a stronger pressure always coming from two opposite directions).

Since we have already used the two scalar degrees of freedom of gauge transformations to simplify the perturbations of the metric, we must keep all the four

[5] Note that cosmologists often use other gauges in their calculations because they may offer other technical advantages (stability of numerical calculations, similarity with observable quantities, etc.)

[6] Some of the relations below can be derived carefully while assuming a perfect fluid with $T^\mu_{\ \nu} = p\,\delta^\mu_{\ \nu} + (\rho+p)U^\mu U_\nu$. We start from writing the bulk velocity as $U^\mu = (U^0, v^i)$. The normalisation condition $\vec{U}\cdot\vec{U} = -1$ gives $U^0 = \sqrt{-\bar{g}^{00}}(1-\psi)$ if we only keep terms of order zero and one in perturbations. Then $U^\mu = (\sqrt{-\bar{g}^{00}}(1-\psi), v^i)$ and, after lowering indices, $U_\nu = (-\sqrt{-\bar{g}_{00}}(1+\psi), a^2 v^i)$ at the same order. This gives, first, $T^0_{\ 0} = p - (\rho+p)(1-\psi)(1+\psi) = \rho = \bar{\rho} + \delta\rho$ at order one; second, $T^1_{\ 1} = T^2_{\ 2} = T^3_{\ 3} = p + (\rho+p)\times\mathcal{O}(v^2) = p = \bar{p} + \delta p$; and third, $T^0_{\ i} = (\rho+p)(1-\psi)\sqrt{-\bar{g}^{00}}a^2 v^i = (\bar{\rho}+\bar{p})\sqrt{-\bar{g}^{00}}a^2 v^i$.

components of the perturbed stress-energy tensor. For simplicity, we will reformulate these four variables in terms of four dimensionless quantities. To describe density fluctuations, we introduce the relative fluctuation $\delta \equiv \delta\rho/\bar{\rho}$. For pressure fluctuations, we use the square of the sound speed $c_s^2(t,\vec{x}) \equiv \delta p(t,\vec{x})/\delta\rho(t,\vec{x})$. To describe the third scalar degree of freedom of the stress-energy tensor, instead of referring to the velocity potential v, one can use the velocity divergence θ such that

$$(\bar{\rho}+\bar{p})\theta \equiv \bar{g}^{ij}\partial_j(\delta T^0{}_i)_{\rm cf} = (\bar{\rho}+\bar{p})a^{-2}\sum_i(\partial_i)^2 v = (\bar{\rho}+\bar{p})\frac{\Delta v}{a^2} \ . \tag{3.17}$$

Note that Δ represents the Laplacian with respect to comoving coordinates, and $\frac{\Delta}{a^2}$ the physical Laplacian. The relation between v and $\theta = \frac{\Delta v}{a^2}$ is unique[7]. Similarly, to describe the fourth degree of freedom, it is usual to define a quantity σ such that

$$(\bar{\rho}+\bar{p})\frac{\Delta\sigma}{a^2} \equiv -(\partial_i\partial_k - \frac{1}{3}\delta_{ik}\Delta)\bar{g}^{jk}(\delta T^i{}_j)_{\rm lon.} = \frac{2}{3}(\bar{\rho}+\bar{p})\frac{\Delta(\Delta s)}{a^4} \ . \tag{3.18}$$

Again, the relationship between s and $\sigma = \frac{2}{3}\frac{\Delta s}{a^2}$ is unique. As the σ function describes the presence of a pressure with a quadrupolar angular dependence, it is customary to call it directly the anisotropic stress. In summary, the scalar perturbations of the total stress-energy tensor of the universe are contained in four dimensionless scalar functions $\{\delta, c_s^2, \theta, \sigma\}$. In the following we will write the total perturbed stress-energy tensor $\delta T^\mu_{\ \nu}$ as a sum over several components $\delta T_X{}^\mu_{\ \nu}$ for each species X of matter and radiation (photons, electrons, baryons, neutrinos, dark matter). Thus we will introduce separate variables $\{\delta_X, c_{s\,X}^2, \theta_X, \sigma_X\}$ for each of them.[8]

3.5 Linearised Einstein equations

To describe the effects of gravitation in a perturbed universe, one has to insert the perturbed Friedmann metric and stress-energy tensor into the Einstein equation, and calculate the Einstein tensor to order one. This gives ten linear equations which can be separated into four equations for the Bardeen scalars, four for the Bardeen vectors and two for the Bardeen tensors. These equations are not all independent of each other, due to Bianchi identities.

For the scalar sector, we obtain four equations. The most useful ones in this chapter will be (using conformal time η and in comoving Fourier space):

$$\frac{2}{a^2}\left[k^2\phi + 3\frac{a'}{a}\left(\phi' + \frac{a'}{a}\psi\right)\right] = -8\pi G\sum_X \bar{\rho}_X\delta_X \ , \tag{3.19}$$

$$\frac{2}{3}\frac{k^2}{a^2}(\phi-\psi) = 8\pi G\sum_X(\bar{\rho}_X+\bar{p}_X)\sigma_X \ . \tag{3.20}$$

These equations correspond respectively to $\delta G^0{}_0$ (equal to $8\pi G\,\delta T^0{}_0$) and $-\frac{a^2}{k^2}(k_ik_j - \frac{1}{3}\delta_{ij}k^2)\delta G_{\rm lon.}^{\ ij}$ (equal to its counterpart in T^{ij}). For completeness, let us also

[7] Indeed, the only non-zero solutions of the homogeneous equation $\Delta v = 0$ diverge at infinity. Perturbation theory is only valid when all functions remain small on the whole space, which excludes such solutions.

[8] Note that the extensive quantities that can be added up between the different fluids are not directly $\{\delta_X,\, c_{s\,X}^2,\, \theta_X,\, \sigma_X\}$ but $\{\delta\rho_X,\, \delta p_X,\, (\bar{\rho}_X+\bar{p}_X)\theta_X,\, (\bar{\rho}_X+\bar{p}_X)\sigma_X\}$.

give the equations associated to $\delta G^i{}_i$ (with fixed i) and $ik_i\,\delta G_{\rm cf}{}^{0i}$:

$$\frac{2}{a^2}\left[\phi'' + \frac{a'}{a}(\psi' + 2\phi') + \left(2\frac{a''}{a} - \frac{a'^2}{a^2}\right)\psi\right] = 8\pi G \sum_X \bar\rho_X c_{s\,X}^2 \delta_X \ , \quad (3.21)$$

$$2\frac{k^2}{a^2}\left(\phi' + \frac{a'}{a}\psi\right) = 8\pi G \sum_X (\bar\rho_X + \bar p_X)\theta_X \ , \quad (3.22)$$

The second equation 3.20 shows that the difference between the two fluctuations of the metric is caused by the anisotropic pressure of matter. This anisotropic pressure plays a very small role in cosmological perturbation theory. It has a very minor impact on CMB anisotropies and on the formation of large scale structures[9]. In this chapter, for simplicity, we will neglect the σ_X terms, which amounts to working in the $\phi = \psi$ approximation. The two fluctuations of the metric then play the role of a generalized gravitational potential.

The first equation 3.19 shows how this potential responds to the presence of matter density fluctuations. For sub–Hubble modes (with $k \gg a'/a$), the left-hand side of the equation 3.19 is dominated by the first term, which implies in first approximation

$$-\frac{k^2}{a^2}\phi = 4\pi G \sum_X \bar\rho_X \delta_X = 4\pi G\, \delta\rho_{\rm tot.} \ . \quad (3.23)$$

This is precisely the Poisson equation of Newtonian gravitation, $\Delta\psi_{\rm Newton} = 4\pi G\,\rho_{\rm Newton}$, transposed into the cosmological context. Indeed, the gravitational potential can only be generated by the density fluctuations $\delta\rho_{\rm tot.}$, while the average density $\bar\rho_{\rm tot.}$ contributes to the homogeneous expansion. In real space, the term on the left-hand side is the physical gradient of the gravitational potential, $\frac{\Delta}{a^2}$, while Δ is just the comoving gradient.

3.6 Equations of motion for matter fluctuations

The Bianchi identities correspond to the conservation equations of the total stress-energy tensor of the universe, $D_\mu T^\mu_{\ \nu} = 0$. They give two equations for (Bardeen) scalar modes and two for (Bardeen) vector modes. In a universe containing several decoupled fluids labelled by X, the same equations apply individually to the stress-energy tensor $T_X{}^\mu_{\ \nu}$ of each species. We then obtain two (Bardeen) scalar equations of motion per fluid X, the continuity (or conservation of energy) equation and the Euler equation:

$$\delta'_X = -(1+w_X)(\theta_X - 3\phi') - 3\frac{a'}{a}(c_{s\,X}^2 - w_X)\delta_X \ , \quad (3.24)$$

$$\theta'_X = -\frac{a'}{a}(1 - 2w_X - c_{a\,X}^2)\theta_X + \frac{c_{s\,X}^2}{1+w_X}k^2\delta_X - k^2\sigma_X + k^2\psi \ . \quad (3.25)$$

In section 3.4, we have already given the definition of the speed of sound for each fluid, $c_{s\,X}^2(\eta,\vec k) \equiv \delta p_X(\eta,\vec k)/\delta\rho_X(\eta,\vec k)$. The equations feature another quantity

[9] The photons do not contribute to this term as long as they are strongly coupled to the baryons: they then form a perfect fluid with an isotropic pressure. When the photons decouple, σ_γ increases, but then it rapidly tends towards zero in the sub–Hubble limit. Since baryons and electrons are always in thermal equilibrium, their pressure remains isotropic. At epochs important for CMB physics, dark matter and neutrinos are decoupled. Dark matter has negligible pressure (isotropic and anisotropic), but neutrinos do not. The anisotropic pressure σ_ν is maximal around the Hubble radius crossing, but its physical impact is negligible in the super–Hubble and sub–Hubble limits.

with the same dimension, the adiabatic sound speed, $c^2_{a\,X}(\eta) \equiv \dot{\bar{p}}_X(\eta)/\dot{\bar{\rho}}_X(\eta)$. In these equations we have also introduced the equation of state parameter $w_X(\eta) \equiv \bar{p}_X(\eta)/\bar{\rho}_X(\eta)$.

For perfect fluids, one can show that the anisotropic pressure σ_X always vanishes, while $c^2_{s\,X} = c^2_{a\,X} = w_X$. Then, the equations 3.24 and 3.25 only involve two variables (δ_X, θ_X) and form a closed system (together with the Einstein equation).

For decoupled species, the anisotropic pressure σ_X is non-zero and the equations 3.24, 3.25 do not form a closed system. This shows the need for a more complete description using kinetic theory, i.e. the distribution function in phase space and the Boltzmann equation. In real space, the distribution function $f_X(\eta, \vec{x}, \vec{p})$ depends on conformal time, position and momentum. For a decoupled species like neutrinos, $X = \nu$, the Boltzmann equation indicates that along each space-time geodesic, this distribution evolves as $\frac{d}{d\eta} f_\nu = 0$. For photons, we must add the Thomson scattering term, which also depends on the electron distribution: formally, this term is a C functional of the f_γ and f_e functions. We then obtain a linearised Boltzmann equation of the form:

$$\frac{d}{d\eta} f_\gamma = C\left[f_\gamma, f_e\right] \ . \tag{3.26}$$

This last equation will be the most important one in Chapter V. We will develop it at that time.

Chapter 4

Inflation

4.1 Motivations for cosmological inflation

So far, we discussed an incomplete version of the ΛCDM model in which the universe is dominated first by radiation, then matter and then Λ. In this section, we will show that this model contains some inconsistencies. We will prove that these are solved if we assume that radiation domination was preceded by a stage of accelerated expansion. A model that could explain such a stage will be presented in the following section.

4.1.1 Flatness problem

Today, Ω_k is measured to be at most of order 10^{-2}, possibly much smaller, while $\Omega_r \equiv \rho_r/\rho_{\rm crit} \simeq \rho_r/(\rho_\Lambda + \rho_{\rm m})$ is of the order of 10^{-4}. Since $\rho_k^{\rm eff} = -\frac{3k}{8\pi G a^2}$ 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

$$\frac{\rho_k^{\rm eff}(t_i)}{\rho_r(t_i)} \leq \left(\frac{a(t_i)}{a(t_0)}\right)^2 = \left(\frac{\rho_r(t_0)}{\rho_r(t_i)}\right)^{1/2} . \tag{4.1}$$

Today, the radiation energy density $\rho_r(t_0)$ is of the order of $(10^{-4}{\rm eV})^4$. If the early universe reached the order of the Planck density $(10^{18}{\rm GeV})^4$ at the Planck time t_P, then at that time the ratio was

$$\frac{\rho_k^{\rm eff}(t_P)}{\rho_r(t_P)} = \frac{(10^{-4}{\rm eV})^2}{(10^{18}{\rm GeV})^2} \sim 10^{-62} . \tag{4.2}$$

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 terms on the right-hand side of the Friedmann equation: matter and spatial curvature. 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 $\rho_k^{\rm eff}/\rho_r$ 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 notice that $\Omega_k \equiv \rho_k^{\rm eff}/\rho_{\rm crit}$ 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 amount 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, $|\Omega_k| \sim 1$ today at t_0, and the expansion is exactly exponential between t_i and t_f. 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. Finally, we could assume that inflation is not exactly exponential, but this would make it less efficient and more inflation would be needed, as we shall see below.

A stage of expansion with an exponential scale factor is called a De Sitter stage. During such a stage, $H = \frac{\dot{a}}{a}$ is constant. Actually, H is not allowed to increase with time in an expanding universe: it would mean that $\rho_{\rm tot}$ increases while comoving volumes are stretched, and thus it would violate energy conservation. So H can either be constant during a De Sitter stage, or decrease in any other situation. We are now interested in De Sitter inflation because keeping H constant during inflation is the most efficient way to solve the flatness problem. Any other model for inflation (with $\ddot{a} > 0$ but a decreasing H) is possible, but requires a bigger amount of inflation in order to solve the flatness problem.

So, we assume that between t_i and t_f the scale factor grows exponentially from a_i to a_f, 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

$$\frac{\rho_k^{\rm eff}(a_0)}{\rho_k^{\rm eff}(a_i)} = \left(\frac{a_i}{a_0}\right)^2 = \frac{\rho_r(a_0)}{\rho_{\rm inf}(a_i)} = \frac{\rho_r(a_0)}{\rho_{\rm inf}(a_f)} = \frac{\rho_r(a_0)}{\rho_r(a_f)} = \left(\frac{a_f}{a_0}\right)^4 \tag{4.3}$$

and we finally obtain the relation

$$\frac{a_f}{a_i} = \frac{a_0}{a_f} \ . \tag{4.4}$$

So, the condition for the minimal duration of inflation reads

$$\frac{a_f}{a_i} \geq \frac{a_0}{a_f} \ , \tag{4.5}$$

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

$$N \equiv \ln\, a\ . \tag{4.6}$$

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

$$(N_f - N_i) \geq (N_0 - N_f) \tag{4.7}$$

i.e., the number of e-folds during inflation should be greater or equal to the number of e-folds after inflation, $\Delta N \equiv N_0 - N_f$. 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 $(10^{-4}\mathrm{eV})^4$. We further admit that the inflationary energy scale is at most of the order of $(10^{16}\mathrm{GeV})^4$ (we will see in Chapter 5 that, otherwise, current observations of CMB anisotropies would have revealed a signature of primordial gravitational waves). This gives

$$\Delta N = \ln\frac{a_0}{a_f} = \ln\left(\frac{\rho_r(a_f)}{\rho_r(a_0)}\right)^{1/4} \leq \ln 10^{29} \sim 67\ . \tag{4.8}$$

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 e-folds after inflation would reduce to

$$\Delta N \sim \ln 10^{16} \sim 37 \tag{4.9}$$

and the flatness problem can be solved with only 37 e-folds of inflation.

4.1.2 Horizon problem

We have already seen 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,

$$d_H(t_1, t_2) = a(t_2)\int_{t_1}^{t_2}\frac{dt}{a(t)}\ . \tag{4.10}$$

Noticing that $dt = da/(aH)$,

$$d_H(a_1, a_2) = a_2\int_{a_1}^{a_2}\frac{da}{a^2 H(a)}\ , \tag{4.11}$$

where the Hubble parameter is seen now as a function of a. Finally, we already found that if the expansion is decelerated between t_1 and t_2 and if $t_2 \gg t_1$, then $d_H(t_1, t_2) \simeq R_H(t_2)$.

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 of 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 unnatural in the framework of the Hot Big Bang scenario.

What is the reason for this problem? When one computes the horizon, one integrates $(a^2 H)^{-1}$ over da. The convergence of the integral

$$\int_{a_1}^{a_2} \frac{da}{a^2 H(a)} = \int_{a_1}^{a_2} \frac{da}{a\dot{a}} \tag{4.12}$$

with respect to the lower boundary $a_1 \to 0$ depends on the fact that the expansion is accelerated or decelerated. For linear expansion ($\dot{a} = $ constant), the integrand is $1/a$, the limiting case between convergence and divergence. If the expansion is decelerated, $\dot{a}$ decreases and the integral converges: this leads to the proof that $d_H(a_1, a_2) \simeq R_H(a_2)$ for $a_2 \gg a_1$. But, if the expansion is accelerated, $\dot{a}$ increases and the integral diverges in the limit $a_1 \to 0$. Then, one can obtain an infinitely large horizon at time t_2, simply by choosing a_1 to be small enough.

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 the exercises, you will prove that the condition for solving this problem is exactly the same as for solving the flatness problem:

$$(N_f - N_i) \geq (N_0 - N_f)\ . \tag{4.13}$$

4.1.3 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

$$\lambda(t) = a(t)\frac{2\pi}{k} \tag{4.14}$$

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., $\lambda > R_H \Leftrightarrow k < 2\pi aH$). 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 at the time, for instance, 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.

Let's try to take the second point of view. Can we find a model such that the largest wavelength observable today, which is $\lambda_{\max}(t_0) \sim R_H(t_0)$, 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,

$$\frac{\lambda(t)}{R_H(t)} = \frac{2\pi a(t)}{k}\frac{\dot{a}(t)}{a(t)} = \frac{2\pi\dot{a}(t)}{k} \ , \tag{4.15}$$

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 4.1).

It is easy to show that once again, the minimal number of e-folds of inflation requested for solving this problem should be at least equal to the number of e-folds after inflation.

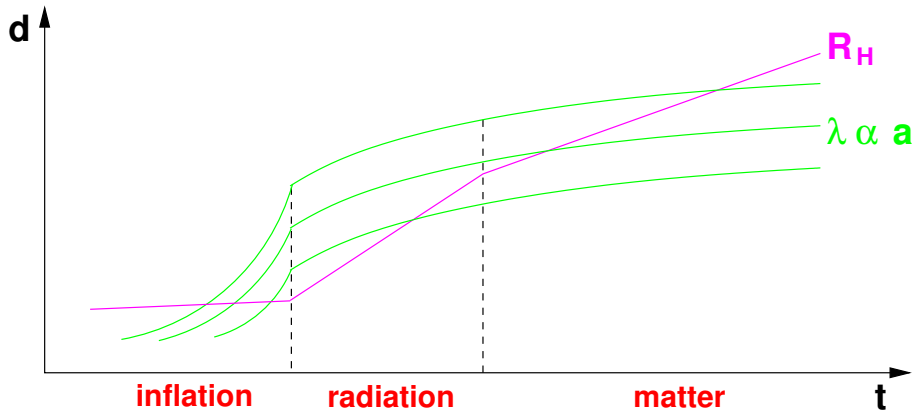

Figure 4.1: *Comparison of the Hubble radius with the physical wavelength of a few cosmological perturbations. During the initial stage of accelerated expansion (called inflation), the Hubble radius grows more slowly than each wavelength. So, cosmological perturbations originate from inside R_H. Then, the wavelength of each mode grows larger than the Hubble radius during inflation and re–enters during radiation or matter domination.*

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 of each other.

4.1.4 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 "topologically stable 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.

Inflation can solve the problem provided that it takes place after the creation of monopoles. 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 the unwanted relics decay faster than the effective curvature density ($\rho_k^{\rm eff} \propto a^{-2}$).

4.2 Slow-roll inflation

We have seen that the problems of section 4.1 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 (1.40) in a flat universe with the conservation equation (1.43), it is easy to find that

$$\ddot{a} > 0 \qquad \Rightarrow \qquad \rho + 3p < 0. \tag{4.16}$$

What type of matter corresponds to such an unusual relation between density and pressure? A positive cosmological constant can do the job:

$$p_\Lambda = -\rho_\Lambda \qquad \Rightarrow \qquad \rho_\Lambda + 3p_\Lambda = -2\rho_\Lambda < 0. \tag{4.17}$$

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

$$S = \int d^4x \sqrt{|g|}\,(\mathcal{L}_g + \mathcal{L}_\varphi) \tag{4.18}$$

involves the Einstein-Hilbert Lagrangian of gravity

$$\mathcal{L}_g = \frac{R}{16\pi\mathcal{G}} \tag{4.19}$$

and that of the scalar field

$$\mathcal{L}_\varphi = -\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - V(\varphi) = -\frac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - V(\varphi) \tag{4.20}$$

where $V(\varphi)$ is the scalar potential. The variation of the action with respect to $g_{\mu\nu}$ enables to define the energy-momentum tensor

$$T_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi + \mathcal{L}_\varphi g_{\mu\nu} \tag{4.21}$$

and the Einstein tensor $G_{\mu\nu}$, which are related through the Einstein equations

$$G_{\mu\nu} = 8\pi\mathcal{G}\,T_{\mu\nu}\ . \tag{4.22}$$

Instead, the variation of the action with respect to φ gives Klein-Gordon equation

$$-\frac{1}{\sqrt{|g|}}\partial_\mu\left[\sqrt{|g|}\partial^\mu\varphi\right] + \frac{\partial V}{\partial\varphi} = 0\ . \tag{4.23}$$

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

$$ds^2 = -dt^2 + a(t)^2[dx^2 + dy^2 + dz^2] \tag{4.24}$$

is filled by a homogeneous classical scalar field $\bar{\varphi}(t)$ that we will call the *inflaton*. One can show that the corresponding energy-momentum tensor is diagonal, $T^\nu_\mu = \mathrm{diag}(-\rho, p, p, p)$, with

$$\rho = \frac{1}{2}\dot{\bar{\varphi}}^2 + V(\varphi)\ , \tag{4.25}$$

$$p = \frac{1}{2}\dot{\bar{\varphi}}^2 - V(\varphi)\ . \tag{4.26}$$

The Friedmann equation reads

$$-G_0^0 = 3H^2 = 8\pi\mathcal{G}\,\rho = 8\pi\mathcal{G}\left(\frac{1}{2}\dot{\bar{\varphi}}^2 + V(\varphi)\right) \tag{4.27}$$

and the Klein-Gordon equation

$$\ddot{\bar{\varphi}} + 3H\dot{\bar{\varphi}} + \frac{\partial V}{\partial\varphi}(\bar{\varphi}) = 0\ . \tag{4.28}$$

These two independent equations specify completely the evolution of the system.[1] Finally, the combination $G_i^i - G_0^0 = 8\pi\mathcal{G}(\rho + p)$ provides a very useful relation

$$\dot{H} = -4\pi\mathcal{G}\,\dot{\bar{\varphi}}^2 \tag{4.30}$$

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.

To get a long stage of inflation, we must require that the *first slow-roll condition*

$$\frac{1}{2}\dot{\bar{\varphi}}^2 \ll V(\bar{\varphi}) \tag{4.31}$$

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*

$$|\ddot{\bar{\varphi}}| \ll \left|\frac{\partial V}{\partial\varphi}(\bar{\varphi})\right|\ . \tag{4.32}$$

You may wonder what "$\ll$" means here. In practise, we will see that successful models of inflation just require the quantities on the left to be at least one order of magnitude smaller than quantities on the right during most of inflation: this is sufficient to ensure the minimal number of e-folds $\Delta N_{\rm inflation} \sim \mathcal{O}(60)$.

The two slow-roll conditions can be rewritten as conditions either on the slowliness of the variation of $H(t)$, or on the flatness of the potential $V(\varphi)$. You will prove this in the exercises. In particular, one can define two parameters ϵ and η quantifying the "flatness of the potential", such that

$$\epsilon \equiv \frac{M_P^2}{16\pi}\left(\frac{V'}{V}\right)^2, \qquad \eta \equiv \frac{M_P^2}{8\pi}\frac{V''}{V}\ , \tag{4.33}$$

in which, exceptionally, the prime denotes a derivative with respect to φ rather than to conformal time. These are called the *slow-roll parameters.* You will find in the exercises that the two previous slow-roll conditions are equivalent to

$$\epsilon \ll 1\ , \qquad |\eta| \ll 1\ . \tag{4.34}$$

[1] It is worth mentioning that the full Einstein equation provides another relation

$$G_i^i = -\left(2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right) = 8\pi\mathcal{G}\,p = 8\pi\mathcal{G}\left(\frac{1}{2}\dot{\bar{\varphi}}^2 - V(\varphi)\right)\ . \tag{4.29}$$

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.

Again, having slow-roll parameters at least as small as $\mathcal{O}(10^{-1})$ during most of inflation is sufficient to ensure the minimal number of e-folds $\Delta N_{\rm inflation} \sim \mathcal{O}(60)$, as we shall check on a few concrete examples near the end of this chapter.

When these two conditions hold, the Friedmann and Klein-Gordon equations become

$$3H^2 \simeq 8\pi\mathcal{G}\,V(\bar{\varphi}) \,, \tag{4.35}$$

$$3H\dot{\bar{\varphi}} \simeq -\frac{\partial V}{\partial \varphi}(\bar{\varphi}) \,. \tag{4.36}$$

The two slow-roll conditions can be rewritten as conditions either on the slowliness 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 independently by the Russian Alexei Starobinsky and the American Alan Guth (after some precursor work by the Belgians Robert Brout and François Englert). 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 at some point in the very early universe, the energy density is dominated by a scalar field, with a potential flat enough for slow–roll. (This could be right after the quantum gravity era with some kind of random initial conditions, or a bit later, at the beginning of a phase transition, when a scalar field is close to a local maximum of the potential). 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 then decay into fermions and bosons through a non-linear resonance process. This can explain the origin of all the particles filling our universe. Then these particles interact and reach thermal equilibrium: this is why this stage is called "reheating". There is no thermal equilibrium and no possible definition of temperature during inflation: temperature emerges from 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 to large scale structures in the universe. We will study this in the next sections.

4.3 Inflationary perturbations

4.3.1 Quantisation of perturbations during inflation

We decompose the inflaton field and the metric in a background term and some small perturbations,

$$\varphi(t,\vec{x}) = \bar{\varphi}(t) + \delta\varphi(t,\vec{x}) \ , \qquad g_{\mu\nu}(t,\vec{x}) = \bar{g}_{\mu\nu}(t) + \delta g_{\mu\nu}(t,\vec{x}) \ . \tag{4.37}$$

The fluctuations can be expanded in comoving Fourier space, that is, in wavenumbers $\vec{k}$,

$$\delta\varphi(t,\vec{k}) \ , \qquad \delta g_{\mu\nu}(t,\vec{k}) \ . \tag{4.38}$$

These perturbations include several "propagating degrees of freedom" obeying to a wave equations, namely: the scalar field fluctuations, which are part of the sector of "Bardeen scalars", and which are coupled to the scalar metric perturbations, and the two degrees of freedom of gravitational waves, which are part of the sector of "Bardeen tensors".

We start studying each propagating degree of freedom when their wavelength is sub-Hubble, i.e., when $k \gg aH$ or $k\eta \gg 1$. Then, we can quantise them using the laws of quantum field theory. In the sub-Hubble limit, the effects of spacetime curvature are negligible, and in the limit of small perturbations, non-linear couplings in the equations of motion (e.g. terms like $\lambda(\delta\varphi)^3$ that would potentially couple different Fourier modes with each other) are also negligible. Thus we can just apply the rules for the "quantisation of free fields in flat spacetime", which are very easy and well-known. Essentially, we must treat each independent Fourier mode (e.g. $\delta\varphi(t,\vec{k})$ for the scalar field) as a harmonic oscillator. Assuming that this mode is in its fundamental (vaccum) state, it has a Gaussian wave function and a variance

$$\langle|\delta\varphi(t,\vec{k})|^2\rangle = \frac{\hbar}{2k} \ . \tag{4.39}$$

Later on, the Fourier modes cross the Hubble radius and become super-Hubble. Then, one prove several important features:

- the quantum statistics of the modes becomes indistinguishable from that of a classical stochastic system, that is, a stochastic system described by standard probabilities (e.g. $P(\delta\varphi)$) rather than wave functions (e.g. $\Psi(\delta\varphi)$.
- the probability $P(\delta\varphi)$ that a given mode has a value $\delta\varphi$ is still a gaussian centred in zero, like in the sub-Hubble regime (due to the Gaussianity of the wave function of the fundamental state).
- the variance $\langle|\delta\varphi(t,\vec{k})|^2\rangle$ can be followed as a function of time by looking at the solution of the classical equation of motion. For the scalar field, this equation is the linearised Klein-Gordon equation that has been derived in the exercises (combined with the Einstein equations in order to eliminate metric fluctuations). Thus, this variance may depend on the background evolution $H(t)$ and on the potential $V(\varphi)$, which both appear in the equation of motion.

Thus, at the end of inflation, we expect that all propagating degrees of freedom (like scalar field fluctuations and gravitational waves) have random fluctuations, described by Gaussian statistics, with a variance that can be computed once we know the details of the background evolution for the Hubble rate $H(t)$ and the potential $V(\bar{\varphi})$. More precisely, the field can be shown to have a universal solution, independent of $H(t)$ and $V(\bar{\varphi}(t))$, in the two asymptotic limits $k \gg aH$ (sub-Hubble) and $k \ll aH$ (super-Hubble). Instead, the evolution does depend on the details of $H(t)$ and $V(\bar{\varphi}(t))$ near the time of Hubble crossing, when $k \simeq aH$. Thus, a detailed calculation would show that the variance of $\delta\varphi(t, \vec{k})$ at the end of inflation depends on $H(t)$ and $V(\bar{\varphi}(t))$ near the time $t = t_k$ at which $k = aH$. Using the slow-roll equations for the background evolution, the result can then be expressed as a function of the potential $V(\bar{\varphi})$ and its derivative $V'(\bar{\varphi})$ evaluated at the value $\bar{\varphi} = \bar{\varphi}_k$ reached by the background field at the time t_k.

We don't have time to enter more into the details of this fascinating topic, which is one of the most concrete known applications of quantum field theory in curved space time.[2] We will jump directly to the main results, first for the scalar modes (that is, the prerturbations of the field $\delta\varphi$ coupled to the perturbations of the metric in the Newtonian gauge ϕ, ψ), and then for the tensor modes (that is, the degrees of freedom of $\delta g_{\mu\nu}$ accounting for gravitational waves).

4.3.2 Scalar perturbations

In the scalar sector, during inflation, the degrees of freedom in the Newtonian gauge are $\delta\varphi$, ϕ and ψ (but there is only one "propagating degree of freedom", that is, one independent combination of $\delta\varphi$, ϕ and ψ obeying to a wave equation). After inflation, the inflaton field has decayed into several species and the scalar degrees of freedom are (δ_x, δ_x, $c^2_{\mathrm{s},x}$, σ_x) for each species x, plus still (ϕ, ψ) for the metric.

In linear perturbation theory, as long as all these degrees of freedom are linearly coupled, it is enough to know that we generated Gaussian random perturbations for $\delta\varphi$ in order to be sure that all perturbations are Gaussian random variables at any time. The various linear equations of motion (linearised Einstein, linearised Klein-Gordon, continuity, Euler...) allow us to compute the evolution of the root mean square of each perturbations (the square rot of the variance) as well as the relation between the variance of the different quantities. One often says that in a stochastic theory described by linear equations of evolution, there is a "linear transport of probability", which cannot change the shape of of the distributions of probability. For a probability with zero mean, the evolution can only rescale the root mean square of a perturbation $A(t, \vec{k})$ by a factor $\alpha(t)$, its variance by $\alpha^2(t)$, and more generally, $\langle |A|^n \rangle$ by a factor $\alpha^n(t)$. For a Gaussian distribution, knowing the evolution of the root mean square (or of the variance) is sufficient. $\alpha(t)$ is just given by the solution of the equations of motion.

The evolution of each stochastic variable $A(t, \vec{k})$ can be conveniently decomposed into a stochastic number at initial time, plus a function of time:

$$A(t, \vec{k}) = A(t_i, \vec{k})\, \alpha(t) \ . \tag{4.40}$$

[2] For a more detailed discussion, but still not too technical, you may look at "Inflationary Cosmology" (lecture notes from my 2006 course at EPFL, https://lesgourg.github.io/courses/Inflation_EPFL.pdf, sections 2.4 - 2.6), or at "Inflation" by D. Baumann, arXiv:[0907.5424] (Chapter 2).

Here, we decomposed the gaussian stochastic time-dependent variable $A(t,\vec{k})$ into a gaussian stochastic variable at initial time, $A(t_i,\vec{k})$, plus a deterministic function $\alpha(t)$ which is the solution of the equations of motion correctly normalised to the initial condition $\alpha(t_i)=1$, in order to satisfy

$$\langle|A(t_i,\vec{k})|^2 = \langle|A(t_i,\vec{k})|^2\rangle\,\alpha^2(t_i)\ . \tag{4.41}$$

In our actual problem, we know the variance of the inflaton field $\delta\varphi$ throughout inflation, and in particular, in the super-Hubble regime at the end of inflation. Then, we want to know the variance of the matter fields (δ_x, δ_x, $c^2_{\mathrm{s},x}$, σ_x) after inflation, in particular, in the super-Hubble regime at the beginning of radiation domination. This will provide us with the correct in initial condition for studying CMB anisotropies and the formation of the large scale structure of the universe. But at the end of inflation, the inflaton decays into matter fields. So, do we need to study the details of this process in order to relate e.g. each δ_x during RD to $\delta\varphi$ during inflation? Fortunately, the answer is no, because there exists a linear combination of all perturbations that is conserved in the super-Hubble regime under some very generic conditions. This is the curvature perturbation $\mathcal{R}$, which is related to other perturbations in the newtonian gauge through:

$$\mathcal{R} \equiv \psi - \frac{1}{3}\frac{\delta\rho_{\mathrm{tot}}}{\bar{\rho}_{\mathrm{tot}}+\bar{p}_{\mathrm{tot}}}\ . \tag{4.42}$$

The fact that $\mathcal{R}$ is conserved on super-Hubble scales and under very generic assumptions will be proved in the next subsection.

We won't show it here, but there is a way to define $\mathcal{R}$ as a gauge-independent combination of metric and matter fluctuations, such that it takes the above expression in the newtonian gauge. However, in another gauge called the total comoving gauge (in which the total energy transfer flux δT_{0i} vanishes), $\mathcal{R}$ coincides with the perturbation of the spatial radius of curvature of the universe. Thus, in any gauge, $\mathcal{R}$ is usually called the *curvature perturbation.*

The crucial point is that this quantity is also conserved on super-Hubble scales at the transition between inflation and RD. Thus it allows to related trivially the perturbations before and after inflation. Knowing the root mean square (or the variance) of $\delta\varphi$, ϕ and ψ at the end of inflation in the super-Hubble regime, one can use equation (4.42) to compute the root mean square of $\mathcal{R}(\vec{k})$. But this root mean square is the same at the beginning of RD. Then one can use the inverse of equation (4.42) to compute matter density fluctuations at the beginning of RD. We will admit that a detailed study of perturbations during inflation, based mainly on equation (4.39) for the initial quantisation, on all the linearised equations of motion for the later evolution, and finally on equation (4.42), gives the following prediction for the super-Hubble variance of the curvature perturbation:

$$\langle|\mathcal{R}(\vec{k})|^2\rangle = \frac{256\pi^3}{3k^3}\frac{V(t_k)^3}{M_P^6 V'(t_k)^2}\ . \tag{4.43}$$

Here, t_k means "the time when the wavevectors $\vec{k}$ of wavenumber k crossed the Hubble radius during inflation", that is, the time when $k\simeq aH$. So $V(t_k)$ stands for "the inflaton potential when $k\simeq aH$" and $V'(t_k)$ for "the inflaton potential derivative $\frac{\partial V}{\partial\varphi}$ when $k\simeq aH$". The symbol $\simeq$ is not very precise but this does not matter, because quantities like $V(t_k)$ and $V'(t_k)$ vary very slowly with time during the slow-roll regime.

We still have a problem to solve. We now have a prediction for the variance of $\langle|\mathcal{R}(\vec{k})|^2\rangle$ that is valid at the beginning of RD in the super-Hubble regime, but equation (4.42) only relates $\mathcal{R}$ to ψ and $\delta\rho_{\rm tot}$. So how can we infer the variance of the individual free functions of each species, $(\delta_x,\, \delta_x,\, c^2_{{\rm s},x},\, \sigma_x)$, to the variance of $\mathcal{R}$? We will solve this in the next section.

4.3.3 Adiabatic initial conditions on super-Hubble scales

Let us consider a toy model for a perturbed universe, in any perturbed quantity $A(t,\vec{x})$ (which could be anything: density, metric coefficient, etc.) is related to its homogeneous background solution $\bar{A}(t)$ via a single time-shift function $\delta t(\vec{x})$. A priori, there is no reason that the universe is as simple as this. But if it it was, we would have:

$$A(t,\vec{x}) = \bar{A}(t+\delta t(\vec{x})) = \bar{A}(t) + \dot{\bar{A}}(t)\,\delta t(\vec{x}) + \mathcal{O}(\delta t(\vec{x})^2)\ , \tag{4.44}$$

and thus, in linear perturbation theory, we would have

$$\delta A(t,\vec{x}) = \dot{\bar{A}}(t)\,\delta t(\vec{x})\ . \tag{4.45}$$

After giving a second thought, we realise that there *is* a good reason for arguing that we live in such a universe. If we study the sector of Bardeen scalar perturbations and if we assume that *all* these perturbations originate entirely from just the fluctuation of the inflaton field during inflation, $\delta\varphi(\vec{x})$, then we see that this fluctuation does play the role of a unique time-shifting function, and that all perturbations must indeed be of the previous form! This famous assumption (which can also be challenged and tested against observations) will make our life considerably easier, because it means that all scalar fluctuations are fully *correlated*. This brings many simplification. And most importantly, we will see that this assumption is also confirmed experimentally (through the observation of CMB acoustic peaks).

However this simplified mathematical description is expected to describe very well the evolution of perturbations on super-Hubble scales. Indeed, when particles are created after inflation, they start to interact in many ways (at least with electroweak, strong and gravitational interactions). These interactions have an impact on scales that are in causal contact, where the causal horizon is defined with respect to a time at the beginning of Radiation Domination. We know that this causal horizon is given in order of magnitude by the Hubble radius. Thus, on super-Hubble scales, the perturbations of different species can be thought to be decoupled and the previous picture applies. On sub-Hubble scales, interactions change the evolution and we should no longer use equations like (4.44, 4.45).

Then we can use equation (4.45) for the density perturbation of any species X,

$$\delta\rho_X(t,\vec{x}) = \rho_X(t,\vec{x}) - \bar{\rho}_X(t) = \dot{\bar{A}}(t)\,\delta t(\vec{x}) = -3\frac{\dot{a}}{a}(\bar{\rho}_X(t)+\bar{p}_X(t))\,\delta t(\vec{x})\ . \tag{4.46}$$

We group all the terms that depend on the index X on the left,

$$\forall X,\qquad \frac{\delta\rho_X(t,\vec{x})}{\bar{\rho}_X(t)+\bar{p}_X(t)} = -3\frac{\dot{a}}{a}\,\delta t(\vec{x})\ . \tag{4.47}$$

Since the right-hand side is independent of X, this ratio is the same for all species X, and thus, for any pair (X,Y), we have

$$\forall (X,Y),\qquad \frac{\delta\rho_X(t,\vec{x})}{\bar{\rho}_X(t)+\bar{p}_X(t)} = \frac{\delta\rho_Y(t,\vec{x})}{\bar{\rho}_Y(t)+\bar{p}_Y(t)}\ , \tag{4.48}$$

or equivalently

$$\forall(X,Y), \qquad \frac{\delta_X(t,\vec{x})}{1+w_X} = \frac{\delta_Y(t,\vec{x})}{1+w_Y(t)} \ . \tag{4.49}$$

In a universe containing non-relativistic baryons and CDM with $w = 0$, and ultra-relativistic photons and neutrinos with $w = 1/3$, this gives a simple relation,

$$\delta_{\rm b} = \delta_{\rm cdm} = \frac{3}{4}\delta_\gamma = \frac{3}{4}\delta_\nu \ , \tag{4.50}$$

that must be satisfied at any time by super-Hubble Fourier modes.

Besides, you will prove in the exercises that equation (4.50) guarantees that the sound speed of the multi-fluid made up of baryons, cold dark matter, photons and neutrinos has a very peculiar sound speed, which is the same in every point, and whose square is given by

$$\frac{\delta p_{\rm tot}(t,\vec{x})}{\delta\rho_{\rm tot}(t,\vec{x})} = \frac{\dot{\bar{p}}_{\rm tot}(t)}{\dot{\bar{\rho}}_{\rm tot}(t)} \ . \tag{4.51}$$

The square root of the ratio of the right-hand side is called the adiabatic sound speed of the multi-fluid. Any fluid with $\frac{\delta p}{\delta\rho} = \frac{\dot{\bar{p}}}{\dot{\bar{\rho}}}$ is said to have adiabatic fluctuations. Thus, in cosmology, the famous relation of equation (4.50) is referred to as "adiabatic initial conditions". (Here "initial conditions" means "conditions imposed on fluctuations at the beginning of radiation domination, on super-Hubble scales").

In the last sub-section, we said that the curvature perturbation $\mathcal{R}$ is conserved on Super-Hubble scales. This is actually true only in a universe with adiabatic initial condition.[3] However, the standard cosmological model always assumes adiabatic initial conditions, since this has been proved to be correct by observations.

In the exercises, you have used the linearised Einstein equation $G^0_0 = 8\pi T^0_0$ and $G^i_i = 8\pi T^i_i$ to prove that on Super-Hubble scales, during both RD and MD, the fluctuations $\delta_{\rm tot} = -2\phi = -2\psi$ is constant in time (up to a decaying mode that can be neglected). We will now put together all the information we have from the adiabatic initial condition (4.50), the definition of the curvature perturbation (4.42), and the fact that $\delta_{\rm tot} = -2\phi = -2\psi$=constant. We will study the consequences of these relations during RD and during MD on super-Hubble scales.

First, during RD, that is, when $\bar{\rho}_{\rm b}$ and $\bar{\rho}_{\rm cdm}$ are negligible compared to $\bar{\rho}_\nu$ and $\bar{\rho}_\gamma$:

- The total relative density perturbation is given approximately by

$$\begin{aligned}
\delta_{\rm tot} &= \frac{\delta\rho_X}{\bar{\rho}_X} = \frac{\sum_X \delta\rho_X}{\sum_X \bar{\rho}_X} = \frac{\sum_X \bar{\rho}_X\delta_X}{\sum_X \bar{\rho}_X} \\
&= \frac{\bar{\rho}_{\rm b}\delta_{\rm b} + \bar{\rho}_{\rm cdm}\delta_{\rm cdm} + \bar{\rho}_\gamma\delta_\gamma + \bar{\rho}_\nu\delta_\nu}{\bar{\rho}_{\rm b} + \bar{\rho}_{\rm cdm} + \bar{\rho}_\gamma + \bar{\rho}_\nu} \\
&\simeq \frac{\bar{\rho}_\gamma\delta_\gamma + \bar{\rho}_\nu\delta_\nu}{\bar{\rho}_\gamma + \bar{\rho}_\nu} \\
&= \frac{\bar{\rho}_\gamma + \bar{\rho}_\nu}{\bar{\rho}_\gamma + \bar{\rho}_\nu}\,\delta_\gamma = \delta_\gamma = \delta_\nu \ .
\end{aligned} \tag{4.55}$$

[3] We can now prove this fact explicitely. Here we just give the brute-force proof, which is conceptually easy but computationally cumbersome (this calculation is not part of the programme of the course). It just uses the Friedmann equation, the linearised Einstein equation

and the fact that the total fluid has an adiabatic sound speed:

$$\begin{aligned}
\mathcal{R}' &= \phi' - \frac{1}{3}\frac{\delta\rho'_{\rm tot}}{\bar\rho_{\rm tot}+\bar p_{\rm tot}} + \frac{1}{3}\frac{\delta\rho_{\rm tot}(\bar\rho'_{\rm tot}+\bar p'_{\rm tot})}{(\bar\rho_{\rm tot}+\bar p_{\rm tot})^2} \\
&= \phi' - \frac{1}{3}\frac{\delta\rho'_{\rm tot}}{\bar\rho_{\rm tot}+\bar p_{\rm tot}} + \frac{1}{3}\frac{\delta\rho_{\rm tot}(1+c_s^2)\bar\rho'_{\rm tot}}{(\bar\rho_{\rm tot}+\bar p_{\rm tot})^2} \\
&= \phi' - \frac{1}{3}\frac{\delta\rho'_{\rm tot}}{\bar\rho_{\rm tot}+\bar p_{\rm tot}} - \frac{1}{3}\frac{\delta\rho_{\rm tot}(1+c_s^2)3\frac{a'}{a}(\bar\rho_{\rm tot}+\bar p_{\rm tot})}{(\bar\rho_{\rm tot}+\bar p_{\rm tot})^2} \\
&= \phi' - \frac{1}{3}\frac{\delta\rho'_{\rm tot}}{\bar\rho_{\rm tot}+\bar p_{\rm tot}} - \frac{a'}{a}\frac{(\delta\rho_{\rm tot}+c_s^2\delta\rho_{\rm tot})}{\bar\rho_{\rm tot}+\bar p_{\rm tot}} \\
&= \phi' - \frac{1}{3}\frac{\delta\rho'_{\rm tot}}{\bar\rho_{\rm tot}+\bar p_{\rm tot}} - \frac{a'}{a}\frac{(\delta\rho_{\rm tot}+\delta p_{\rm tot})}{\bar\rho_{\rm tot}+\bar p_{\rm tot}} \\
&= \phi' - \frac{1}{3(\bar\rho_{\rm tot}+\bar p_{\rm tot})}\left[\delta\rho'_{\rm tot} + 3\frac{a'}{a}\delta\rho_{\rm tot} + 3\frac{a'}{a}\delta p_{\rm tot}\right] \\
&= \phi' - \frac{1}{3(\bar\rho_{\rm tot}+\bar p_{\rm tot})}\left[\delta\rho'_{\rm tot} + 2\frac{a'}{a}\delta\rho_{\rm tot} + \frac{a'}{a}\delta\rho_{\rm tot} + 3\frac{a'}{a}\delta p_{\rm tot}\right] \\
&= \phi' - \frac{1}{12\pi G a^2(\bar\rho_{\rm tot}+\bar p_{\rm tot})}\left[\left(4\pi G a^2\delta\rho_{\rm tot}\right)' + \frac{a'}{a}4\pi G a^2\delta\rho_{\rm tot} + 3\frac{a'}{a}4\pi G a^2\delta p_{\rm tot}\right] \\
&= \phi' - \frac{1}{12\pi G a^2(\bar\rho_{\rm tot}+\bar p_{\rm tot})}\left[\left(-3\frac{a'}{a}\phi' - 3\left(\frac{a'}{a}\right)^2\phi\right)'\right. \\
&\qquad\qquad + \frac{a'}{a}\left(-3\frac{a'}{a}\phi' - 3\left(\frac{a'}{a}\right)^2\phi\right) \\
&\qquad\qquad \left. + 3\frac{a'}{a}\left(\phi'' + 3\frac{a'}{a}\phi' + \left(2\frac{a''}{a} - \frac{a'^2}{a^2}\right)\phi\right)\right] \\
&= \phi' - \frac{1}{12\pi G a^2(\bar\rho_{\rm tot}+\bar p_{\rm tot})}\left[-3\frac{a'}{a}\phi'' - 3\frac{a''}{a}\phi' + 3\left(\frac{a'}{a}\right)^2\phi'\right. \\
&\quad -3\left(\frac{a'}{a}\right)^2\phi' - 6\frac{a''}{a}\frac{a'}{a}\phi + 6\left(\frac{a'}{a}\right)^3\phi - 3\left(\frac{a'}{a}\right)^2\phi' - 3\left(\frac{a'}{a}\right)^3\phi \\
&\quad \left. +3\frac{a'}{a}\phi'' + 9\left(\frac{a'}{a}\right)^2\phi' + 3\frac{a'}{a}\left(2\frac{a''}{a} - \frac{a'^2}{a^2}\right)\phi\right] \\
&= \phi' - \frac{1}{12\pi G a^2(\bar\rho_{\rm tot}+\bar p_{\rm tot})}\left[-3\frac{a''}{a}\phi' + 6\left(\frac{a'}{a}\right)^2\phi'\right] \\
&= \phi' + \frac{\frac{a''}{a} - 2\left(\frac{a'}{a}\right)^2}{4\pi G a^2(\bar\rho_{\rm tot}+\bar p_{\rm tot})}\phi' \ . \qquad (4.52)
\end{aligned}$$

To further simplify, we suspect that we should use the derivative of the Friedmann equation, combined with the equation of conservation of energy:

$$\begin{aligned}
\left\{\left(\frac{a'}{a^2}\right)^2\right\}' &= \frac{8\pi G}{3}\bar\rho'_{\rm tot} \\
2\left(\frac{a''}{a^2} - 2\frac{a'^2}{a^3}\right)\frac{a'}{a^2} &= -\frac{8\pi G}{3}3\frac{a'}{a}(\bar\rho_{\rm tot}+p_{\rm tot}) \\
\left(\frac{a''}{a} - 2\frac{a'^2}{a^2}\right) &= -4\pi G a^2(\bar\rho_{\rm tot}+p_{\rm tot}) \ . \qquad (4.53)
\end{aligned}$$

Finally we get

$$\mathcal{R}' = \phi' - \frac{4\pi G a^2(\bar\rho_{\rm tot}+p_{\rm tot})}{4\pi G a^2(\bar\rho_{\rm tot}+p_{\rm tot})}\phi' = 0 \ . \qquad (4.54)$$

- The curvature perturbation is given by

$$\begin{aligned}\mathcal{R} &\equiv \psi - \frac{1}{3}\frac{\delta\rho_{\rm tot}}{\bar\rho_{\rm tot}+\bar p_{\rm tot}} = \psi - \frac{1}{3}\frac{\delta_{\rm tot}}{1+w_{\rm tot}} \\ &= -\frac{1}{2}\delta_{\rm tot} - \frac{1}{3}\frac{\delta_{\rm tot}}{1+\frac{1}{3}} = -\frac{1}{2}\delta_{\rm tot} - \frac{1}{4}\delta_{\rm tot} = -\frac{3}{4}\delta_{\rm tot}\ . \end{aligned} \tag{4.56}$$

- Putting everything together,

$$\delta_{\rm tot} = \delta_\gamma = \delta_\nu = \frac{4}{3}\delta_{\rm b} = \frac{4}{3}\delta_{\rm cdm} = -2\phi = -2\psi = -\frac{4}{3}\mathcal{R} = \text{constant}\ . \tag{4.57}$$

Thus, during RD and for any super-Hubble wavevector $\vec{k}$, all variables are linearly related to each other by simple time-independent coefficients of order one. This means that they are all statistically correlated.

Second, during MD, that is, when $\bar\rho_{\rm b}$ and $\bar\rho_{\rm cdm}$ are much larger than $\bar\rho_\nu$ and $\bar\rho_\gamma$:

- The total relative density perturbation is given approximately by

$$\begin{aligned}\delta_{\rm tot} &= \frac{\delta\rho_X}{\bar\rho_X} = \frac{\sum_X \delta\rho_X}{\sum_X \bar\rho_X} = \frac{\sum_X \bar\rho_X\delta_X}{\sum_X \bar\rho_X} \\ &= \frac{\bar\rho_{\rm b}\delta_{\rm b} + \bar\rho_{\rm cdm}\delta_{\rm cdm} + \bar\rho_\gamma\delta_\gamma + \bar\rho_\nu\delta_\nu}{\bar\rho_{\rm b}+\bar\rho_{\rm cdm}+\bar\rho_\gamma+\bar\rho_\nu} \\ &\simeq \frac{\bar\rho_{\rm b}\delta_{\rm b} + \bar\rho_{\rm cdm}\delta_{\rm cdm}}{\bar\rho_{\rm b}+\bar\rho_{\rm cdm}} \\ &= \frac{\bar\rho_{\rm b}+\bar\rho_{\rm cdm}}{\bar\rho_{\rm b}+\bar\rho_{\rm cdm}}\,\delta_{\rm b} = \delta_{\rm b} = \delta_{\rm cdm}\ . \end{aligned} \tag{4.58}$$

- The curvature perturbation is given by

$$\begin{aligned}\mathcal{R} &\equiv \psi - \frac{1}{3}\frac{\delta\rho_{\rm tot}}{\bar\rho_{\rm tot}+\bar p_{\rm tot}} = \psi - \frac{1}{3}\frac{\delta_{\rm tot}}{1+w_{\rm tot}} \\ &= -\frac{1}{2}\delta_{\rm tot} - \frac{1}{3}\frac{\delta_{\rm tot}}{1+0} = -\frac{1}{2}\delta_{\rm tot} - \frac{1}{3}\delta_{\rm tot} = -\frac{5}{6}\delta_{\rm tot}\ . \end{aligned} \tag{4.59}$$

- Putting everything together,

$$\delta_{\rm tot} = \delta_{\rm b} = \delta_{\rm cdm} = \frac{3}{4}\delta_\gamma = \frac{3}{4}\delta_\nu = -2\phi = -2\psi = -\frac{6}{5}\mathcal{R} = \text{constant}\ . \tag{4.60}$$

Thus, during MD and for any super-Hubble wavevector $\vec{k}$, all variables are also linearly related to each other by simple time-independent coefficients of order one. Again, they are all statistically correlated.

Finally, we come back to our statement that, thanks to the linearity of the equations of evolution, for any generic perturbation $A(t,\vec{k})$, the solution can be decomposed in a stochastic initial condition multiplied by a deterministic time-evolution function. We see that all super-Hubble fluctuations can be simply related to the curvature perturbation $\mathcal{R}$, for which we have predictions from the inflation era (we know that each $\mathcal{R}(\vec{k})$ is a random number, Gaussian distributed, with zero mean and with a variance given by equation (4.43)). Thus, $\mathcal{R}(\vec{k})$ can be conveniently chosen as a reference quantity in the expression of all the

perturbations with the same wavenumber $\vec{k}$. In other words, we can write for any generic perturbation A:

$$A(t,\vec{k}) = T_A(t,k)\ \mathcal{R}(\vec{k})\ , \tag{4.61}$$

where $T_A(t,k)$ is the deterministic solution of the equation of motion of A for all modes $\vec{k}$ of wavenumber k, normalised initially to $T_A(t_i,k) = A(t_i,\vec{k})/\mathcal{R}(\vec{k})$. The quantity $T_A(t,k)$ is called the transfer function of the perturbation A.

Let us illustrate this with a couple of concrete examples. Take for instance the metric fluctuation for a given Fourier mode, $\phi(t,\vec{k})$. This random number can be decomposed as

$$\phi(t,\vec{k}) = T_\phi(t,k)\ \mathcal{R}(\vec{k})\ , \tag{4.62}$$

where $T_\phi(t,k)$ is the solution of the equation of motion for ϕ, normalised on super-Hubble scales and during RD to $T_\phi(t_i,k) = \phi(t_i,\vec{k})/\mathcal{R}(\vec{k}) = \frac{2}{3}$, where we have inferred the coefficient $\frac{2}{3}$ from equation (4.57). Instead, the transfer function of the photon perturbation is normalised on super-Hubble scales and during RD to $T_\gamma(t_i,k) = \delta_\gamma(t_i,\vec{k})/\mathcal{R}(\vec{k}) = -\frac{4}{3}$, where we have used again equation (4.57).

The variance of a generic perturbation A is then given by

$$\langle|A(t,\vec{k})|^2\rangle = (T_A(t,k))^2\ \langle|\mathcal{R}(\vec{k})|^2\rangle\ . \tag{4.63}$$

As long as the mode $\vec{k}$ is super-Hubble, $(T_A(t,k))^2$ is just a numerical coefficient of order one that can be inferred from equation (4.57) or (4.60), but as soon as a mode enters the Hubble radius, it undergoes a non-trivial evolution caused by different forces (gravity, pressure, interaction between particles) that we will study in chapters 5 and 6.

In chapters 5 and 6, we will see that the most interesting cosmological observations that can be performed and compared to theoretical predictions are summarised by variances. The measurement of CMB anisotropies leads essentially to a measurement of the variance $\langle|\delta_\gamma(t,\vec{k})|^2\rangle$ (corrected by a few other variances), while the measurement of the galaxy distribution around us leads essentially to a measurement of the variance $\langle|\delta_{\rm cdm}(t,\vec{k})|^2\rangle$ (corrected by the variance of baryons). Thus, the goal of a theoretical cosmologist is to make predictions for these variances, compare them with the data, and find which cosmological model best describes our universe. But equation (4.63) shows that the work of such a theoretical cosmologist can be decomposed in two parts: first, make predictions for $\langle|\mathcal{R}(\vec{k})|^2\rangle$, and second, make predictions for $T_A(t,k)$. We have seen that the first problem is solved by inflation, which leads to the prediction of equation (4.43). The second task will be addressed by us in chapters 5 and 6.

4.3.4 Tensor perturbations

The FLRW metric fluctuation $\delta g_{ij}(t,\vec{x}) = a(t)^2 H_{ij}(t,\vec{x})$ contains a traceless transverse component $H^T_{ij}(t,\vec{x})$ that encodes the two polarisation degrees of freedom of gravitational waves. One can always define two independent fields $(h_1(t,\vec{x}),\, h_2(t,\vec{x}))$ and two independent traceless transverse matrices $(H^1_{ij},\, H^2_{ij})$ such that

$$H^T_{ij}(t,\vec{x}) = h_1(t,\vec{x})\ H^1_{ij} + h_2(t,\vec{x})\ H^2_{ij} \tag{4.64}$$

accounts for the propagation of GWs in a given direction.

The equation of motion of the fields $(h_1(t,\vec{x}),\ h_2(t,\vec{x}))$ is given by the variation of the Einstein-Hilbert action. After a long calculation, one sees that the two degrees of freedom h_1, h_2 obey to the same equation of motion,

$$h'' + 2\frac{a'}{a}h' - \Delta h = 0 \ , \tag{4.65}$$

where the prime denotes a derivative with respect to conformal time. This equation turns out to be the same as that of an inhomogeneous massless scalar field in FLRW. (In fact, in GR and for whatever metric, it is a generic fact that the same evolution equation describes scalar fields and tensor modes).

The degrees of freedom $(h_1(t,\vec{x}),\ h_2(t,\vec{x}))$ are propagating degrees of freedom that should be quantised during inflation in the same way as the inflaton perturbations. This leads to the following expression for the total variance of the tensor $H^T_{ij}(t,\vec{k})$ summed up over all indices:

$$\sum_{i,j}\langle|H^T_{ij}(t,\vec{k})|^2\rangle = \frac{256\pi^2}{3k^3}\frac{V(t_k)}{M_P^4} \ . \tag{4.66}$$

Like in equation (4.43), the notation t_k means "the time when the wavevectors $\vec{k}$ of wavenumber k crossed the Hubble radius during inflation", that is, the time when $k \simeq aH$.

4.3.5 Definition of the power spectrum

In cosmology and astrophysics, one calls "power spectrum" the variance of the Fourier modes of a field A:

$$P_A(t,k) = \langle|A(t,\vec{k})|^2\rangle \ . \tag{4.67}$$

There is a wavevector on the right-hand side because this is really the variance of each mode $\vec{k}$. But there is only a wavenumber on the left-hand side, because the assumption of a statistically isotropic universe implies that the variance is the same for all modes $\vec{k}$ with the same k. The power spectrum can be computed at each given time t, but in what follows we will omit the time argument for concision.

Lots of measurable quantities in real space derive from an integral of the type

$$M = \int \frac{d^3\vec{k}}{(2\pi)^3}\ P_A(k)\ K(k) \ , \tag{4.68}$$

where $K(k)$ is a dimensionless kernel. If A is a dimensionless perturbation, the measurable quantity M is dimensionless too. For instance:

- you will check in the exercises that with $K(k) = \frac{\sin kR}{kR}$, one gets the two-point correlation function between pairs of points separated by R, that is, $\langle A(t,\vec{x})A(t,\vec{x}')\rangle$ for any pair $\vec{x}$, $\vec{x}'$ such that $|\vec{x}' - \vec{x}'| = R$.
- Taking the limit $R \longrightarrow 0$, we see that with $K(k) = 1$, one gets the variance of the fluctuation A in real space, $\langle|A(t,\vec{x})|^2\rangle$, which must be the same at every point $\vec{x}$ as a consequence of the assumption of statistical homogeneity.
- There are many other examples that we will not use hin this lecture. For instance, with $K(k) = 3\left(\frac{\sin kR}{(kR)^3} - \frac{\cos kR}{(kR)^2}\right)$, one gets the variance of A smoothed over a spheres of radius R.

In these integrals, we have $\vec{k} = k^2\,dk\,d\Omega$. Given that the integrand does not depend on the direction, we can integrate over the solid angle and get a factor 4π. Then,

$$M = \int d\ln k \, \left(\frac{k^3}{2\pi^2} P_A(k)\right) K(k) \; . \tag{4.69}$$

The quantity between parentheses plays a particular role. It is "the contribution of each logarithmic interval" (i.e. of each scale) to the result. Intuitively, in a universe where all scales have the same weight (a bit like in a fractal picture), this quantity should be independent of k. Otherwise, the result can be dominated either by large scales or by small scales. If A is dimensionless, the quantity between parentheses must be a dimensionless quantity. It is commonly called the *dimensionless power spectrum* $\mathcal{P}_A$:

$$\mathcal{P}_A(k) = \frac{k^3}{2\pi^2} P_A(k) \; . \tag{4.70}$$

When $\mathcal{P}_A(k)$ is independent of k, the power spectrum is called "scale-invariant".

4.3.6 Primordial spectrum of scalar and tensor perturbations

According to equations (4.43) and (4.66), the power spectrum of primordial curvature perturbations and of primordial tensor perturbations (computed on super-Hubble scales) read

$$\mathcal{P}_\mathcal{R}(k) = \frac{128\pi}{3} \frac{V(\varphi(t_k))^3}{M_P^6 V'(\varphi(t_k))^2} \; , \tag{4.71}$$

$$\mathcal{P}_h(k) = \frac{128}{3} \frac{V(\varphi(t_k))}{M_P^4} \; , \tag{4.72}$$

where t_k is the time when $k = aH$. It is striking that the remaining dependance on k is only through the terms $V(\varphi(t_k))$ and $V'(\varphi(t_k))$. Due to the SR conditions, these quantities must be varying very slowly with time and with k, and thus, the two spectra must be nearly scale-invariant. Then, at leading order in a Taylor expansion of $\ln\mathcal{P}$ versus $\ln k$, one gets

$$\mathcal{P}_\mathcal{R} = A_s \, (k/k_*)^{e_s} \tag{4.73}$$

$$\mathcal{P}_h = A_t \, (k/k_*)^{e_t} \tag{4.74}$$

where k_* is an arbitrary *pivot scale* chosen for doing the expansion. Usually, people just pick up randomly one of the typical scales that we observe in CMB maps – thus the pivot scale $2\pi/k_*$ is of the same order of magnitude as the current comoving Hubble radius, such that $k_* \sim a_0 H_0$. A_s and A_t are the scalar and tensor amplitude parameters, while e_s and e_t are some scalar and tensor exponents that we are going to precise. The amplitude parameters are then given by

$$A_s = \frac{128\pi}{3} \frac{V(\varphi(t_{k_*}))^3}{M_P^6 V'(\varphi(t_{k_*}))^2} \; , \tag{4.75}$$

$$A_t = \frac{128}{3} \frac{V(\varphi(t_{k_*}))}{M_P^4} \; . \tag{4.76}$$

According to the definition of equations (4.73, 4.74), the exponents should be given by the calculation of

$$e_s = \left.\frac{d\ln\mathcal{P}_{\mathcal{R}}(k)}{d\ln k}\right|_{k_*} , \tag{4.77}$$

$$e_t = \left.\frac{d\ln\mathcal{P}_h(k)}{d\ln k}\right|_{k_*} . \tag{4.78}$$

The derivatives with respect to k can be computed bearing in mind that the spectra depend on k due to the fact that the potential and its derivatives are evaluated at the time t_k when $k = aH$. Using the chain rule and slow-roll approximate formulas, one can do these calculations explicitly. For instance, for the tensor exponent, we compute

$$\frac{d\ln\mathcal{P}_h}{d\ln k} = \frac{k}{V}\frac{dV(\varphi(t_k))}{dk} = \frac{aH}{V}\frac{dV(\varphi(t_k))}{d(aH)} = \frac{aH}{V}\frac{dV}{d\varphi}\frac{d\varphi}{dt}\frac{dt}{d(aH)} , \tag{4.79}$$

where

$$\frac{d(aH)}{dt} = \dot{a}H + a\dot{H} = a(H^2 + \dot{H}) . \tag{4.80}$$

The second term is suppressed with respect of the first one by one power of the slow-roll parameter ϵ. So, at leading order in a slow-roll expansion, we get $d(aH)/dt \simeq aH^2$. Substituting in the original expression, we now find

$$\frac{d\ln\mathcal{P}_h}{d\ln k} = \frac{aH}{V}V'\dot{\varphi}\frac{1}{aH^2} . \tag{4.81}$$

We now use the slow-roll formula $3H\dot{\varphi} = -V'$. We get:

$$\frac{d\ln\mathcal{P}_h}{d\ln k} = -\frac{1}{V}V'\frac{V'}{3H^2} . \tag{4.82}$$

With the slow-roll formula $3H^2 = 8\pi GV$, this simplifies into

$$\frac{d\ln\mathcal{P}_h}{d\ln k} = -\frac{1}{8\pi G}\frac{(V')^2}{V^2} = -2\epsilon . \tag{4.83}$$

Finally,

$$e_t = -2\epsilon(t_{k_*}) = -2\epsilon_* . \tag{4.84}$$

A similar calculation gives the scalar tilt as a function of the two slow-roll parameters:

$$e_s = -6\epsilon(t_{k_*}) + 2\eta(t_{k_*}) \equiv -6\epsilon_* + 2\eta_* , \tag{4.85}$$

By convention, people decided to use the *primordial tilts* n_s and n_t instead of the exponents e_s and e_t. They are related to each other through

$$n_s \equiv e_s + 1 , \tag{4.86}$$

$$n_t \equiv e_t . \tag{4.87}$$

There is no deep reason behind the term $+1$, only habits and conventions. So, we expect that the scalar tilt n_s is close to one, while the tensor tilt n_t is close to zero. We can also define the tensor-to-scalar ratio

$$r \equiv \frac{A_t}{A_s} = \frac{M_P^2}{\pi}\frac{V'(\varphi(t_k))^2}{V(\varphi(t_k))^2} = 16\ \epsilon(t_{k_*}) = 16\epsilon_* . \tag{4.88}$$

4.3.7 Constraints from observations

Inflation makes several non-trivial predictions that can in principle be confronted to observations:

- Negligible spatial curvature.
- Gaussian primordial fluctuations.
- Fluctuations generated extremely early, such that they already exist on super-Hubble scales during RD/MD (rather than being generated just after Hubble crossing during RD/MD).
- Adiabatic initial conditions.
- Nearly scale invariant power spectra.
- Small but non-zero amount of primordial tensor modes (i.e. GWs).

In the late XXth century, inflation was competing with other plausible mechanisms for the generation of primordial fluctuations: for instance, mechanisms based on the generation of topological defects during phase transitions. These mechanisms usually predict different properties for the fluctuations (non-gaussian, generated at Hubble crossing, non-adiabatic, significantly different from scale-invariant, etc.) But observations of the CMB and Large Scale Structures (LSS) over the period 2001-2018 has brought a strong confirmation of all the above predictions, with the exception of the last one. Indeed, current data is not sensitive enough to probe primordial GWs with a tensor-to-scalar ratio r much smaller than 0.1. People are still building more accurate experiments in order to try to detect the small tensor component.

Current observations of CMB anisotropies and LSS (e.g. galaxy and cluster maps) will be detailed in the next chapter. We will see that they probe the variance of density fluctuations, especially of photons and dark matter, over an epoch ranging from redshift $z \sim 10^3$ to $z \sim 0$. LSS can only probe primordial scalar fluctuations. In principle, CMB maps can probe both scalar and tensor fluctuations, with a possibility to measure these two contributions independently from each other.

A combination of recent CMB and LSS data[4] provides accurate measurements of the scalar amplitude and tilt:

$$A_s = (2.105 \pm 0.030) \times 10^{-9} \qquad (68\%\text{C.L.}) \ , \tag{4.89}$$

$$n_s = 0.9665 \pm 0.0038 \qquad (68\%\text{C.L.}) \ . \tag{4.90}$$

Since primordial GWs are not observed yet, we only have an upper bound

$$r < 0.1 \qquad (95\%\text{CL}) \ , \tag{4.91}$$

which can be translated into

$$\epsilon_* < 0.006 \qquad (95\%\text{CL}) \ , \tag{4.92}$$

or, together with the measurement of A_s, into $V < (2 \times 10^{16}\text{GeV})^4$ (95%CL). Combining this with the measurement of n_s, we infer that the second slow-roll parameter can only vary in the range

$$\eta = -0.010^{+0.005}_{-0.009} \qquad (68\%\text{C.L.}) \ . \tag{4.93}$$

[4]These results use $k_* = 0.05$ Mpc^{-1} and are based on Planck and BAO data from 2018, that will be explained in Chapters 5 and 6.

This is sufficient to prove that the SR parameters are at most of order 10^{-1} during inflation (or, at least, around the time at which $k_* \sim aH$), and that the inflaton potential is concave (at least in the same region).

But it would be better to measure primordial tensor modes, in order to:

- test one more predictions of inflation,
- determine the energy scale of inflation (given by $V(t_k)$ and thus by A_t),
- prove the "self-consistency condition" $n_t = -r/8$.

This is the main goal of several future CMB experiments. The same primordial GWs could also be detected in principle in future GW interferometers such as LISA or the Einstein telescope.

4.4 Study of inflationary models

4.4.1 General method

We will present a general method for the study of inflationary potential, that is, for checking that a given potential is compatible with successful inflation, and for studying the constraints set by observations on the parameters of the potential. In the next subsection, we will apply this approach to a particular example.

Let us assume that we are given a parametric form for the potential $V(\varphi)$, depending on a few unkown parameters α_i. Then, we can:

1. *Check the slow-roll conditions.* We compute $(\epsilon(\varphi), \eta(\varphi))$. Then, we find under which condition (on φ and/or on the potential parameters) inflation can take place, that is, under which condition $\epsilon(\varphi) \ll 1$ and $|\eta(\varphi)| \ll 1$ both hold. Then, we can find the condition for inflation to stop, that is, find the value $\varphi_{\rm end}$ such that at least one of the two slow-roll parameters reaches ~ 1. Depending on the potential, $\epsilon \sim 1$ or $|\eta| \sim 1$ can take place first, but in both case, this triggers the end of inflation.

2. *Solve the background evolution.* The goal is to find φ as a function of time t, or even better, as a function of the scale factor a or its logarithm $\log(a)$. We know that $\log(a)$ is called the "e-fold number" and can be defined with respect to an arbitrary time of reference. In the current context, it is particularly convenient to define N as the positive number of e-folds between a given time during inflation and the end of inflation, that is:

$$N \equiv \ln \frac{a_{\rm end}}{a} > 0 \ . \tag{4.94}$$

   With such a definition, N decreases during inflation. It reaches $N = 0$ at the end of inflation. Thus we can say that "an event takes place N e-folds before the end of inflation". Our goal is now to find the function $\varphi(N)$. Normally, one would do this by solving jointly the exact Klein-Gordon and Friedmann equations. However, we can also assume in very good approximation that the slow-roll conditions hold for any $a < a_{\rm end}$, and thus, that $3H^2 \simeq 8\pi M_P^{-2} V$ and $\dot\varphi \simeq -V'/(3H)$ (see equations (4.35, 4.36)). This gives us a way to find an approximation for $\varphi(N)$ very quickly,

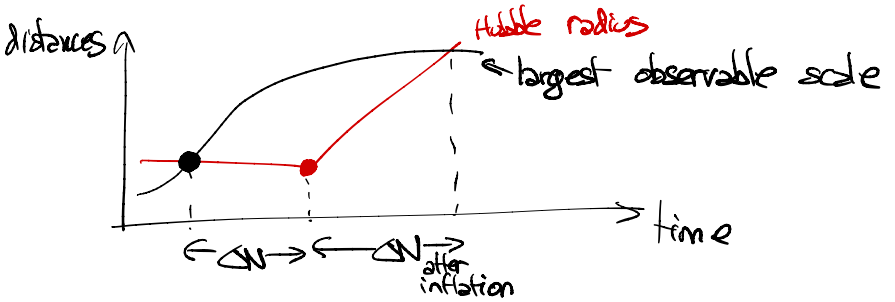

Figure 4.2: Evolution of two physical distances: the Hubble scale, and the wavelength corresponding to the whole observable universe, which is of the order of the Hubble scale today.

without solving the full Klein-Gordon equation. Indeed, we can write:

$$N = \ln a_{\rm end} - \ln a = \int_a^{a_{\rm end}} d\ln\tilde{a} = \int_t^{t_{\rm end}} \frac{da}{a\,dt}\,dt \tag{4.95}$$

$$= \int_\varphi^{\varphi_{\rm end}} H\,\frac{dt}{d\varphi}d\varphi = \int_\varphi^{\varphi_{\rm end}} \frac{H}{\dot\varphi}d\varphi\ . \tag{4.96}$$

Finally, using equations (4.35, 4.36), we get

$$N \simeq -\int_\varphi^{\varphi_{\rm end}} \frac{3H^2}{V'}d\varphi \simeq \int_{\varphi_{\rm end}}^{\varphi} \frac{8\pi M_P^{-2}V}{V'}d\varphi = \frac{8\pi}{M_P^2}\int_{\varphi_{\rm end}}^{\varphi}\frac{V}{V'}d\varphi\ . \tag{4.97}$$

3. *Use the constraint on the scalar spectrum amplitude.* Indeed, we can now compute A_s at an arbitrary pivot scale k_*. The best is to match k_* with one of the scales observable in CMB maps. By definition, these scales are almost as large as the largest observable scale, that is, the radius of our whole observable universe. We don't know yet the value of N when this scale crosses the Hubble radius during inflation. We decide to call this unknown value N_*: we will compute it in the next step. Thus, $\varphi(N_*)$ is the value of the inflaton field when this scale crosses the Hubble radius, that is, at the time that we previously called t_{k_*}. To compute the observable scalar amplitude A_s at t_{k_*}, we need to use equation (4.75) evaluated N_* e-folds before the end of inflation. Then, the observational constraint of equation (4.89) can give us a constraint equation between the parameters of the potential and N_*.

4. *Find the number of e-folds N_*.* We keep in mind a plot like the one we did when discussing the horizon problem (see figure 4.2). Here, the growing wavelength stands for the wavelength of the observable radius of the universe, that we identify with the pivot scale k_*. We know that it crosses the Hubble radius twice: first, N_* e-folds before the end of inflation (according to the definition of N_*), and second, approximately today (since the radius of the observable universe coincides with the current Hubble scale, as studied in Chapter 1). Moreover, the calculations of section 4.1.2 on the horizon problem proved a very simple but important result: the number of e-folds between the first Hubble crossing and the end of inflation, N_*, must be equal to the number of e-folds between the end of

inflation and today.[5] Thus we can write

$$N_* = \ln \frac{a_0}{a_\mathrm{f}} = \ln \frac{\rho_{\mathrm{r,beg}}^{1/4}}{\rho_{\mathrm{r},0}^{1/4}} \ , \tag{4.98}$$

where $\rho_{r,\mathrm{beg}}$ is the radiation density at the very beginning of radiation domination, $\rho_{\mathrm{r},0}$ is the radiation density today, and we used the fact that $\rho_\mathrm{r} \propto a^{-4}$. The first value of the radiation density can be identified with $V(\varphi_\mathrm{end})$, because of the continuity of the total density in the universe at the transition between inflation and radiation domination. The second value is known to be of the order of $(10^{-3}\mathrm{eV})^4$.[6] Then we get:

$$N_* \sim \ln \frac{V(\varphi_\mathrm{end})^{1/4}}{(10^{-3}\mathrm{eV})} \ . \tag{4.99}$$

Since we computed φ_end in step 1, we can use this expression to find N_*.

5. *Use the remaining constraints on the parameters* (n_s, r). Knowing N_*, we can compute the scalar tilt from (4.85, 4.86) and the tensor-to-scalar ratio from (4.88). Finally, we can see if the results match the observational bounds of equations (4.90, 4.91).

4.4.2 A simple example: the quadratic inflaton potential

We now assume explicitely the simplest possible potential: a quadratic potential, that is, a mass term, with no additional self-coupling terms:

$$V(\varphi) = \frac{1}{2} m^2 \varphi^2 \ . \tag{4.100}$$

We follow the steps presented in the previous subsection:

1. *Check the slow-roll conditions.*

$$\epsilon = \frac{M_P^2}{16\pi}\left(\frac{V'}{V}\right)^2 = \frac{M_P^2}{16\pi}\left(\frac{2}{\varphi}\right)^2 = \frac{1}{4\pi}\left(\frac{M_P}{\varphi}\right)^2 \ , \tag{4.101}$$

$$\eta = \frac{M_P^2}{8\pi}\frac{V''}{V} = \frac{M_P^2}{8\pi}\frac{2}{\varphi^2} = \epsilon \ . \tag{4.102}$$

So, coincidentally, for this particular potential, the two slow-roll parameters are equal. We see that inflation requires $\epsilon \ll 1 \Leftrightarrow |\varphi| \gg (4\pi)^{-1/2} M_P$. Note that this only gives a condition on the value fo the field, not on the value of the mass. Inflation ends when $\epsilon = 1 \Leftrightarrow |\varphi| = (4\pi)^{-1/2} M_P$. We now understand that there are two possibilities for inflation:

- Either the field starts from a positive value $\varphi \gg (4\pi)^{-1/2} M_P$ and rolls down the potential towards the minimum, with $\dot\varphi < 0$. Inflation ends when $\varphi = (4\pi)^{-1/2} M_P$. Then, the field oscillates around zero, which leads to reheating.
- Or the field starts from a negative value $\varphi \ll -(4\pi)^{-1/2} M_P$ and rolls down the potential towards the minimum, with $\dot\varphi > 0$. Inflation ends when $\varphi = -(4\pi)^{-1/2} M_P$. Then, in the same way, the field oscillates around zero, which leads to reheating.

[5] Strictly speaking, we proved this result in the case of exact exponential (De Sitter) inflation. But it remains a very good approximation for other models of inflation.

[6] as we know, a more precise number can be inferred from the value of ω_r in equation (2.63).

These two branches follow from the symmetry $\varphi \leftrightarrow -\varphi$ of our potential. They are strictly equivalent, and we will assume for simplicity that we are in the first case, with a positive φ and a negative $\dot{\varphi}$.

2. *Solve the background evolution.* We start from equation (4.97),

$$N = \frac{8\pi}{M_P^2}\int_{\varphi_{\rm end}}^{\varphi}\frac{V}{V'}d\varphi = \frac{8\pi}{M_P^2}\int_{\varphi_{\rm end}}^{\varphi}\frac{\varphi}{2}d\varphi = \frac{2\pi}{M_P^2}\left(\varphi^2-\varphi_{\rm end}^2\right) . \quad (4.103)$$

We can replace $\varphi_{\rm end}$ by $(4\pi)^{-1/2}M_P$:

$$N = \frac{2\pi}{M_P^2}\varphi^2 - \frac{1}{2} . \quad (4.104)$$

From the discussion of section 4.1.1, we expect that when observable scales cross the Hubble radius, N is in the rage from 37 to 67, so the factor $\frac{1}{2}$ can be safely neglected:

$$N \simeq \frac{2\pi}{M_P^2}\varphi^2 . \quad (4.105)$$

Inverting this relation, we get the kind of result that we expected:

$$\varphi(N) = \sqrt{\frac{N}{2\pi}}M_P . \quad (4.106)$$

3. *Use the constraint on the scalar spectrum amplitude.* We compute A_s at the pivot scale t_{k_*}, that is, N_* e-folds before the end of inflation:

$$\begin{aligned} A_s &= \frac{128\pi}{3M_P^6}\frac{V(t_{k_*})^3}{V'(t_{k_*})^2} , \\ &= \frac{128\pi}{3M_P^6}\frac{m^2\,\varphi(t_{k_*})^4}{8} \\ &= \frac{16\pi}{3M_P^6}m^2\,\varphi(N_*)^4 \\ &= \frac{16\pi}{3M_P^6}m^2\left(\frac{N_*}{2\pi}\right)^2 M_P^4 \\ &= \frac{4\,m^2\,N_*^2}{3\pi M_P^2} . \end{aligned} \quad (4.107)$$

We now impose the constraint $A_s \simeq 2\times 10^{-9}$, and invert the previous relation to get a constraint on the mass:

$$m = \left(\frac{3\pi A_s}{4}\right)^{1/2}\frac{M_P}{N_*} = \left(\frac{3\pi\,10^{-9}}{2}\right)^{1/2}\frac{M_P}{N_*} \sim \frac{10^{-4.5}}{N_*}M_P . \quad (4.108)$$

4. *Find the number of e-folds N_*.* We know that

$$V(\varphi_{\rm end}) = \frac{1}{2}m^2\varphi_{\rm end}^2 . \quad (4.109)$$

Replacing m and $\varphi_{\rm end}$ by the values that we just computed, we get

$$\begin{aligned} V(\varphi_{\rm end}) &= \frac{1}{2}\left(\frac{10^{-4.5}}{N_*}M_P\right)^2\left(\frac{M_P}{(4\pi)^{1/2}}\right)^2 = \frac{10^{-9}}{8\pi N_*^2}M_P^4 \sim \frac{10^{-10}}{N_*^2}M_P^4 \\ &= \frac{10^{-10}}{N_*^2}(10^{19}{\rm GeV})^4 = \frac{10^{-10}}{N_*^2}(10^{28}{\rm eV})^4 . \end{aligned} \quad (4.110)$$

We can substitute this result in equation (4.99):

$$\begin{aligned} N_* &\simeq \ln \frac{V(\varphi_{\rm end})^{1/4}}{(10^{-3}{\rm eV})} = \ln\left(\frac{10^{-2.5}}{\sqrt{N_*}}\frac{10^{28}{\rm eV}}{10^{-3}{\rm eV}}\right) \\ &= \ln\left(10^{28.5}\right) - \frac{1}{2}\ln N_* \simeq 65 - \frac{1}{2}\ln N_* \ . \end{aligned} \tag{4.111}$$

The result of this equation is $N_* \simeq 63$. Thus we have proved that in this particular model of inflation, the observable scales cross the Hubble radius approximately 63 e-folds before inflation. Knowing N_*, we can determine completely the value of the mass m:

$$m \simeq \frac{10^{-4.5}}{N_*} M_P = \frac{10^{-4.5}}{63} M_P \simeq 10^{-6} M_P \simeq 10^{13}{\rm GeV} \ . \tag{4.112}$$

In this particular model, a necessary condition for successful inflation is to have an inflation mass of the order of 10^{13}GeV (so, considerably heavier than, for instance, the electroweak Higgs field detected at the LHC. Such masses could appear in completely new sectors of particle physics, which could be important at very high energies, and which would have decayed completely in the early universe).

5. *Use the remaining constraints on the parameters* (n_s, r). Since we know the value of all model parameters like m and N_*, we can compute the value of:

$$\epsilon_* = \frac{1}{4\pi}\left(\frac{M_P}{\varphi(N_*)}\right)^2 = \frac{1}{4\pi}\frac{2\pi}{N_*} = \frac{1}{2\,N_*} \simeq \frac{1}{126} \tag{4.113}$$

$$\eta_* = \epsilon_* \tag{4.114}$$

$$n_s = 1 - 6\epsilon_* + 2\eta_* \simeq 1 - \frac{4}{126} = 0.968 \tag{4.115}$$

$$r = 16\epsilon_* = \frac{16}{126} \simeq 0.13 \tag{4.116}$$

We find that this value of n_s is in very good agreement with the observational bounds of equation (4.90), but the value of r is too large compared to the bounds of equation (4.91). Thus, the model is excluded because it predicts a too high level of tensor modes, that would have already been observed in current CMB experiments. Since for a fixed amplitude A_s, a high r implies a high A_t and thus a high $V(t_{k_*})$, another way to say this is that inflation with a quadratic potential would have to take place at a too high energy scale compared to observational constraints.

We could have guessed that this model was excluded already from the fact that we mentioned in section 4.3.7 that only the potentials that are concave around $\varphi \simeq \varphi(N_*)$ can pass observational tests. The quadratic potential is obviously convex everywhere. We studied it anyway, because it is simple and pedagogical.

Several other potentials with the required property of concavity pass the observational tests successfully. We can mention at least two categories of potentials that work (with appropriate values of their free parameters), see figure 5.1. The first class of potentials is that of symmetry-breaking potentials, in which the inflaton field is rolling away from an unstable equilibrium point. Second, there are models in which the slope of the potential is dominated either by logarithmic or inverse exponential term, $V = V_0(1 + \alpha \ln(\varphi))$ or $V = V_0(1 - e^{-\alpha\varphi})$.

Figure 4.3: Examples of potential shapes $V(\varphi)$. The potential on the left is everywhere convex and fails to pass the observational tests. Typical successful potentials include the symmetry breaking potential in the middle and the potential dominated by a logarithmic term $\ln(\varphi)$ on the right.

This second class of potentials often appears, for instance, in models beyond Einstein gravity, that is, with additional corrections to the Einstein-Hilbert action (for instance, of the order of the squared Ricci scalar R^2), or with a direct coupling term between the Ricci scalar and the inflaton field.

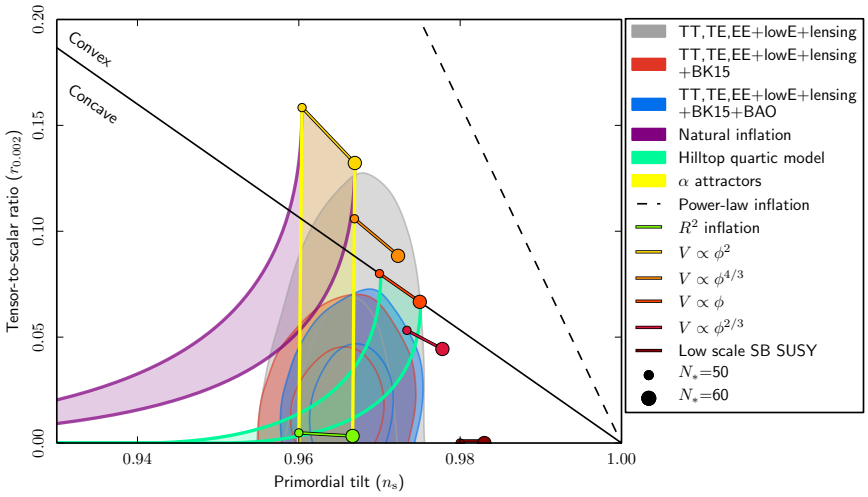

Figure 4.4: Summary of current constraints on inflation. The axes stand for the scalar tilt n_s and the tensor-to-scalar ratio r. The best current bounds are shown by the blue contours. The solid line shows the boundary between potentials that are convex and concave for $\varphi \simeq \varphi(N_*)$. The various coloured lines, dots and shaded regions correspond to the theoretical predictions far (n_s, r) for various models of inflation (after imposing already the constraint on the measured value of A_s). The model that we studied in this section is represented by the large yellow dot. The space between the two green curves corresponds to a model with a symmetry breaking potential. The two green dots and the line between them correspond to a model of modified gravity with a logarithmic shape. *Figure taken from Astron.Astrophys. 641 (2020) A10, arXiv:1807:06211.*

Chapter 5

CMB anisotropies

We know that photons decouple around $z \sim 1080$ with $T_{\rm dec} \sim 0.25\,{\rm eV} \sim 2900\,{\rm K}$, and then travel essentially in straight line up to us. Thus, they bring us an image of an abstract sphere at $z \sim 1080$, centred on our own comoving coordinates, which is called the last-scattering surface. Before decoupling, we know that photons were in thermal equilibrium with a given temperature $T(t, \vec{x})$ in every event. Because of the small density perturbations in all species including photons, there were small perturbations of temperature $\delta T(t, \vec{x})$, obeying to

$$\rho_\gamma \propto T^4 \quad \Rightarrow \quad \delta_\gamma = \frac{\delta\rho_\gamma}{\bar{\rho}_\gamma} = 4\frac{\delta T}{\bar{T}} \ . \tag{5.1}$$

After decoupling, we know that the photons keep a blackbody distribution, but get redshifted along the line of sight, approximately by $a_0/a_{\rm dec} = 1 + z_{\rm dec}$. Thus, the average temperature of the photons is expected to decrease by a factor $1+z_{\rm dec}$, down to 2.7 K, but the relative fluctuations $\frac{\delta T}{\bar{T}}$ should survive. Thus we should see small anisotropies on the map of CMB temperature, $T(\theta, \phi)$. This is indeed the case since 1992, thanks to a series of satellite (COBE, WMAP, Planck) and ground-based CMB experiments.

Since all primordial fluctuations can be described as stochastic gaussian fields, we expect that the two-point correlation function of the map of $T(\theta, \phi)$ contains all the relevant information about the model that describes these fluctuations. In Fourier space (or, since we are talking about a sphere, in multipole space), the information is contained in the variance of the modes, that is in the power spectrum of the map. We even expect that this power spectrum could be related theoretically to the spectra of primordial fluctuations $\mathcal{P}_\mathcal{R}$ and $\mathcal{P}_h$.

The goal of this chapter is to show that there is indeed such a relation, and that, in principle, the measurement of the power spectrum of the temperature anisotropy map allows us to measure:

- the spectra of primordial fluctuations $\mathcal{P}_\mathcal{R}(k)$ and $\mathcal{P}_h(k)$, and thus the parameters describing inflation: A_s, n_s, r, etc.

- the parameters that describe the cosmological model, like the Ω_i's or H_0, which play a role in the evolution of the photon perturbations between the early universe and today, and thus, in the relation between the primordial spectra and the observer temperature anisotropy spectrum.

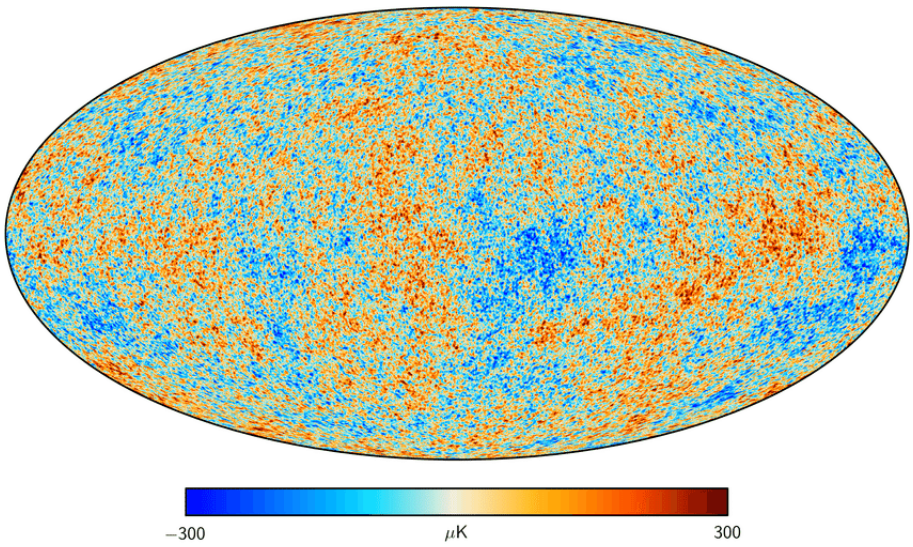

Figure 5.1: Map of CMB anisotropies $\delta T(\theta, \phi)$ measured by the Planck satellite in 2018, in units of micro-Kelvin. The data represents temperature differences on a sphere (that can be thought to be our last scattering surface). Here, the sphere has been projected on an ellipse. In this picture, the resolution has been degraded in order to reduce the size of the file. *Taken from Astron.Astrophys. 641 (2020) A1, arXiv:1807.06205; credits: ESA.*

5.1 Back to Thomson scattering

5.1.1 Diffusion rate

We have seen in Chapter 2, section 2.2.7, that until the time of recombination, there are efficient Coulomb interactions between photons and free electrons. When $T < 0.5\,\text{MeV}$, electrons are non-relativistic. Then we are in the Thomson scattering limit of Coulomb interactions: the photons are just deflected by their electromagnetic interactions with non-relativistic electrons, but the modulus of the photon and electron momentum is unaffected by the interaction. The Thomson scattering rate computed with respect to proper time reads

$$\frac{dN}{dt} = \sigma_T \, n_e^{\text{free}} = \sigma_T \, x_e n_e \ , \tag{5.2}$$

where σ_T is the Thomson scattering rate, n_e^{free} is the free electron number density, n_e is the total electron number density, and x_e is the ionisation fraction studied in section 2.2.7. Here we will call Γ the Thomson scattering rate computed with respect to conformal time η, given by

$$\Gamma_\gamma \equiv \frac{dN}{d\eta} = \frac{dN}{dt}\frac{dt}{d\eta} = \sigma_T \, a \, x_e \, n_e \ . \tag{5.3}$$

After positron annihilation, the conservation of the number of electrons implies $n_e \propto a^{-3}$. Therefore, as long as the electrons are fully ionised, Γ_γ decreases as a^{-2}. At the time of recombination, the fraction of free electrons falls exponentially, and then stabilises around a value of $^{-4}$ (see figure 5.2, top left). During the exponential decay, the scattering rate becomes smaller than the Hubble rate (figure 5.2, top right), making the photon-electron coupling inefficient and the

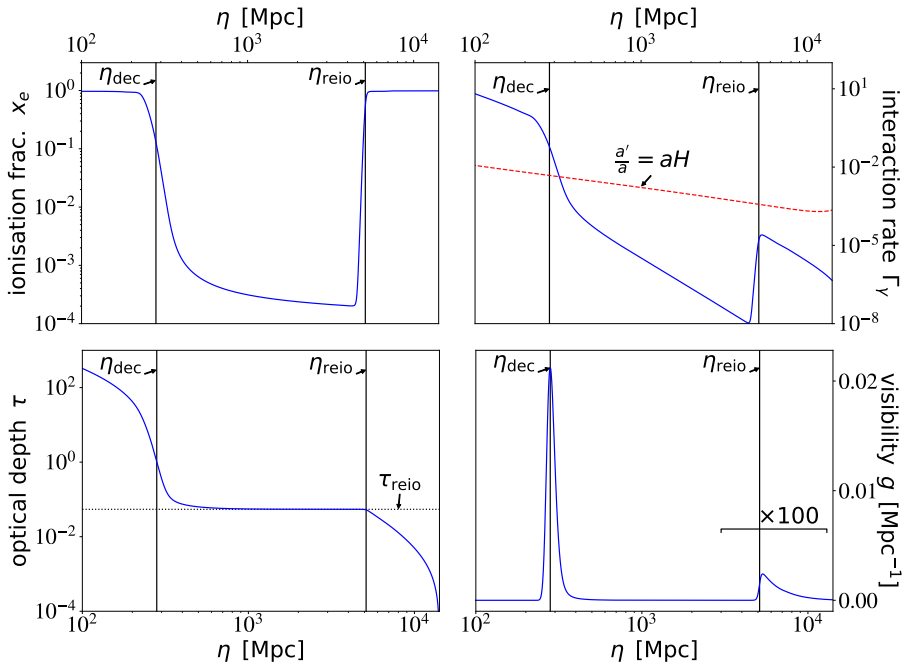

Figure 5.2: Evolution of a few relevant rates and functions from an arbitrary time before recombination until the present day: ionisation fraction x_e (top left), Thomson scattering rate Γ_γ compared to the Hubble rate in conformal time (top right), optical depth τ (bottom left), visibility function (bottom right). The vertical lines indicate the times of decoupling ($\eta_{\rm dec}$) and reionisation ($\eta_{\rm reio}$). Obtained with CLASS (`class-code.net`).

universe transparent. Most photons then travel freely and in a straight line to us. They thus transmit the image of our last scattering surface to us.

After decoupling, the interaction rate continues to decrease as $\Gamma_\gamma \propto a^{-2}$. Much later, when the first stars are formed, the fraction of free electrons increases again to one: this is the epoch of reionisation. The Γ_γ rate also increases, but without reaching the Hubble rate (figure 5.2, top right), because the electrons are now too diluted. The universe thus remains transparent, but with a small but non-zero probability of Thomson interaction of the photons with the free electrons of the inter-galactic medium, which is equivalent to the presence of a very diffuse fog.

5.1.2 Optical depth

The optical depth measures the thickness of this fog. It is obtained by integrating the Thomson scattering rate between the observer and a given point in the universe that has emitted its image at a given time. The optical depth τ expressed in relation to the conformal time η is thus

$$\tau(\eta) \equiv \int_{\eta}^{\eta_0} d\eta\, \Gamma_\gamma(\eta) \ , \tag{5.4}$$

where η_0 is the conformal time today, called the conformal age of the universe. With such a definition, note that the Thomson scattering rate reads $\Gamma_\gamma = -\tau'$.

Starting from today and going back in time, we see that the optical depth increases until the beginning of reionisation, then stabilises around a plateau value due to a very low Thomson scattering rate between recombination and reionisation, and finally increase sharply at recombination (see figure 5.2, bottom left). The plateau value noted $\tau_{\rm reio}$ is called the *reionisation optical depth*. It quantifies the fog effect induced by the reionisation, which prevents us from seeing perfectly the images coming from the recombination period. However the measured value, $\tau_{\rm reio} \simeq 0.06$, is much less than one, in agreement with the fact that the universe is strongly transparent since the recombination.

5.1.3 Visibility function

The visibility function $g(\eta)$ represents the probability that a CMB photon observed today had its last interaction with an electron at the time η. A statistical argument shows that there is a simple relationship between this function and the optical depth:

$$g(\eta) \equiv -\tau' e^{-\tau} \ . \tag{5.5}$$

This probability is correctly normalised to one:

$$\lim_{\eta\to 0} \int_{\eta}^{\eta_0} g(\eta) = e^{-\tau(\eta_0)} - \lim_{\eta\to 0} e^{-\tau(\eta)} = 1 - 0 \ . \tag{5.6}$$

The evolution of the Thomson scattering rate described above is such that the visibility function has two peaks (see figure 5.2, bottom right): a narrow and very high recombination peak, and a broad but much lower reionisation peak (on the figure, its amplitude is multiplied by 100). Between these peaks, the interaction rate is negligible, and the visibility function is zero. The probability that a CMB photon experienced its last interaction either during recombination or during reionisation is given by the integral of $g(\eta)$ under each of the peaks. The integral of the reionisation peak gives a probability of $1 - e^{-\tau_{\rm reio}} \simeq 6\%$: in other words, about 94% of the CMB photons have travelled to us without interacting.

5.1.4 Diffusion length

As the decoupling of the photons is not instantaneous, their mean free path gradually grows from zero to infinity at the moment of recombination. Each Thomson scattering points them in a random direction, so that their trajectory resembles a Brownian motion. The scattering length of the photons is defined as the distance they travel between an arbitrary time $\eta_{\rm ini}$ chosen well before recombination and a later time η. The finite value of this length just before decoupling plays an important role in the CMB spectrum. The comoving scattering length r_d obeys to a first approximation a random walk law,

$$r_{\rm d}^2(\eta) \simeq \int_{\eta_{\rm ini}}^{\eta} d\tilde{\eta}\ \Gamma_\gamma(\tilde{\eta})\ r_{\rm m.f.p.}^2(\tilde{\eta}) \ , \tag{5.7}$$

where $r_{\rm m.f.p.} = c/\Gamma_\gamma$ is the comoving mean free path of the photons, i.e. the average distance travelled between two Thomson interactions, which occur at a frequency given by Γ_γ. Finally the physical scattering length $\lambda_{\rm d}$ is given by:

$$\lambda_{\rm d}(\eta) = a(\eta)\, r_{\rm d}(\eta) \simeq a(\eta) \left[\int_{\eta_{\rm ini}}^{\eta} d\tilde{\eta}\ \Gamma_\gamma^{-1}(\tilde{\eta}) \right]^{1/2} \ . \tag{5.8}$$

5.2 Temperature anisotropy in a given direction

5.2.1 The linearised Boltzmann equation

We have seen in section 3.6 that the continuity and Euler equations do not provide complete equations of evolution for the photon perturbations. Indeed, decoupled photons should have an anisotropic stress σ_γ. Thus the continuity and Euler equations do not form a closed system, and it is necessary to use the full Bolztmann equation

$$\frac{d}{d\eta} f_\gamma = C\left[f_\gamma, f_e\right] \ , \tag{5.9}$$

where $f_\gamma(t, \vec{x}, \vec{p})$ is the phase-space distribution of photons in the perturbed universe, and $C\left[f_\gamma, f_e\right]$ stands for the Thomson scattering term that couples photons with electrons (and thus, implicitly, with baryons, since electrons and baryons form a single fluid). In any Boltzmann equation like (5.9), the total derivative $\frac{d}{d\eta}$ is supposed to be evaluated along geodesics – in this case, along any photon geodesic in the perturbed FLRW universe. For $C = 0$, this equation would simply express the conservation of the number of photons along each geodesic. Our goal will be to find a linearised version of equation (5.9), accounting for the evolution of small photon perturbations $\delta f_\gamma(t, \vec{x}, \vec{p})$.

The first step is to study the geodesic equation that gives the evolution of the photon 4-momentum P^μ defined in the FLRW coordinate system, or, if one prefers, of the *physical* 4-momentum $p^\mu = (E, p^i)$ that would be *measured by comoving observers*. The physical 3-momentum can be parametrised as $p^i = p\,\hat{n}$, where the unit vector $\hat{n}$ gives the direction of propagation. The geodesic equation can be used to compute the evolution of both $p(\eta)$ and $\hat{n}(\eta)$ along geodesics at first order in metric perturbations. The derivative $\frac{d\hat{n}^i}{d\eta}$ is useful to understand how metric fluctuations deflect photons, in order to study gravitational lensing. However, this will not be relevant in this section, as we shall see below. The derivative $\frac{dp}{d\eta}$ shows how metric fluctuations change the energy of photons – producing a local red-shifting or blue-shifting of the photons. This derivative is computed in one of the exercise sheets, and reads

$$\frac{d\ln(a\,p)}{d\eta} = \phi' - \hat{n}\cdot\vec{\nabla}\psi \ . \tag{5.10}$$

Equation (5.10) indicates firstly that in the absence of fluctuations in the metric, the momentum would simply be redshifted due to the expansion stretching the wavelengths and reducing the energy of the photons, $p \propto a^{-1}$. Fluctuations in the metric modulate this average evolution. The dilation effect associated with ϕ' represents a local fluctuation of the expansion rate, and thus of the stretching effect. The gravitational Doppler effect associated with $\hat{n}\cdot\vec{\nabla}\psi$ represents the energy gains and losses experienced by photons falling into or leaving a gravitational potential well.

In the primordial universe, photons are in thermal and chemical equilibrium at every point, with a Bose-Einstein distribution of zero chemical potential, i.e. a blackbody spectrum

$$f_\gamma(\eta, \vec{x}, \vec{p}) = \frac{1}{e^{\frac{p}{T(\eta,\vec{x})}} - 1} \ , \tag{5.11}$$

where $T(\eta, \vec{x})$ is the local value of the photon temperature. This distribution is isotropic, i.e. independent of the $\hat{n}$ direction of $\vec{p}$. In the instantaneous

decoupling approximation, this blackbody distribution freezes at the time of recombination. Thereafter, T no longer has the thermodynamic interpretation of a temperature, but continues to exist as a single parameter of the blackbody distribution. For the sake of simplicity, it will continue to be called "temperature".

If the photons interact only gravitationally after decoupling, their blackbody distribution cannot be altered. This is easily deduced from the geodesic equation (5.10), which shows that even in the presence of metric fluctuations, the relative momentum shift of photons along each geodesic, $dp = dp/p$, is independent of the momentum p. Thus each momentum p evolves, as does the temperature T which represents the typical momentum of the photons, but any distortion with respect to a blackbody is impossible. On the other hand, in a given point, an observer would see photons arriving from different directions along different geodesics, and thus carrying a different temperature. The parameter T thus acquires a dependence on the direction of propagation $\hat{n}$, which means that the photons are described entirely by a variable $T(\eta, \vec{x}, \hat{n})$ (and not only $T(\eta, \vec{x})$ as in thermal equilibrium).

We can therefore insert into the Boltzmann equation (3.26) a blackbody distribution with a direction dependent temperature,

$$f_\gamma(\eta, \vec{x}, p, \hat{n}) = \frac{1}{e^{\frac{p}{T(\eta,\vec{x},\hat{n})}} - 1} \ , \tag{5.12}$$

and then develop it to first order in perturbations. After expressing the relative perturbations of the temperature parameter to first order as $\Theta(\eta, \vec{x}, \hat{n}) \equiv \delta T(\eta, \vec{x}, \hat{n})/\bar{T}(\eta)$, and after writing the first-order total derivative $d/d\eta$ in perturbations of the metric thanks to the equation of the geodesics (5.10), you will find in on of the exercise sheets that the collisionless Bolztmann equation $\frac{d}{d\eta} f_\gamma = 0$ gives at first order in perturbations

$$\Theta' + \hat{n} \cdot \vec{\nabla}\Theta - \phi' + \hat{n} \cdot \vec{\nabla}\psi = 0 \ . \tag{5.13}$$

In the steps of the exercise, you will find that this equation can be obtained without an explicit calculation of $\frac{d\hat{n}^i}{d\eta}$ along photon geodesics, because this term would contribute to the Boltzmann equation at order two in perturbations. This is very important: it shows that at leading order, CMB physics can be studied while neglecting the effect of gravitational lensing, and pretending that $\frac{d\hat{n}^i}{d\eta} = 0$. In other words, we can pretend that photons travel in straight line between the last scattering surface and the observer. We know very well how to calculate the effects of gravitational lensing, but this is a next-to-leading-order effect that can be safely neglected in a course of first introduction to the CMB.

We should now restore the collision term of equation (3.26). Repeating the same exercise without neglecting the collision term is relatively complicated, but the steps can be found in many textbooks, such as for instance [2,5,7]. We will not explain these steps here and we will just give the final linearised Boltzmann equation expressed in terms of Θ,

$$\Theta' + \hat{n} \cdot \vec{\nabla}\Theta - \phi' + \hat{n} \cdot \vec{\nabla}\psi = -\Gamma_\gamma \left(\Theta - \Theta_0 - \hat{n} \cdot \vec{v}_\mathrm{b}\right) \ , \tag{5.14}$$

where $\Gamma_\gamma = -\tau'$ is the Thomson scattering rate, $\Theta_0(\eta, \vec{x})$ is the temperature fluctuation of photons at a point averaged over all $\hat{n}$ directions,

$$\Theta_0(\eta, \vec{x}) = \int \frac{d\hat{n}}{4\pi} \Theta(\eta, \vec{x}, \hat{n}) \ , \tag{5.15}$$

and $\vec{v}_e$ is the group velocity of electrons, equal to that of baryons, $\vec{v}_e = \vec{v}_b$. The right-hand side term of the equation (5.14) is presented here in a simplified version, neglecting the angular dependence of the Thomson cross section: this approximation is sufficient at the level of this chapter, but the angular dependence must be restored to understand the CMB polarization in detail (that will be briefly introduced at the end of this chapter).

By going from (3.26) to (5.14), we have reduced the dimensionality of the problem, since the variable Θ no longer depends on the norm of the momentum p: we have used the fact that the spectrum is of the blackbody type to eliminate the momentum from the equation of motion.

This apparently complex equation is in fact very intuitive. In the absence of gravitational ($\phi = \psi = 0$) and Thomson ($\Gamma_\gamma = 0$) interactions, it is a free diffusion equation, $\Theta' + \hat{n} \cdot \vec{\nabla}\Theta = 0$, solved by plane waves propagating at the speed of light in the direction of $\hat{n}$. Gravitational interactions modulate the individual photon momentum and thus their collective temperature parameter Θ, firstly by the dilation term ϕ', and secondly by the gravitational Doppler effect $\hat{n} \cdot \vec{\nabla}\psi$. When the Thomson interactions are very effective ($\Gamma_\gamma \to \infty$), they force the temperature to find an equilibrium point such that the term in brackets cancels, $(\Theta - \Theta_0 - \hat{n} \cdot \vec{v}_\mathrm{b}) = 0$. Indeed, the interactions push the photons to form a perfect gas in equilibrium with the electron-baryon fluid, and thus with an isotropic temperature in the instantaneous comoving reference frame of this fluid, i.e. $\Theta(\eta, \vec{x}, \hat{n}) = \Theta_0(\eta, \vec{x})$. The other reference frames are obtained by a Lorentz boost according to $\vec{v}_e$, which generates a dipole in the photon temperature given by $\hat{n} \cdot \vec{v}_\mathrm{b}(\eta, \vec{x})$. Finally the equilibrium temperature is given by

$$\Theta(\eta, \vec{x}, \hat{n}) = \Theta_0(\eta, \vec{x}) + \hat{n} \cdot \vec{v}_\mathrm{b}(\eta, \vec{x}) \ . \tag{5.16}$$

5.2.2 Line-of-sight integral

The map of anisotropies visible today ($\eta = \eta_0$) from our position in the universe ($\vec{x} = \vec{o}$ with an appropriate choice of origin) looking in a direction $\hat{n}$ corresponds to the function

$$\frac{\delta T}{\bar{T}}(\hat{n}) = \Theta(\eta_0, \vec{o}, -\hat{n}) - \Theta_0(\eta_0, \vec{o}) \ , \tag{5.17}$$

since a photon observed in the $\hat{n}$ direction propagates in the $-\hat{n}$ direction. The term $-\Theta_0(\eta_0, \vec{o})$ is often omitted in the literature, but in order to be rigorous, we have to put it here, in order to reflect the fact that observers estimate $\bar{T}$ as the angular average of the CMB temperature map. Thus, by definition, the angular average of $\frac{\delta T}{\bar{T}}$ should vanish. Our goal is now to relate this quantity to the cosmological perturbations along the line of sight, and in particular to the point of intersection between this line and our last scattering surface. To do this, we will integrate the Boltzmann equation along the line of sight.

The total derivative of an arbitrary angle-dependent function $\mathcal{F}(\eta, \vec{x}, \hat{n})$ along a geodesic describing photons propagating in a direction $\hat{n}$ decomposes into

$$\frac{d}{d\eta}\mathcal{F}(\eta, \vec{x}, \hat{n}) = \mathcal{F}' + \sum_i \frac{dx_i}{d\eta}\frac{\partial \mathcal{F}}{\partial x_i} + \sum_i \frac{dn_i}{d\eta}\frac{\partial \mathcal{F}}{\partial n_i} \ . \tag{5.18}$$

In the theory of linear perturbations, if $\mathcal{F}$ is already a perturbation of order one, it is sufficient to express $\frac{dx_i}{d\eta}$ and $\frac{dn_i}{d\eta}$ at order zero. This amounts to considering unperturbed geodesics, i.e. propagation in a constant direction $\hat{n}$ with $\frac{dn_i}{d\eta} = 0$ and $\frac{dx_i}{d\eta} = n_i$. The normalisation of this last equality comes from the form of

the Friedmann metric in conformal time and natural units: $ds^2 = 0$ implies $d\eta^2 = d\vec{x}^2$. We thus reduce the equation (5.18) to

$$\frac{d}{d\eta}\mathcal{F}(\eta, \vec{x}, \hat{n}) = \mathcal{F}' + \hat{n} \cdot \vec{\nabla}\mathcal{F} \ . \tag{5.19}$$

We will now replace $\mathcal{F}(\eta, \vec{x}, \hat{n})$ by a particular combination of the optical depth τ, the temperature perturbation Θ and the metric perturbation ψ, whose integral along the line of sight is easy to calculate:

$$\mathcal{F}(\eta, \vec{x}, \hat{n}) = e^{-\tau(\eta)} \left(\Theta(\eta, \vec{x}, \hat{n}) + \psi(\eta, \vec{x})\right) \ . \tag{5.20}$$

The total derivative of this function is

$$\frac{d\left[e^{-\tau}(\Theta + \psi)\right]}{d\eta} = e^{-\tau}\left[\Theta' + \psi' + \hat{n} \cdot \vec{\nabla}(\Theta + \psi)\right] - \tau' e^{-\tau}(\Theta + \psi) \ . \tag{5.21}$$

Up to one term $(\phi' + \psi')$, we recognise between the second brackets the left-hand term of the Boltzmann equation (5.14), which we substitute for the right-hand term (while using $\Gamma_\gamma = -\tau'$ and $-\tau' e^{-\tau} = g$):

$$\frac{d\left[e^{-\tau}(\Theta + \psi)\right]}{d\eta} = g\left(\Theta_0 + \psi + \hat{n} \cdot \vec{v}_\mathrm{b}\right) + e^{-\tau}(\phi' + \psi') \ . \tag{5.22}$$

We can integrate this equation along a geodesic corresponding to a ray of light propagating in the $\hat{n}$ direction, between an arbitrary initial instant chosen well before the recombination ($\eta = \eta_\mathrm{ini}$) and today ($\eta = \eta_0$). The term on the left being a total derivative, its integration simply gives

$$\left[e^{-\tau}(\Theta + \psi)\right]_{\eta_0} - \left[e^{-\tau}(\Theta + \psi)\right]_{\eta_\mathrm{ini}} \ , \tag{5.23}$$

and using $e^{-\tau(\eta_0)} = 1$ and $e^{-\tau(\eta_\mathrm{ini})}$, we find the identity

$$\Theta|_{\eta_0} + \psi|_{\eta_0} = \int_{\eta_\mathrm{ini}}^{\eta_0} d\eta \left[g\left(\Theta_0 + \psi + \hat{n} \cdot \vec{v}_\mathrm{b}\right) + e^{-\tau}(\phi' + \psi')\right] \ . \tag{5.24}$$

For the sake of compactness, we have omitted most of the arguments in this equation, but we will now specify them. The first term is actually $\Theta(\eta_0, \vec{o}, \hat{n})$, i.e. the temperature anisotropy measured by an observer in the $-\hat{n}$ direction of the sky. The second term is the gravitational potential at our point in the universe, $\psi(\eta_0, \vec{o})$, and therefore does not depend on the direction. It is part of $\Theta_0(\eta_0, \vec{o})$ and, as such, it does not contribute to the map $\delta T/\bar{T}$. In the integral, all functions depend only on $(\eta, \vec{x})$ and are evaluated at a current point that moves along the line of sight in the $\hat{n}$ direction at the speed of light and reaches the observer today, i.e. $\vec{x}(\eta) = (\eta - \eta_0)\hat{n}$.

5.2.3 Sachs-Wolfe equation

To simplify the integral along the line of sight, we will neglect the reionization, which will be reintroduced in a later section, and we will perform an instantaneous decoupling approximation. The combination of these two hypotheses corresponds to $\Gamma_\gamma(\eta) = 0$ for $\eta > \eta_\mathrm{dec}$ and $\Gamma_\gamma(\eta) = \infty$ for $\eta < \eta_\mathrm{dec}$. This is equivalent to replacing $g(\eta)$ by the Dirac distribution $(\eta - \eta_\mathrm{dec})$, which preserves

the normalization condition $\int_{\eta_{\rm ini}}^{\eta_0} d\eta\, g(\eta) = 1$, and $e^{-\tau(\eta)}$ by the Heaviside distribution $H(\eta - \eta_{\rm dec})$, equal to 1 for $\eta \geq \eta_{\rm dec}$. In this limit, the equation (5.24) simplifies to

$$\frac{\delta T}{\bar{T}}(-\hat{n}) = [\Theta_0 + \psi + \hat{n}\cdot\vec{v}_{\rm e}]_{\rm dec} + \int_{\eta_{\rm dec}}^{\eta_0} d\eta\,(\phi' + \psi')\ , \tag{5.25}$$

where the notation $|_{\rm dec}$ is a shortcut for the argument $(\eta_{\rm dec}, (\eta_{\rm dec} - \eta_0)\hat{n})$, i.e. at the point of our last scattering surface seen in the $-\hat{n}$ direction of the sky. The functions ϕ' and ψ' remain evaluated at the current point of coordinates $(\eta, x(\eta) = (\eta - \eta_0)\hat{n})$. An equivalent form of this equation was first derived by Reiner Sachs and Arthur Wolfe in 1967.

This equation illuminates the different terms that contribute to the CMB anisotropy map. The contribution in brackets corresponds to perturbations present at a point on the last scattering surface, while the integral takes into account corrections that accumulate along the line of sight. Let us detail each of these terms.

Intrinsic anisotropy

The term $\Theta_0|_{\rm dec}$ is the most obvious one: the anisotropy observed in a given direction of the sky comes from a temperature fluctuation present at a point of the last scattering surface. Indeed, between the decoupling and today, the individual momentum of each photon is multiplied on average by a factor $(a_{\rm dec}/a_0)$. Consequently, the blackbody temperature is multiplied on average by the same factor. The relative perturbation $\Theta = \frac{\delta T}{\bar{T}}$ is left invariant by this transformation. It would therefore be logical to observe:

$$\frac{\delta T}{\bar{T}}(-\hat{n}) = \Theta_0|_{\rm dec}\ . \tag{5.26}$$

The other terms in (5.25) represent corrections to this relationship from arising from several Doppler and gravitational effects in each direction.

Doppler effect

In the instantaneous decoupling approximation, one can imagine that the CMB photons are ejected from the last scattering surface with a group velocity. Due to the Doppler effect, this velocity shifts their frequency and individual momentum p, thus contributing to the observed blackbody temperature. The relative shift is given by the group velocity of the photon-electron-baryon fluid projected along the line of sight and divided by the speed of light, i.e. in natural units $\hat{n}\cdot\vec{v}_\gamma|_{\rm dec} = \hat{n}\cdot\vec{v}_{\rm e}|_{\rm dec} = \hat{n}\cdot\vec{v}_{\rm b}|_{\rm dec}$. A speed pointing towards the observer ($\hat{n}\cdot\vec{v}_{\rm e}|_{\rm dec} > 0$) corresponds to an increase in the perceived frequency, and therefore in the measured temperature.

Sachs-Wolfe effect

Along the line of sight, the photons are continuously subjected to a gravitational Doppler effect from the term $\hat{n}\cdot\vec{\nabla}\psi$ in the geodesic equation (5.10) and in the Boltzmann equation (5.14). If the gravitational potential were static, i.e. dependent on $\vec{x}$ but not on η, the effect would be conservative: it would depend only on the difference between the values of the gravitational potential at the interaction point and at our location. We see this in equation (5.24): taking $\phi' =$

$\psi' = 0$, neglecting the Doppler effect and using the instantaneous decoupling approximation, the remaining contribution is

$$\Theta_0|_{\eta_0} = \Theta_0|_{\rm dec} - \left(\psi|_{\eta_0} - \psi|_{\rm dec}\right) \ . \tag{5.27}$$

We have seen that the term $\psi|_{\eta_0}$ is of no importance. On the other hand, the term $\psi|_{\rm dec}$ is different in each direction, and contributes to the observed anisotropy map. This term can be seen as the contribution of a gravitational Doppler effect at the point of photon emission: a photon leaving the last scattering surface from a gravitational potential well ($\psi|_{\rm dec} < 0$) loses energy to leave this well, and thus arrives with a lower temperature. This is called the Sachs-Wolfe effect.

Integrated Sachs–Wolfe effect

On the other hand, when the perturbations of the metric depend on time, the effect is no longer conservative, as is proved by the integral in equation (5.25), which depends on the value of ϕ' and ψ' at each point along the line of sight. Imagine a large gravitational potential well that is widening over time due to the gravitational collapse of structures in the universe. A photon passing through it would gain less energy entering the well than it would lose energy leaving it, as the well deepens over time. This results in a net redshift effect for the photon. This cumulative effect is taken into account in the integral $\int_{\eta_{\rm dec}}^{\eta_0} d\eta\,(\phi' + \psi')$. This is called the Integrated Sachs-Wolfe (ISW) effect.

In equation (5.25), the Integrated Sachs-Wolfe effect is the only contribution to the CMB temperature map that is not related to the quantities at photon decoupling. If this term dominated, the CMB map would not give a picture of the primordial universe, but rather of local structure formation. Fortunately, this contribution remains small because the fluctuations of the metric are almost static during matter domination (in the linear perturbation regime), as seen in the exercises (we will also see it again in the chapter on Large Scale Structures). Variations can only take place near the beginning and end of matter domination, during Λ domination and/or in the non-linear perturbation regime. The corresponding effects remain small. It is thanks to this property that the CMB map effectively represents a picture of the baby universe.

The equation (5.25) is very useful for understanding the physics of the CMB, but it only indicates the observable temperature in a given direction, whereas the theoretical predictions concern the spectrum of temperature anisotropies. We will develop this in the next section.

5.3 Spectrum of temperature anisotropies

5.3.1 Boltzmann equation in multipole space

Fourier transform

We saw in section 3.4 that in the case of scalar perturbations in the Bardeen sense, all group velocities are the gradients of scalar quantities. For electrons and baryons, according to previous definitions, $\vec{v}_{\rm b} = a^{-2}\vec{\nabla} v_{\rm b}$. Therefore, in the linearized Boltzmann equation (5.14), the direction $\hat{n}$ always forms a scalar product with the operator $\vec{\nabla}$. In Fourier space, this scalar product depends on

the cosine of the angle between the propagation direction and the wave vector, $\hat{n} \cdot \vec{k} \equiv k \cos\alpha$. Then, equation (5.14) becomes:

$$\Theta' + ik\cos\alpha\ \Theta - \phi' + ik\cos\alpha\ \psi = \tau'\left(\Theta - \Theta_0 + ik^{-1}\cos\alpha\ \theta_{\rm b}\right) , \qquad (5.28)$$

where we used equation (3.17), $\theta_b = (\Delta v_b)/a^2$, such that in Fourier space $\vec{v}_{\rm b} = i\vec{k}v_{\rm b}/a^2 = -i\vec{k}\,\theta_{\rm b}/k^2$. Although the vector $\hat{n}$ depends in general on two angles, we see that only one angle α appears in this equation. If the initial conditions for $\Theta(\eta, \vec{k}, \hat{n})$ also depends on $\vec{n}$ only through α, we can conclude that at any time, Θ only depends on α, and can be written as $\Theta(\eta, \vec{k}, \alpha)$. This is indeed the case, because, in the initial tightly coupled regime, Θ is given by equation (5.16), which reads in Fourier space:

$$\begin{aligned} & \Theta(\eta_{\rm ini}, \vec{x}, \hat{n}) &&= \Theta_0(\eta_{\rm ini}, \vec{x}) + \hat{n}\cdot\vec{v}_{\rm b}(\eta_{\rm ini}, \vec{x}) && (5.29)\\ \Rightarrow\ & \Theta(\eta_{\rm ini}, \vec{k}, \hat{n}) &&= \Theta_0(\eta_{\rm ini}, \vec{k}) - i\,\hat{n}\cdot\vec{k}\ \theta_{\rm b}(\eta_{\rm ini}, \vec{k})\,/\,k^2 && \\ & &&= \Theta_0(\eta_{\rm ini}, \vec{k}) - ik^{-1}\cos\alpha\ \theta_{\rm b}(\eta_{\rm ini}, \vec{k}) . && (5.30) \end{aligned}$$

Once more, we have reduced the dimensionality of the problem: from $f(\eta, \vec{x}, \vec{p})$ to $\Theta(\eta, \vec{x}, \hat{n})$, then $\Theta(\eta, \vec{k}, \hat{n})$, and finally $\Theta(\eta, \vec{k}, \alpha)$.[1]

Expansion in Legendre series

When a function depends on a single angle, it is useful to perform a transformation based on Legendre polynomials $P_l(x)$. We define the expansion of $\Theta(\alpha)$ into Legendre multipoles Θ_l as:

$$\Theta(\eta, \vec{k}, \alpha) = \sum_l (-i)^l (2l+1)\Theta_l(\eta, \vec{k}) P_l(\cos\alpha) . \qquad (5.31)$$

Given the orthogonality condition of Legendre polynomials,

$$\int_{-1}^{1} dx P_l(x) P_m(x) = \frac{2}{2l+1}\delta_{lm} , \qquad (5.32)$$

the inverse transformation is given by

$$\Theta_l(\eta, \vec{k}) = \frac{i^l}{2}\int_{-1}^{1} d(\cos\alpha)\,\Theta(\eta, \vec{k}, \alpha) P_l(\cos\alpha) . \qquad (5.33)$$

The prefactor $(-i)^l(2l+1)$ in (5.31) is just a convention for this expansion. It will lead to simplifications in some calculations. It is worth noting that in previous sections, we already defined a monopole Θ_0 for a general function of two angles, just taking the average over a sphere:

$$\Theta_0^{\rm sphere} = \frac{1}{4\pi}\int d\hat{n}\,\Theta(\hat{n}) = \frac{1}{4\pi}\int_0^{\pi} d\theta\,\sin\theta \int_0^{2\pi} d\phi\,\Theta(\theta, \phi) . \qquad (5.34)$$

This definition can also be applied to the particular case of a function with axial symmetry around $\vec{e}_z$, not depending on ϕ. However, in such a case, this

[1] This is a consequence of the isotropy of the FLRW background metric, which contains no preferred direction. In a Fourier expansion, the only direction that appears is that of the considered wavevector $\vec{k}$. It is therefore possible to form only one angle: the angle relative to $\vec{k}$.

previous defintion of the monopole exactly coincides with that of the $l = 0$ Legendre multipole, since:

$$\Theta_0^{\rm Legendre} = \frac{1}{2}\int_{-1}^{1} d(\cos\theta)\,\Theta(\theta) = \frac{1}{2}\int_0^{\pi} d\theta\,\sin\theta\,\Theta(\theta) \tag{5.35}$$

$$= \frac{1}{4\pi}\int_0^{\pi} d\theta\,\sin\theta\int_0^{2\pi} d\phi\,\Theta(\theta) = \Theta_0^{\rm sphere}\ . \tag{5.36}$$

The perturbations $\delta T_\gamma^{\mu\nu}$ of the photon stress-energy tensor are related to the Legendre multipoles. Indeed, in kinetic theory, $T_X^{\mu\nu}$ is obtained by integrating f over the spatial part of the 4-momentum $\vec{P}^\mu$ as

$$T_X^{\mu\nu} = \int d^3P_j\,(-g)^{-1/2}\,\frac{P^\mu P^\nu}{P^0} f_X(x^0, x^i, P_j)\ , \tag{5.37}$$

where d^3P_j is the integration volume in the space of the momentum *covector* P_j (in cartesian coordinates, $d^3P = dP_1 dP_2 dP_3$). In our case, after having replaced f_γ by its expression (5.12), T by $\bar{T}(1+\Theta)$, and Θ by its expansion in Legendre multipoles, we can integrate with respect to P_j by keeping only the terms of order one in perturbations. We can then identify an explicit expression for the four generic scalar perturbations $\{\delta\rho_\gamma, \delta p_\gamma, v_\gamma, \sigma_\gamma\}$ introduced in section 3.4, and finally with the four equivalent degrees of freedom $\{\delta_\gamma, c^2_{\gamma\,s}, \theta_\gamma, \sigma_\gamma\}$:

$$\delta_\gamma = 4\Theta_0\ , \qquad c^2_{\gamma\,s} = \frac{1}{3}\ , \tag{5.38}$$

$$\theta_\gamma = 3k\Theta_1\ , \qquad \sigma_\gamma = 2\Theta_2\ . \tag{5.39}$$

Note that the first two relations are obvious for ultra-relativistic particles in thermal equilibrium, which obey $\rho_\gamma = 3p_\gamma \propto T^4$ at each point. But these relations are also valid after photon decoupling. Using the above relationship between θ_γ and Θ_1, we can reformulate the right-hand side of our Boltzmann equation in Fourier space (5.28) as:

$$\Theta' + ik\cos\alpha\ \Theta - \phi' + ik\cos\alpha\ \psi = \tau'\left[\sum_{l\geq 2}(-i)^l(2l+1)\Theta_l(\eta,\vec{k})P_l(\cos\alpha) - \frac{i}{k}\cos\alpha\ (\theta_\gamma - \theta_{\rm b})\right]\ . \tag{5.40}$$

We see again that in the limit of a very efficient Thomson scattering, $|\tau'| \to \infty$, the Boltzmann equation tends to cancel the content of the brackets, that is, to align the group velocity of the photons with that of the electron-baryon fluid ($\theta_\gamma = \theta_{\rm b}$), which fixes the temperature dipole Θ_1, and to make the fluid isotropic beyond this dipole ($\Theta_{l\geq 2} = 0$).

Boltzmann hierarchy

The equation (5.40) can be expanded into mutipoles using the orthogonality property of Legendre polynomials. Using the expansion (5.31), the definitions $P_1(\cos\alpha) = \cos\alpha$ and $P_0(\cos\alpha) = 1$, the recurrence relation $(2l+1)xP_l(x) = (l+1)P_{l+1}(x) + lP_{l-1}(x)$[2] and the relations (5.38), one obtains an independent

[2]Note that this relation gives, for $l = 0$, $xP_0(x) = P_1(x)$, and for $l = 1$, $3xP_1(x) = 2P_2(x) + P_0(x)$.

relation for each coefficient of the Legendre expansion:

$$\delta'_\gamma + \frac{4}{3}\theta_\gamma - 4\phi' = 0 \ , \tag{5.41}$$

$$\theta'_\gamma + k^2\left(-\frac{1}{4}\delta_\gamma + \sigma_\gamma\right) - k^2\psi = \tau'(\theta_\gamma - \theta_b) \ , \tag{5.42}$$

$$\Theta'_l - \frac{kl}{2l+1}\Theta_{l-1} + \frac{k(l+1)}{2l+1}\Theta_{l+1} = \tau'\Theta_l \quad \forall l \geq 2 \ . \tag{5.43}$$

These equations form the Boltzmann hierarchy. The first two equations match perfectly the stress-energy tensor conservation equations (3.24, 3.25) of a relativistic fluid, with the addition of the Thomson diffusion term. When τ' is very large with respect to a'/a, the last equation forces Θ_l to remain zero for $l \geq 2$. When τ' decreases around the time of photon decoupling, the amplitude of the perturbations is transferred from $l = 0, 1$ (i.e. from δ_γ and θ_γ) to higher order mutipoles, thanks to the coupling between neighbouring mutipoles.

Physically, this corresponds to the fictional experience of an observer in the universe. As long as the photons are strongly coupled, the observer perceives temperature anisotropies only as a dipole corresponding to its velocity relative to the photon-electron fluid. Then, as time passes by and as the mean free path increases, the observer sees photons coming from more and more remote points, or more precisely from a spherical last scattering surface of increasing radius. These photons give him an image of this surface with inhomogeneities seen from a smaller and smaller angle, corresponding to a larger and larger multipole moment l.

The equations (5.41-5.43) are considered as the most important ones in the precise study of CMB physics. They play a central role in the numerical calculation of the CMB spectrum by so-called "Einstein–Boltzmann solvers". However, they will hardly intervene in the qualitative reasonings of the next sections.

5.3.2 CMB statistics in multipole space

Multipoles a_{lm}

To construct a quantity that can be both predicted theoretically and observed experimentally, it is useful to expand the map of CMB temperature anisotropies –introduced in equation (5.17)– in spherical harmonics[3]:

$$\frac{\delta T}{\bar{T}}(\hat{n}) = \Theta(\eta_0, \vec{o}, -\hat{n}) - \Theta_0(\eta_0, \vec{o}) = \sum_{l\geq 1}\sum_{-l\leq m\leq l} a_{lm}Y_{lm}(\hat{n}) \ . \tag{5.44}$$

The term $-\Theta_0(\eta_0, \vec{o}, -\hat{n})$ is often omitted in the literature, but in order to be rigorous, we have to put it here, in order to reflect the fact that observers estimate $\bar{T}$ as the angular average of the CMB temperature map. Thus, by definition, the angular average of $\frac{\delta T}{\bar{T}}$ should vanish.

We recall that $\hat{n}$ is a unit vector, consequently described by two angles (θ, ϕ). Thus the function $\frac{\delta T}{\bar{T}}(\hat{n})$ is defined on a sphere. The expansion in spherical harmonics plays the same role as a Fourier transformation for functions defined on a sphere: for larger l's, the multipoles a_{lm} (with $-m \leq l \leq m$) represent anisotropies at smaller angular scales. The fact that the temperature map is a

[3]By convention, we will always use the usual complex representation of spherical harmonics, such tht $Y_{lm}(\theta, \varphi) \propto e^{im\varphi}$. However, we will write both indices down, Y_{lm}, instead of using another frequent notation, Y_l^m.

real function and that $Y^*_{lm} = (-1)^m Y_{l-m}$ gives the constraint $a^*_{lm} = (-1)^m a_{l-m}$. To obtain an inverse relationship, one uses the orthogonality relationship of spherical harmonics,

$$\int d\hat{n}\, Y_{lm}(\hat{n}) Y^*_{l'm'}(\hat{n}) = \delta^K_{ll'}\delta^K_{mm'} \ , \tag{5.45}$$

where $\delta^K_{ll'}$ is the Kroenecker symbol, as well as the expansion of the temperature perturbation into Legendre multipoles (equation 5.31), the relation between Legendre polynomials and spherical harmonics,

$$P_l(\hat{n}\cdot\hat{k}) = \sum_{m=-l}^{l} \frac{4\pi}{2l+1} Y_{lm}(\hat{n}) Y_{lm}(\hat{k}) \ , \tag{5.46}$$

and the parity relation of spherical harmonic $Y_{lm}(-\hat{n}) = (-1)^l Y_{lm}(\hat{n})$. After about ten lines of calculation, we arrive at an expression for each multipole a_{lm} of the temperature map as a function of the Legendre multipole Θ_l evaluated today (by defining the unit vector $\hat{k} \equiv \vec{k}/k$),

$$a_{lm} = \sqrt{\frac{2}{\pi}} i^l \int d^3\vec{k}\ Y_{lm}(\hat{k})\, \Theta_l(\eta_0, \vec{k}) \ . \tag{5.47}$$

Spectrum C_l

We have seen in the previous section that each Fourier mode of a perturbation can be considered as a Gaussian random variable. This is the case for the multipole $\Theta_l(\eta_0, \vec{k})$ for each $\vec{k}$. The multipole a_{lm} is thus given by a sum of Gaussian independent variables. Thus, it is also a Gaussian random variable, whose properties are entirely described by a variance. The latter can be deduced from

$$\langle a_{lm} a^*_{l'm'} \rangle = \frac{2}{\pi}\ i^{l-l'} \int d^3\vec{k}\, d^3\vec{k}'\, Y_{lm}(\hat{k}) Y^*_{l'm'}(\hat{k}') \langle \Theta_l(\eta_0, \vec{k}) \Theta^*_{l'}(\eta_0, \vec{k}') \rangle \ . \tag{5.48}$$

Using the definition of the primordial spectrum and transfer functions,

$$\langle \Theta_l(\eta_0, \vec{k}) \Theta^*_{l'}(\eta_0, \vec{k}') \rangle = \Theta_l(\eta_0, \vec{k})\ \Theta_{l'}(\eta_0, \vec{k})\ \delta_D(\vec{k}' - \vec{k})\, P_R(k) \ , \tag{5.49}$$

as well as the relationships (4.70) and (5.45), we arrive at a simpler relationship for the variance[4],

$$\langle a_{lm} a^*_{l'm'} \rangle = \delta^K_{ll'}\delta^K_{mm'} \left[4\pi \int \frac{dk}{k} \Theta^2_l(\eta_0, k) \mathcal{P}_{\mathcal{R}}(k) \right] \ . \tag{5.50}$$

The quantity between the square brackets plays a fundamental role for three reasons:

- it contains all the statistical information about the CMB temperature map, since it represents the variance of Gaussian random quantities;

[4] In this calculation, the fact that $\langle a_{lm} a^*_{l'm'} \rangle$ is zero for $l \neq l'$ or $m \neq m'$ comes from the orthogonality relation. But more fundamentally, it is the consequence of the hypothesis of statistical homogeneity of the Universe – just like in Fourier space the correlation functions are zero for $\vec{k} \neq \vec{k}'$. Statistical isotropy implies that the part between square brackets can only depend on l and not on m: the statistical properties can depend on the angular scale θ considered, and thus on $l = 2\pi/\theta$, but not on the configuration and orientation of the axis system, and thus on m.

- it can be inferred from observations, since the multipoles a_{lm} can be measured (we will then denote the multipole $a_{lm}^{\rm obs}$);
- it can be calculated in a given theoretical model, since the transfer functions $\Theta_l(\eta_0, k)$ and the primordial spectrum $\mathcal{P}_\mathcal{R}(k)$ can be predicted.

This quantity is therefore the most appropriate one to test cosmological models based on the observation of CMB temperature anisotropies. It is called the spectrum of temperature anisotropies C_l,

$$C_l \equiv \langle |a_{lm}|^2 \rangle = 4\pi \int \frac{dk}{k} \Theta_l^2(\eta_0, k) \mathcal{P}_\mathcal{R}(k) \ . \tag{5.51}$$

Cosmic variance

Let us return to the meaning of the symbol $\langle ... \rangle$ in equation (5.51). In cosmological perturbation theory, each a_{lm} is considered a stochastic number, and the average is taken over all realizations of the theory, i.e. over the CMB maps of all possible universes corresponding to a given cosmological model. But we observe only one CMB map, i.e. one realisation of the theory. We cannot therefore know exactly the average theoretical value C_l. At best, we can estimate it from each observed realisation $|a_{lm}^{\rm obs}|^2$. Fortunately, all a_{lm} with l fixed and m in the interval $-l \leq m \leq l$ obey the same probability distribution, i.e. a Gaussian of variance C_l. Therefore, we have the opportunity to approximate C_l by calculating the average value of $|a_{lm}^{\rm obs}|^2$ over all m. In statistical language, one says that the theoretical C_l can be estimated from the observed multipoles using the "estimator" $\hat{C}_l(a_{lm})$:

$$\hat{C}_l(a_{lm}) \equiv \frac{1}{2l+1} \sum_{-l \leq m \leq l} |a_{lm}|^2 \ . \tag{5.52}$$

This is an unbiased estimator of the theoretical C_l, as its mean (considered again with respect to all realisations of the theory) is given by

$$\langle \hat{C}_l(a_{lm}) \rangle = \frac{1}{2l+1} \sum_{-l \leq m \leq l} \langle |a_{lm}|^2 \rangle = C_l \ . \tag{5.53}$$

The quantity $\hat{C}_l(a_{lm}^{\rm obs})$ is closer to the coefficient C_l of the underlying theory than each term $|a_{lm}^{\rm obs}|^2$ taken individually. This can be quantified by calculating the standard deviation between C_l and $\hat{C}_l$ (again considered with respect to all realisations of the theory). Using Wick's theorem, $\langle abcd \rangle = \langle ab \rangle \langle cd \rangle + \langle ac \rangle \langle bd \rangle + \langle ad \rangle \langle bc \rangle$, one can show in a few lines that:

$$\langle (\hat{C}_l - C_l)^2 \rangle \quad = \quad \frac{2}{2l+1} C_l^2 \ . \tag{5.54}$$

Therefore, the larger l is, the smaller the standard deviation is, and the more accurate the estimate of C_l is. Indeed, at smaller angular scales, we observe more $a_{lm}^{\rm obs}$ multipoles, i.e. more theory-independent realisations. This standard deviation is clearly visible in the observational results: the measured C_l are highly scattered at small l and less scattered at large l (see figure 5.3). The standard deviation of equation (5.54) is called *cosmic variance* and plays the role of a theoretical error. A given cosmological model can be considered as providing an excellent explanation of observations if the data are close to the theoretical predictions *within cosmic variance*.

Since cosmic variance is large at small l, it will always be difficult to characterise with great accuracy the aspects of the cosmological model that control the shape of C_l only at large angular scales.[5]. Fortunately, cosmic variance it is sufficiently small at large l for measuring most cosmological parameters with high accuracy. The measurement of the temperature anisotropy spectrum is said to be "cosmic variance limited" when the experimental errors on each C_l become smaller than the cosmic variance.

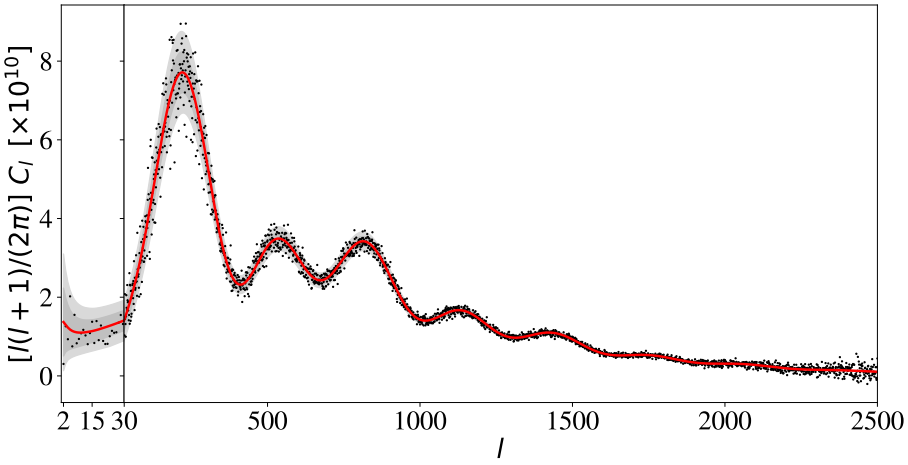

Figure 5.3: Observation of the CMB temperature anisotropies spectrum by the Planck satellite (2018 data). The points represent the measured central value for each C_l, while the curve shows the theoretical C_l for the ΛCDM model that best explains the data. The scatter of the points increases at small $l < 1500$ due to the cosmic variance, and also at large $l > 1500$ due to instrumental errors. The light (resp. dark) grey band shows the cosmic variance predicted by the theoretical model at 1σ (resp. 2σ). A different scale is used on the x-axis at $l < 30$ to highlight the first data points.

5.3.3 Line-of-sight integral in Fourier space

Equation 5.51 shows that the knowledge of the photon transfer functions at the present time, $\Theta_l(\eta_0, k)$, is of crucial importance for the theoretical calculation of the CMB temperature spectrum. These functions can be calculated by a "brute force method". Their evolution equation is given by the Boltzmann hierarchy 5.41–5.43 for each k. We can truncate this hierarchy to a multipole $l_{\text{trunc.}}$ slightly larger than the last multipole l_{max} for which we want to compute the C_l's. This gives a system of $l_{\text{trunc.}}$ coupled differential equations, which can be integrated between an initial time where $k\eta \ll 1$ (that is, in the super-Hubble regime) and the current time.

However, there is another method which is both more efficient for numerical calculations and more pedagogical for the purpose of a CMB course: the line-of-sight integral in Fourier space. This method is based on an integral similar

[5] To improve this knowledge in a really significant way, it would be necessary to travel at close to the speed of light to observe other last scattering surfaces, i.e. other independent CMB maps, and return to Earth to pool all this information.

to the one of section 5.2.2, but this time starting from the Boltzmann equation in Fourier space (5.28). It is no longer an integral along the line of sight, since the quantities are now expressed in Fourier space. Rather, it is an integral with respect to time, which allows us to express the Fourier mode $\vec{k}$ of a current multipole $\Theta_l(\eta_0, \vec{k})$ as a function of the same Fourier mode for other perturbations evaluated at various times between the decoupling of the photons and today. By writing the formal solution of the equation 5.28 in integral form, and performing Legendre transformations, we arrive at the following expression for the photon transfer function:

$$\Theta_l(\eta_0, k) = \int_{\eta_{\rm ini}}^{\eta_0} d\eta \; S_T(\eta, k) \; j_l(k(\eta_0 - \eta)) \; , \tag{5.55}$$

$$S_T(\eta, k) \equiv \underbrace{g \, (\Theta_0 + \psi)}_{\text{intrins. + SW}} + \underbrace{\left(g \, k^{-2} \theta_{\rm b}\right)'}_{\text{Doppler}} + \underbrace{e^{-\tau} (\phi' + \psi')}_{\text{ISW}} \; , \tag{5.56}$$

where $j_l(x)$ is the spherical Bessel function[6], and $S_T(\eta, k)$ is called the temperature source function.

The formal similarity between this result and the true integral along the line of sight (5.24 is obvious). Three contributions are again identified as playing the same role as in section 5.2.2: the intrinsic anisotropy corrected by the Sachs–Wolfe effect (SW), the Doppler contribution, and the integrated Sachs–Wolfe effect (ISW). This similarity is such that the equation (5.55) is also called – somewhat abusively – the line of sight integral in Fourier space.

This integral can be put into other forms by means of integrations by parts. All terms that depend on the visibility function $g(\eta)$ and its derivatives are zero today and negligible at $\eta_{\rm ini}$. Therefore, their integration by part does not generate edge terms. For example, we also have

$$\begin{aligned} \Theta_l(\eta_0, k) = \int_{\eta_{\rm ini}}^{\eta_0} d\eta \qquad & \{g \, (\Theta_0 + \psi) \; j_l(k(\eta_0 - \eta)) \\ & + g \, k^{-1} \theta_{\rm b} \; j_l'(k(\eta_0 - \eta)) \\ & + e^{-\tau} (\phi' + \psi') \, j_l(k(\eta_0 - \eta))\} \; . \end{aligned} \tag{5.57}$$

Thanks to the line-of-sight integral (5.55), to compute the transfer functions $\Theta_l(\eta_0, k)$ up to an arbitrarily large l, it is sufficient to solve the Boltzmann hierarchy just in order to obtain the transfer functions $\Theta_0(\eta, k)$, $\theta_{\rm b}(\eta, k)$, $\phi(\eta, k)$ and $\psi(\eta, k)$. The aim is therefore no longer to have a precise solution of the Boltzmann hierarchy up to $l = l_{\rm max}$, but only up to $l = 2$ (because the Einstein equation, which is necessary to obtain ϕ and ψ, involves the multipoles up to $l = 2$). For this, a truncation around $l_{\rm trunc.} \sim \mathcal{O}(10)$ is sufficient. The number of equations is thus considerably reduced compared to the brutal method.

In addition to its numerical efficiency, the line-of-sight integral gives a much better analytical insight into the shape of the final result and the dependence of the spectrum on the cosmological parameters.

Angular projection

The previous calculation involves a spherical Bessel function evaluated in $j_l(k(\eta_0 - \eta))$. Mathematically, this function appears in the calculation of the line-of-sight

[6] The spherical Bessel functions $j_l(x)$ are related to the usual Bessel functions of the first kind $J_l(x)$ by $j_l = \sqrt{\frac{\pi}{2x}} J_{l+\frac{1}{2}}(x)$

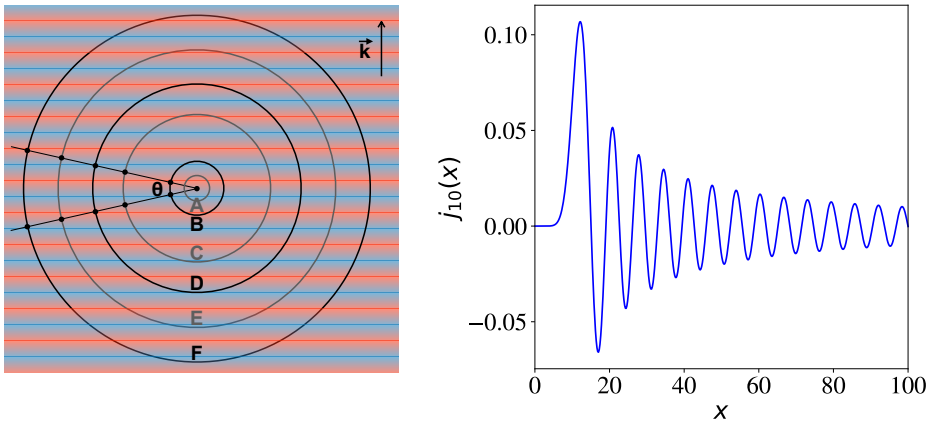

Figure 5.4: *Left:* Contribution of a Fourier mode $A(\eta, \vec{k})$ with fixed $\vec{k}$ fixed and unspecified η to a multipole $\Theta_l(\eta_0, k)$ for $l = \pi/\theta$ fixed. *Right:* Spherical Bessel function $j_{10}(x)$.

integral when we perform a Legendre transformation of the plane wave function $e^{-i\vec{k}\cdot\vec{x}}$. But more intuitively, it plays the role of a projection coefficient from Fourier space to l-multipole space. In the multipole space, each l corresponds to a configuration on a sphere such that the angle between a maximum and an adjacent minimum is given by $\theta = \pi/l$. Let us try to answer the following question: can a Fourier mode $\vec{k}$ of a given perturbation $A(\eta, \vec{k})$ contribute to the multipole l of the transfer function $\theta_l(\eta_0, k)$?

To understand this, one must have in mind the diagram on the left of figure 5.4. If the mode propagates at time η, the observer perceives a cross-sectional view along a sphere whose radius is given by the angular distance d_a until time η,

$$d_\mathrm{a} = a(t) \int_t^{t_0} \frac{dt}{a} = a(\eta) \int_\eta^{\eta_0} d\eta = a(\eta)\ (\eta_0 - \eta)\ . \tag{5.58}$$

The contribution of this Fourier mode to the l-multipole is non-zero if the observer perceives differences between the values of the perturbation at two points on the sphere separated by an angle $\theta = l/\pi$. The perceived contribution is the average over all possible directions of the difference between such pairs of points. The largest contribution to this average always comes from pairs of points arranged parallel to the wave vector, such as the vertically aligned pairs of points on the left of the figure.

There is a unique value of the distance $d_\mathrm{a}(\eta)$, and therefore of the time η, such that these points correspond to an adjacent maximum and minimum of the Fourier mode. On the figure, it is the sphere B. This value of η is the solution of the equation $\lambda/2 = \theta\, d_\mathrm{a}$, where $\lambda/2$ is the half-wavelength, related to the comoving Fourier mode k by $\lambda/2 = a(\eta)\pi/k$. Thus the contribution is maximum if the relation

$$\theta = \frac{\pi}{l} = \frac{\lambda/2}{d_\mathrm{a}} = \frac{a(\eta)\,\pi/k}{a(\eta)\,(\eta_0 - \eta)} \quad \Leftrightarrow \quad l = k(\eta_0 - \eta) \tag{5.59}$$

is satisfied.[7]

[7] If a mode of the same k propagates later, the observer perceives a cross-sectional view

In a completely equivalent way, we could have fixed the diameter of the sphere $d_a(\eta)$ and thus the time η, and looked for the values of k for which $A(\eta, \vec{k})$ gives a contribution to the multipole l. The largest contribution comes from the modes verifying $l = k(\eta_0 - \eta)$: this is the main harmonic. With larger k, we then find the first harmonic, which contributes with an opposite sign, the second harmonic, which contributes with the same sign, and so on.

The spherical Bessel function $j_l(k(\eta_0 - \eta))$ can be seen as a projection coefficient which takes all these effects exactly into account. Its characteristic form appears in the diagram on the right of figure 5.4. It always has a maximum near $k(\eta - \eta_0) = l$, giving the contribution of the main harmonic. For $k(\eta - \eta_0) < l$ it tends rapidly to zero, and for $k(\eta - \eta_0) > l$ it has an oscillatory behaviour taking into account all harmonics.

Instantaneous decoupling approximation

In section 5.2.2, we introduced a double approximation: absence of reionisation and instantaneous decoupling. This allowed us to simplify the integral (5.24) and to obtain the Sachs–Wolfe relation (5.25). This double approximation amounts to considering the photons as strongly coupled with the baryons and electrons until the moment of decoupling, $\eta_{\rm dec}$, and totally free afterwards. In this limit, all photons have their last interaction at the same time, and thus really come from a surface of last scattering (which is a two-dimensional sphere). In reality, the decoupling is progressive and the photons come from a last scattering area with a certain thickness. Moreover, some photons interact again at the moment of reionisation. We will take these two effects into account later.

To better understand the shape of the temperature spectrum, we can apply the same approximation to the line-of-sight integral (5.55–5.57), i.e. replace $g(\eta)$ by $\delta_D(\eta - \eta_{\rm dec})$, and $e^{-\tau(\eta)}$ by $H(\eta - \eta_{\rm dec})$. Starting from (5.57), we then obtain:

$$\begin{aligned}\Theta_l(\eta_0, k) &= (\Theta_0(\eta_{\rm dec}, k) + \psi(\eta_{\rm dec}, k))\; j_l(k(\eta_0 - \eta_{\rm dec})) && \text{(SW)}\\ &\quad + k^{-1}\theta_{\rm b}(\eta_{\rm dec}, k)\; j_l'(k(\eta_0 - \eta_{\rm dec})) && \text{(Doppler)}\\ &\quad + \int_{\eta_{\rm dec}}^{\eta_0} d\eta\; (\phi'(\eta, k) + \psi'(\eta, k))\, j_l(k(\eta_0 - \eta))\;. && \text{(ISW)}\end{aligned} \tag{5.60}$$

The interpretation of this result is very close to the discussion in section (5.2.2). As expected, the Sachs–Wolfe term contains the photon density ($\delta_\gamma = 4\Theta_0$) corrected by the value of the gravitational potential ψ at the time of decoupling: the sum $[\Theta_0 + \psi]$ is often called the effective temperature. The Doppler term contains the bulk velocity of the strongly coupled baryon-electron-photon fluid, described by $\theta_{\rm b}$ up to a factor k[8] (see equation (3.17)). Since the Bessel function

along a smaller sphere, such as sphere A in the figure. It is immediately apparent that this mode cannot contribute at the desired angle (it can only contribute at larger angles). If the mode propagates earlier, the observer perceives a cross-sectional view along a larger sphere, such as spheres C, D, E, F in the figure. For a certain value of η corresponding to sphere D, the angle subtends a min–max–min–max contribution which also contributes, but with an opposite sign. For sphere F, the angle θ subtends a max–min–max–min–max–min contribution which is of the same sign. Sphere C, halfway between B and D, gives a zero contribution (the compared points have the same value), as does sphere E between D and F.

[8] An additional complexity arises from the fact that in real space the Doppler effect is deduced from a gradient term projected along the line of sight. In the line-of-sight formalism, this translates into the presence of a derivative of the Bessel function. Since $j_l'(x)$ has the same oscillations as $j_l(x)$ but with a phase shift, if $X(\eta, k) j_l(k(\eta_0 - \eta))$ represents the projection into multipole space of the Fourier mode $X(\eta, k)$ seen on a sphere, $kX(\eta, k) j_l'(k(\eta_0 - \eta))$ represents the projection of its gradient calculated at any point of the sphere in the direction orthogonal to it.

peaks around $x \simeq l$, the Fourier modes that contribute the most to the SW and Doppler terms are of the order of $k = l/(\eta_0 - \eta_{\rm dec})$. Our discussion of angular projection effects in the previous section allows us to understand this result: the main contribution comes from the main harmonic, i.e. from Fourier modes whose adjacent minima and maxima are seen by the observer at an angle $\theta = \pi/l$ when the mode propagates tangentially to the last scattering surface. Finally, as we have already seen in real space, the integrated Sachs–Wolfe term depends on the time variation of $(\phi+\psi)$ along the line of sight. The integral is not trivial since it involves a partial derivative and not a total derivative.

5.4 CMB physics

5.4.1 Evolution of perturbations

According to the previous section, to understand the form of the photon transfer function $\Theta_l(\eta_0, k)$, and then of the temperature spectrum C_l, we need to know the evolution of the transfer functions Θ_0, $\theta_{\rm b}$, ϕ and ψ. We will simply give a graphical summary of this evolution in figure 5.5, and then comment on the most important aspects and their physical origin. The upper plot of figure 5.5 shows the evolution of the transfer function $\psi(\eta, k)$ for each k, calculated almost exactly by an Einstein–Boltzmann solver. The evolution of ϕ is very similar due to the negligible role of the anisotropic neutrino pressure (see section 3.5). The lower plot shows the evolution of the effective temperature transfer function, $[\Theta_0 + \psi]$. That of $\theta_{\rm b}$ is not shown in the figure, but we will describe it qualitatively in a later section.

Evolution of metric perturbations

The evolution of $\psi(\eta, k)$ is quite simple. We have seen that the perturbations of the metric remain constant in the super-Hubble regime for adiabatic initial conditions. When a wavelength becomes smaller than the Hubble radius, the total energy density contrast $\delta_{\rm tot}$ experiences either:

- oscillations of constant amplitude due to pressure forces during radiation domination,
- or some enhancement $|\delta_{\rm tot}| \nearrow$ due gravity forces during matter or Λ domination.

The metric fluctuations react to this behaviour on sub-Hubble scales according to the Poisson equation (3.23) – which is the limit of the Einstein $\binom{0}{0}$ equation on sub-Hubble scales:

$$-\frac{k^2}{a^2}\phi = 4\pi G\,\delta\rho_{\rm tot} \quad \Rightarrow \quad \phi = -\frac{4\pi G}{k^2}\,(a^2\bar{\rho}_{\rm tot})\;\delta_{\rm tot}\;. \tag{5.61}$$

This shows that during radiation domination ($a^2\bar{\rho}_{\rm tot} \propto a^{-2}$) and matter domination ($a^2\bar{\rho}_{\rm tot} \propto a^{-1}$), $|\phi|$ tends to get damped even if $|\delta_{\rm tot}|$ is constant: the expansion of the universe tends to smooth out metric fluctuations on sub-Hubble scale. As a result:

1. during radiation domination, as a combined effect of pressure forces and of the expansion, ϕ experiences damped oscillations on sub-Hubble scales.

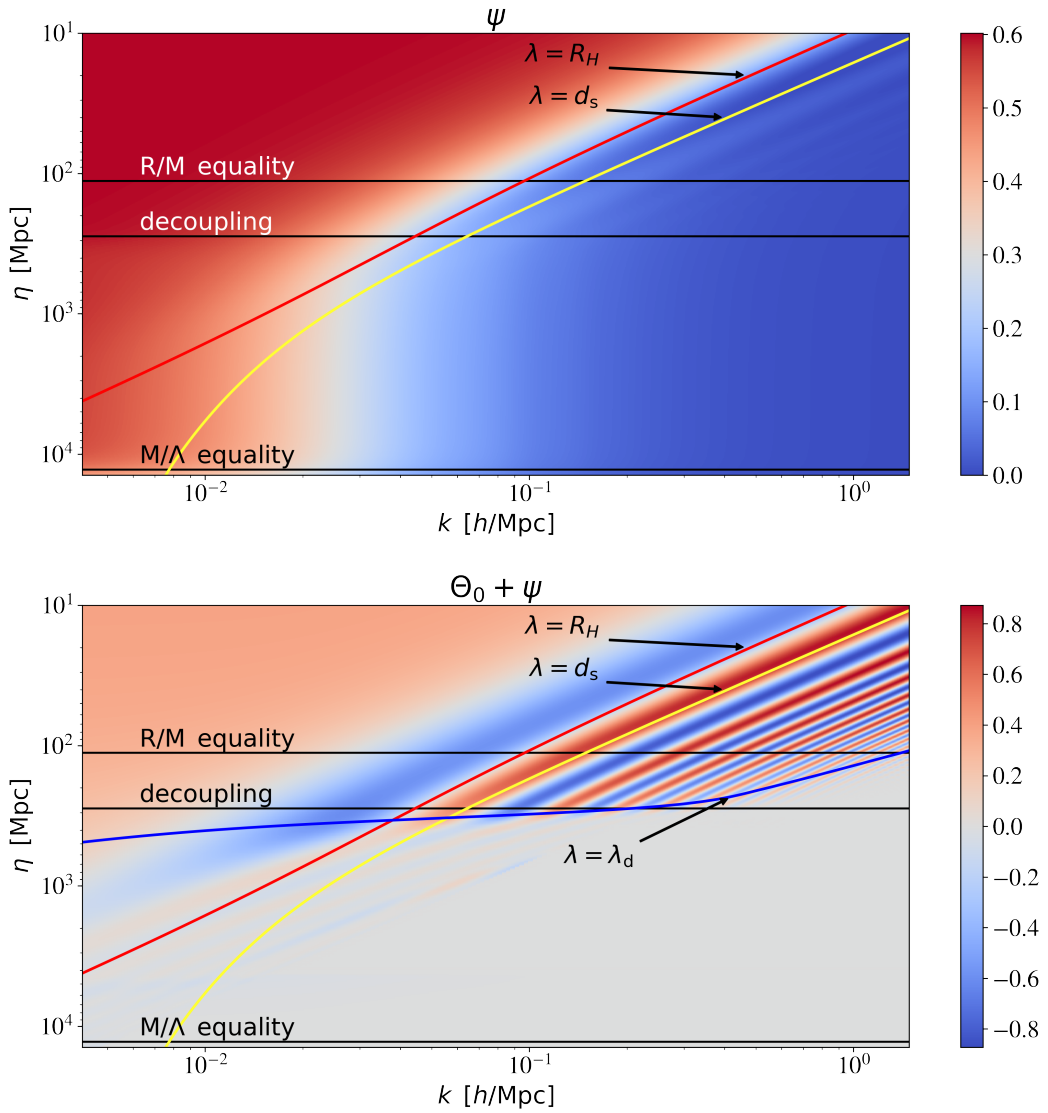

Figure 5.5: Transfer functions of the metric fluctuation $\psi(\eta, k)$ (top) and of the effective temperature $[\Theta_0 + \psi](\eta, k)$ (bottom). The vertical axis represents conformal time, flowing from the primordial universe (top) to the present day (bottom). The horizontal lines indicate characteristic times: radiation/matter equality, photon decoupling and matter/Λ equality. The oblique lines represent the location of the points where the wavelength $\lambda = a\,2\pi/k$ crosses the Hubble radius R_H, the sound horizon d_s or the damping scale d_d. Obtained with CLASS (`class-code.net`).

2. during matter domination, gravity forces lead to $|\delta_\mathrm{tot}| \propto a$, but this factor is exactly canceled by another factor a^{-1} coming from the expansion, and ϕ remains constant in time. Another way to see this is through the $\binom{i}{i}$ Einstein equation: as long as total pressure perturbations in the universe are negligible ($\delta p_\mathrm{tot} \ll \delta\rho_\mathrm{tot}$, which is true during MD and ΛD)

the Einstein equation for $\binom{i}{i}$ gives a closed equation of evolution for ϕ,

$$\phi'' + 3\frac{a'}{a}\phi' + \left(2\frac{a''}{a} - \frac{a'^2}{a^2}\right)\phi = 0 \ . \tag{5.62}$$

During MD, $a(\eta) \propto \eta^2$, the effective mass term in the previous equation vanishes, and ϕ=constant is a solution. This happens because the effect of gravitational collapse and of the expansion exactly cancel each other.

3. Instead, during Λ domination (ΛD), an explicit calculation shows that $(2\frac{a''}{a} - \frac{a'^2}{a^2})$ is positive, which implies and that ϕ decays: expansion wins against gravitational collapse, because the expansion is boosted by the contribution of Λ.

All these behaviours are easy to identify on figure 5.5 (upper plot).

Sound waves

The evolution of $[\Theta_0 + \psi](\eta, k)$ is more interesting because of the phenomenon of acoustic oscillations. Until decoupling, photons, electrons and baryons form a tightly coupled fluid, with

$$\delta_{\rm b} = \frac{3}{4}\delta_\gamma \ , \quad \theta_\gamma = \theta_b \ , \quad \Theta_{l\geq 2} = 0 \ , \tag{5.63}$$

as demonstrated earlier.

The pressure of the photon-electron-baryon fluid is dominated by the contribution of ultra-relativistic photons: at each point, $p_\gamma(t, \vec{x}) = \frac{1}{3}\rho_\gamma(t, \vec{x})$. Therefore, any over-density or over-pressure initially present in the fluid propagates as a sound wave, with a speed of sound whose square is given by:

$$c_{\rm s}^2 = \frac{\delta p_\gamma + \delta p_{\rm b}}{\delta\rho_\gamma + \delta\rho_{\rm b}} \simeq \frac{\frac{1}{3}\delta\rho_\gamma}{\delta\rho_\gamma + \delta\rho_{\rm b}} = \frac{\frac{1}{3}}{1 + \frac{\delta\rho_{\rm b}}{\delta\rho_\gamma}} = \frac{1}{3(1 + \frac{\bar{\rho}_{\rm b}\delta_{\rm b}}{\bar{\rho}_\gamma\delta_\gamma})} \ . \tag{5.64}$$

We have seen that $\delta_{\rm b}/\delta_\gamma$ is equal to 3/4. Finally, after having defined the baryon-photon ratio $R \equiv (3\bar{\rho}_{\rm b})/(4\bar{\rho}_\gamma)$, we obtain

$$c_{\rm s}^2 = \frac{1}{3(1+R)} \ . \tag{5.65}$$

As $\bar{\rho}_{\rm b} \propto a^{-3}$ and $\bar{\rho}_\gamma \propto a^{-4}$, the ratio R grows proportionally to the scale factor. During radiation domination, R is negligible compared to one. Indeed, using the index m for the non-relativistic matter components and r for the ultra-relativistic radiation components, $\bar{\rho}_{\rm m} \ll \bar{\rho}_{\rm r}$ implies $\bar{\rho}_{\rm b} \ll \bar{\rho}_\gamma$. The square of the sound speed is then given by 1/3, because the fluid is entirely dominated by photons. During the domination of matter, R becomes important. The sound speed tends towards zero as the fluid becomes progressively dominated by the baryons, which constitute an ultra–relativistic and pressureless species. The propagation of sound waves then gradually ceases. Finally, the sound horizon is defined as the physical distance travelled by an acoustic wave between the primordial universe and an instant η:

$$d_{\rm s}(\eta) = a(\eta)\int_0^\eta c_s \, d\tilde{\eta} \ . \tag{5.66}$$

Acoustic oscillations

In general, the propagation of a real-valued plane wave with frequency ω and wave vector $\vec{k}$ is given by a function of the form $\Psi(t,\vec{x}) = \Re[Be^{-i(\omega t - \vec{k}\cdot\vec{x})}]$. Therefore, when a field $A(t,\vec{x})$ consists of a superposition of plane waves, each Fourier mode $A(t,\vec{k})$ is of the form $A(t,\vec{k}) = \Re[Be^{i\omega t}] = |B|\cos(\omega t + \varphi)$: plane wave propagation in real space corresponds to oscillating solutions in Fourier space. In our case, the propagation of sound waves in the photon–electron–baryon fluid must correspond to oscillating solutions for the Fourier transform of the density fluctuations of photons and baryons, i.e. for $\delta_\mathrm{b} = \frac{3}{4}\delta_\gamma$: this is what we call the acoustic oscillations in the primordial universe.

To show this, one can combine the Euler equations of photons [5.41, 5.42] and of baryons. You proved in the exercises that, after a few calculations, one obtain an inhomogeneous linear differential equation of the second order concerning the effective temperature $[\Theta_0 + \psi]$,

$$\Theta_0'' + \frac{R}{1+R}\frac{a'}{a}\Theta_0' + k^2 c_\mathrm{s}^2 \Theta_0 = -\frac{k^2}{3}\psi + \frac{R}{1+R}\frac{a'}{a}\phi' + \phi'' \ . \tag{5.67}$$

This is the equation of a driven oscillator. In the left-hand side:

- The third term contains – up to a factor $\bar{\rho}_\gamma$ – the product $k^2 c_\mathrm{s}^2 \delta\rho_\gamma = k^2 \delta p_\gamma$, which comes from the Laplacian of pressure in real space: it represents the pressure force, which resists to compression and thus allows the propagation of sound waves.

- The second term is related to a famous gravitational effect in an expanding universe. In general, an inhomogeneous non-relativistic fluid is subject to the gravitational collapse mechanism, but the expansion slows down this collapse. In the equations, this is reflected in a Hubble friction term propertional to $\frac{a'}{a} = aH$. In the present case, this term is only important when the contribution of non-relativistic baryons to the density of the photon-electron-baryon fluid is large: it is therefore multiplied by $R/(1+R)$.

The right-hand side constitutes the source term of the driven oscillator. It shows how fluctuations in the metric can generate density fluctuations in the photon-electron-baryon fluid, or amplify existing fluctuations.

- The term $-k^2\psi$ comes from the Laplacian of the gravitational potential in real space: it represents the gravitational force, which may cause gravitational collapse.

- The second term of the right-hand side has the same physical origin as the second term of the left-hand side: it represents a local modulation of the Hubble friction effect.

- The term ϕ'' takes into account the dilation effect (see section 5.2.1). This effect can attenuate (respectively amplify) the overdensities at the places where the local expansion accelerates (resp. slows down), by shifting the photon wavelengths towards the red (resp. blue).

If we consider in first approximation that the metric fluctuation $\phi = \psi$ is constant in time, and that the functions R and c_s^2 vary over a timescale that is much longer than one period of oscillation, we can find an approximate solution

using the WKB approximation scheme for second order differential equations. You will do this in the exercises and find:

$$\Theta_0 + \psi \simeq \frac{1}{3}(1+R)^{-1/4} \cos\left(k \int_0^\eta c_s(\tilde{\eta})\, d\tilde{\eta}\right) - R\psi \ . \tag{5.68}$$

As expected, this is an oscillating solution, that corresponds in real space to the propagation of sound waves. The phase of the oscillation depends on the ratio between the wavelength and the sound horizon, since the argument of the cosine is equal to $2\pi\, d_{\rm s}(\eta)/\lambda(\eta)$ (where we used the expression of the sound horizon (5.66) and $\lambda = a\frac{2\pi}{k}$). Since $d_{\rm s}(\eta)$ increases faster than $\lambda(\eta)$, the phase increases with time. In other words, the modes start to oscillate as they approach the sound horizon. When they reach $\lambda(\eta) = d_{\rm s}(\eta)$, that is, $k\int_0^\eta c_s(\tilde{\eta})\, d\tilde{\eta} = 2\pi$, they complete their first oscillation. Then, the further they are within the sound horizon, the more times they have oscillated. This behaviour is clearly visible in the bottom graph of figure 5.5: it corresponds to the positive and negative bands parallel to the sound horizon crossing line. When baryons dominate, the sound waves are gradually attenuated by the factor $(1+R)^{-1/4}$, while they gradually stop to propagate due to $k\int_0^\eta c_s(\tilde{\eta})\, d\tilde{\eta} \longrightarrow$ constant.

Note that the oscillations do not take place around a zero mean value, but around $-R\psi$. Indeed, when the density of baryons becomes important ($R \geq 1$), pressure forces become weaker relative to gravity forces. The areas of over-density therefore become denser and those of under-density emptier. In Fourier space, this results in a shift in the mean value of the $\Theta_0+\psi$ oscillations. Thus, for an over-density ($\Theta_0 > 0$) associated with a gravitational potential well ($\psi < 0$), the mean value $-R\psi > 0$ increases, making the over-density denser (and vice versa).

To understand qualitatively the additional impact of the time variation of metric fluctuations on the right-hand side of the forced oscillator equation (5.67), we have to remember that the fluctuations of the metric remain constant in the super–Hubble regime, and then decrease in the sub–Hubble regime during radiation domination, until they tend towards a constant for $k\eta \ll 1$. Consequently, the source terms containing time derivatives of ϕ and ψ are non-zero only around the crossing of the Hubble radius during radiation domination, when $\lambda \simeq R_H$. They play the role of an amplifier for the forced oscillator. This effect is called the gravitational boost of the acoustic oscillations. Thus, within the Hubble radius, the approximate solution (5.68) underestimates the amplitude of the oscillations.

Diffusion damping

The equation (5.67) is based on the strong coupling limit. In reality, photon decoupling is not instantaneous: the mean free path of the photons increases progressively. The photon–electron–baryon fluid thus passes through a regime of weaker and weaker coupling, until the last photon scatterings occur around the time $\eta \simeq \eta_{\rm dec}$.

In section 5.1.4, we found an analytical approximation for the diffusion length $\lambda_d(\eta) = a(\eta) r_d(\eta) = a(\eta)\frac{2\pi}{k_d(\eta)}$ (see equation (5.8)), which gives the average distance travelled by a photon at a time η. At this time, in each sphere of radius $\lambda_{\rm d}(\eta)$, the direction of propagation of the photons is randomly redistributed, which erases all perturbations of comoving wavelength smaller than

$$\frac{2\pi}{k} < \frac{2\pi}{k_{\rm d}(\eta)} = r_{\rm d}(\eta) \ . \tag{5.69}$$

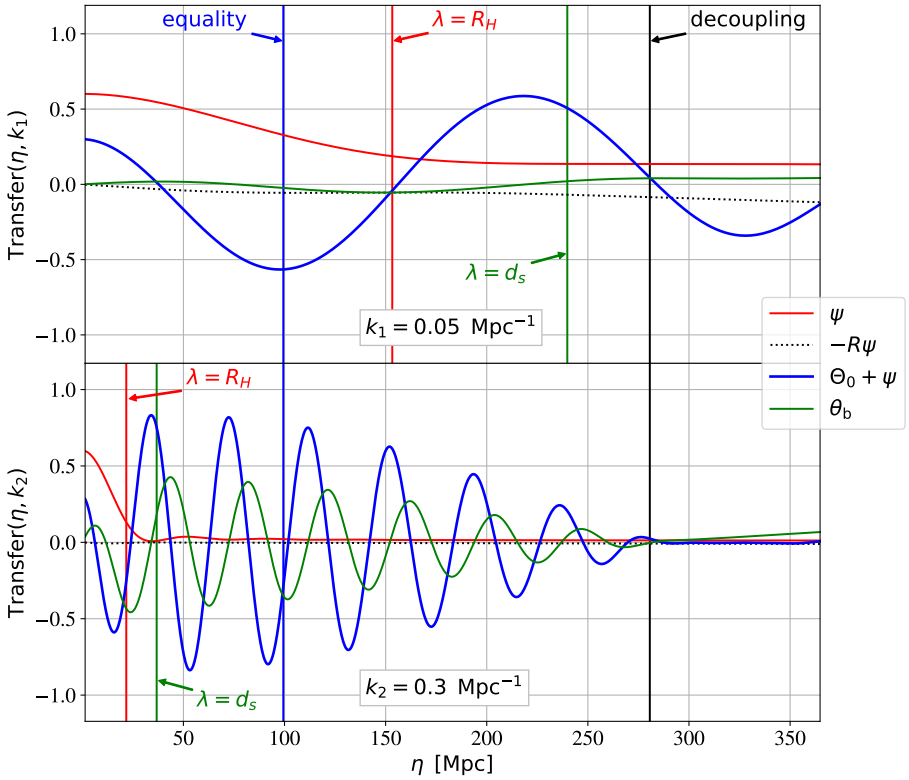

Figure 5.6: Evolution of the transfer functions ψ and $[\Theta_0 + \psi]$ as a function of the conformal time η for the comoving modes $k_1 = 0.05\,\mathrm{Mpc}^{-1}$ (top) and $k_2 = 0.3\,\mathrm{Mpc}^{-1}$ (bottom). The vertical lines show four characteristic times: radiation/matter equality, decoupling, crossing of the Hubble radius ($\lambda = R_H$), crossing of the sound horizon ($\lambda = d_\mathrm{s}$). Obtained with CLASS (`class-code.net`).

This effect is only accurately described by the exact equations of motion, that is, the full Boltzmann hierarchy with the precise time evolution of $\Gamma_\gamma(\eta) = -\tau'(\eta)$. However, it can be shown that, as a first approximation, it can be taken into account by preserving the predictions of the strongly coupled acoustic oscillation equation, and by multiplying the solution obtained for Θ_0 by an exponential cut-off function of the form $\exp[-(k/k_\mathrm{d}(\eta))^2]$.

This damping effect of small-scale fluctuations is called Silk damping (after the physicist Joseph Silk). It is clearly visible on the bottom graph of figure 5.5: the perturbations are negligible below the line where $\lambda = \lambda_\mathrm{d}$.

Summary of the evolution of transfer functions

Figure 5.6 shows the evolution of ψ and $[\Theta_0 + \psi]$ for two comoving Fourier modes, between an initial time close to the end of inflation and a final time chosen shortly after photon decoupling. Since the fluctuations are normalized to $\mathcal{R} = 1$, these graphs represent transfer functions: they are vertical slices (at constant k) of figure 5.5. The velocity of the θ_b baryons is also plotted because it will play a role in the next section.

For the longest wavelength (top graph), the crossing of the Hubble radius

occurs shortly after the time of equality between radiation and matter, and the potential ψ falls only slightly before stabilising around a constant value. The effective temperature $[\Theta_0 + \psi]$ begins to oscillate as the mode approaches the Hubble radius, and ends its first period of oscillation when $\lambda \simeq d_{\rm s}$. The oscillations are centred around the equilibrium value $-R\,\psi$, and not around zero. The amplitude of the oscillations increases somewhat during the first half-period (corresponding to the end of radiation domination) due to the gravitational boost effect, and then decreases due to the role of baryons and the factor $(1 + R)^{-1/4}$. This mode has a too long a wavelength to be affected by the diffusion damping (Silk damping) before decoupling.

For the smallest wavelength (bottom graph), Hubble radius crossing takes place well before equality between radiation and matter. Around the time when $\lambda = R_H$, the potential ψ tends rapidly towards zero. The effective temperature $[\Theta_0 + \psi]$ begins to oscillate as it approaches the Hubble radius and ends its first period of oscillation around the time when $\lambda = d_{\rm s}$. In this case, R is only important when ψ is almost zero. Consequently the product $-R\,\psi$ is always very small, and the oscillations are approximately centred around zero. The amplitude of the oscillations increases very sharply during the first oscillation due to the gravitational boost effect. Then it slowly decreases due to the predominance of baryons. Finally, the fluctuations are almost completely erased before decoupling by the effect of diffusion damping.

The curves in Figure 5.6 are simply two vertical cuts (at constant k) of the surfaces in Figure 5.5. In the next section, we will use these results to establish the shape of the temperature spectrum C_l: we will then be more interested in horizontal cuts of these surfaces at constant time η, and in particular, at $\eta = \eta_{\rm dec}$.

5.4.2 Contributions to the temperature spectrum

We have seen that the photon transfer function $\Theta_l(\eta_0, k)$ is composed of the three Sachs-Wolfe (SW), Doppler and Integrated Sachs-Wolfe (ISW) terms (see equations (5.55-5.56) or (5.60)). However, the calculation of the spectrum C_l is based on an integral over k whose argument is $(\Theta_l(\eta_0, k))^2$ multiplied by the primordial spectrum. Therefore, the spectrum is given by the sum of six contributions: the SW spectrum, the Doppler spectrum, the ISW spectrum, and three cross terms: SW–Doppler, SW–ISW, Doppler–ISW. We will describe the first three contributions using figure 5.7, which shows the most important transfer functions, their square, the total spectrum C_l, and the contributions $C_l^{(\rm SW)}$, $C_l^{(\rm Doppler)}$, $C_l^{(\rm ISW)}$.

Sachs-Wolfe contribution (SW): large-scale plateau

The SW contribution to the temperature spectrum comes from the transfer function of the effective temperature $[\Theta_0+\psi]$ evaluated at the time of decoupling (see equation (5.60)). In this subsection, we will focus on the contribution of the modes that are still in the super-Hubble regime at $\eta = \eta_{\rm dec}$.[9]

Since photon decoupling occurs during matter domination, one can use equation (4.60) to express the transfer functions that give the Sachs–Wolfe term as:

[9] We saw in 5.3.3 that a multipole l receives a contribution from the Fourier modes at $\eta = \eta_{\rm dec}$ mainly for wave numbers k such that $l \simeq k(\eta_0 - \eta_{\rm dec})$. Now, if these modes satisfy $k\eta_{\rm dec} < 1$, they were in the super–Hubble regime at the time of decoupling. Therefore the multipoles obeying $l < (\eta_0 - \eta_{\rm dec})/\eta_{\rm dec} \sim 50$, seen at an angle $\theta = \pi/l$ of the order of a few degrees or more, are determined by the perturbations in the super–Hubble regime.

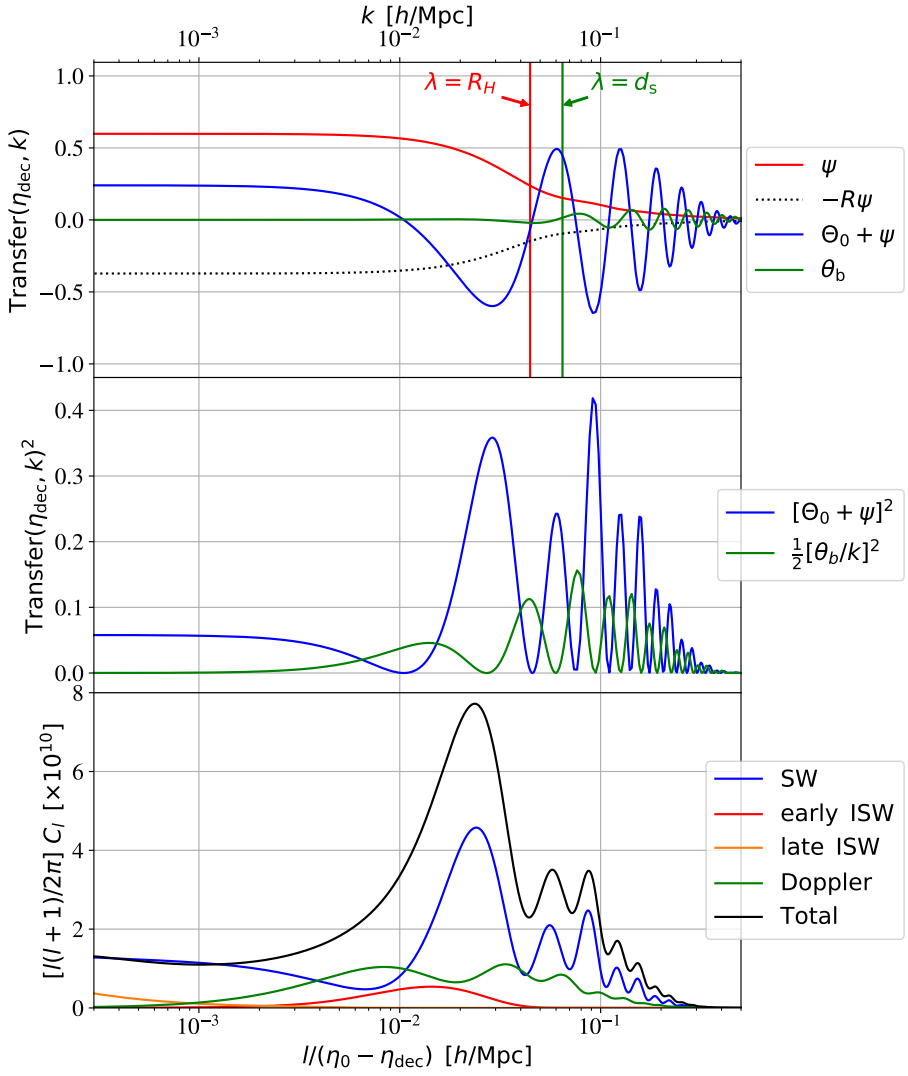

Figure 5.7: (Top:) Transfer functions that are important for the calculation of the temperature spectrum C_l, evaluated at decoupling and plotted as a function of k. The vertical lines show the k modes whose wavelength is equal to the Hubble radius or the sound horizon. (Middle:) square of the transfer function $[\Theta_0 + \psi]$, which gives the Sachs–Wolfe (SW) contribution, and $[\theta_\mathrm{b}/k]$ (divided arbitrarily by two for better readability), which gives approximately the Doppler contribution. (Bottom:) Temperature spectrum C_l plotted as a function of $l/(\eta_0 - \eta_\mathrm{dec})$, and contribution of the SW, early and late ISW, and Doppler spectra. Obtained with CLASS (`class-code.net`).

$$[\Theta_0+\psi]_{(\eta_{\rm dec},k)} = \frac{[\frac{1}{4}\delta_\gamma+\psi](\eta_{\rm dec},\vec{k})}{\mathcal{R}(\eta_{\rm ini},\vec{k})} = \frac{\left(-\frac{2}{3}+1\right)\psi(\eta_{\rm dec},\vec{k})}{\frac{5}{3}\psi(\eta_{\rm dec},\vec{k})} = \frac{1}{5}\ . \qquad (5.70)$$

This calculation has a striking physical interpretation. The contributions of the two terms Θ_0 and ψ are of opposite sign, because when photons leave an overdensity on the last scattering surface (with a high temperature), they have to climb a gravitational potential well and lose energy (which lowers their temperature), or vice-versa: this is the Sachs-Wolfe effect. But in the equation (5.70), we see that the second effect prevails, since in absolute value $1 > \frac{2}{3}$. So the hot spots on the last scattering surface are perceived by the observer as cold spots: on large angular scales, the CMB temperature map shows a negative picture of the last scattering surface. (On small scales, it is the opposite, since in that case $\psi(k,\eta_{\rm dec}) \simeq 0$. Then, the intrinsic temperature fluctuation does not receive a Sachs-Wolfe correction, and the small hot spots on the map also correspond to small hot spots on the last scattering surface.)

We can now calculate the C_l spectrum given by this contribution using equation (5.51). Parameterising the primordial spectrum by a power law as in equations (4.73, 4.86), we obtain:

$$C_l = 4\pi\int\frac{dk}{k}\frac{A_{\rm s}}{25}\left(\frac{k}{k_*}\right)^{n_{\rm s}-1} j_l^2(k(\eta_0-\eta_{\rm dec})) = 4\pi\frac{A_{\rm s}}{25}\int\frac{dx}{x}\left(\frac{x}{x_*}\right)^{n_{\rm s}-1} j_l^2(x), \qquad (5.71)$$

where we have made a change of variable from k to $x \equiv k(\eta_0-\eta_{\rm dec})$, and defined $x_* \equiv k_*(\eta_0-\eta_{\rm dec})$. When the spectrum is exactly scale invariant, $n_{\rm s}=1$, we can use the exact formula $\int_0^\infty dx x^{-1} j_l(x)^2 = [2l(l+1)]^{-1}$ and get a simple solution,

$$C_l = \frac{2\pi A_{\rm s}}{25 l(l+1)}\ . \qquad (5.72)$$

Consequently, the quantity $l(l+1)C_l$ is constant. The case of scale-invariant primordial fluctuations corresponds to $\mathcal{P}_\mathcal{R}(k)$ being constant in Fourier space, and $l(l+1)C_l$ being constant in multipole space. For this reason, when showing the CMB spectrum on a figure, it is customary to plot $l(l+1)C_l$.

When $n_{\rm s} \neq 1$, there is also an analytical solution involving the Euler function $\Gamma(x)$. We will not give it here: it is enough to know that $l(l+1)C_l$ then depends slightly on l. With $n_{\rm s} > 1$ (spectrum $\mathcal{P}_\mathcal{R}(k)$ larger for large k or small wavelengths, called a "red spectrum"), the $l(l+1)C_l$ is larger for large l or small angles. With $n_{\rm s} < 1$ (a "blue spectrum"), it is the opposite. Therefore, the SW contribution to $l(l+1)C_l$ simply reflects the shape of the primordial dimensionless spectrum $\mathcal{P}_\mathcal{R}(k)$.

We will see that at these scales, the other contributions (Doppler and ISW) are weaker (see also the graph at the bottom of the figure 5.7). This quasi-horizontal branch of the CMB spectrum is therefore called the Sachs-Wolfe plateau. A low angular resolution experiment like COBE–DMR (1992-94) could only see this region of the spectrum, but this was sufficient to obtain an approximate measurement of $A_{\rm s}$ and $n_{\rm s}$, and to obtain a first confirmation of one of the predictions of inflation: $n_{\rm s} \simeq 1$.

Sachs–Wolfe (SW) contribution: acoustic peaks

The previous calculation shows that if the fluctuations remained constant and nearly scale-invariant at all epochs and wavelengths (and not only in the super-Hubble regime), the $l(l+1)C_l$ spectrum would be nearly flat. But physical

phenomena occurring at sub-Hubble scales modulate the fluctuations. These modulations are reflected in the CMB spectrum.

In figure 5.7, the upper plot shows the transfer function $[\Theta_0 + \psi]_{(\eta_{\rm dec},k)}$, calculated by an Einstein–Boltzmann code. This function is simply a horizontal cut of the lower plot of figure 5.5 at time $\eta_{\rm dec}$. Its behaviour is in agreement with the discussion in section 5.4.1. The modes that are still in the super–Hubble regime at the time of decoupling simply reflect the initial adiabatic conditions (see equation (5.70)). Smaller wavelengths keep the imprint of acoustic oscillations, for which we have seen an approximate expression in equation (5.68). At the time of decoupling, some modes have oscillated for half a period (first negative peak of the transfer function), or one period (second positive peak), or one and a half period (third negative peak), etc. The oscillations are centred on a mean value $-R\,\psi(\eta_{\rm dec}, k)$ which tends to zero in the sub–Hubble limit. The amplitude of the oscillations relative to this mean value increases between the first and second peaks, because the gravitational boost effect mentioned at the end of the section 5.4.1 is weaker for modes that cross the Hubble radius during matter domination, which is the case for the modes contributing to the first peak (such as the k_1 mode in figure 5.6). The amplitude is then roughly constant for the next three peaks, and then decreases due to the exponential cut-off $\exp(-(k/k_{\rm d})^2)$ caused by diffusion damping 5.4.1.

The spectrum C_l involves the square of the function $[\Theta_0 + \psi]_{(\eta_{\rm dec},k)}$ (see (5.51)). This quantity appears on the middle plot of figure 5.7. We can observe the consequence of the shift in the mean value of the oscillations by $-R\,\psi(\eta_{\rm dec}, k)$. As long as the transfer function $-R\,\psi(\eta_{\rm dec}, k)$ is non-negligible, i.e. for the first five or six peaks, this shift makes the square of the negative peaks higher than the square of the positive peaks. We therefore have an alternation of large peaks (odd peaks: first, third, etc.) and small peaks (even peaks: second, fourth). Beyond, the effect of this shift is negligible, and the exponential cut-off in $\exp(-2(k/k_{\rm d})^2)$ takes over: the amplitude of the peaks decreases monotonically.

Finally, according to the discussion in section 5.3.3, the $C_l^{\rm SW}$ contribution of the SW term to the temperature spectrum should look like $[\Theta_0 + \psi]^2_{(\eta_{\rm dec},k)}$, applying the $k \longleftrightarrow l = k(\eta_0 - \eta_{\rm dec})$ mapping. To illustrate this, the graph at the bottom of figure 5.7 shows $C_l^{\rm SW}$ not as a function of l but of $l/(\eta_0 - \eta_{\rm dec})$. In this way, we can effectively identify the same behaviour in the functions $[\Theta_0 + \psi]^2$ and $C_l^{\rm SW}$, with the same structures at the same positions: a Sachs–Wolfe plateau, some peaks called the acoustic peaks, and an exponential cut-off.[10]

Equation (5.68) gives us an analytical approximation for the position of the peaks of the C_l^{SW} spectrum in mutipole space: as the extrema of $[\Theta_0 + \psi]_{(\eta_{\rm dec},k)} \sim \cos(k\int_0^{\eta_{\rm dec}} c_s d\tilde{\eta})$ are given by $k\int_0^{\eta_{\rm dec}} c_s d\tilde{\eta} = N\pi$, with $N = 1, 2, 3, ...$, the peaks correspond to

$$l_N \simeq k(\eta_0 - \eta_{\rm dec}) = N\pi \frac{\eta_0 - \eta_{\rm dec}}{\int_0^{\eta_{\rm dec}} c_s d\tilde{\eta}} = N\pi \frac{d_{\rm a}(\eta_{\rm dec})}{d_{\rm s}(\eta_{\rm dec})}\ , \tag{5.73}$$

where $d_{\rm a}(\eta)$ is the angular diameter (5.58) and $d_s(\eta)$ the sound horizon (5.66). By substituting $d_s(\eta_{\rm dec})$ with $\lambda_{\rm d}(\eta_{\rm dec})$, we could get a similar estimate of the

[10] The mapping $k \longleftrightarrow l = k(\eta_0 - \eta_{\rm dec})$ would only be perfect if we only considered the main harmonic and used the instantaneous decoupling approximation; but the exact result takes into account the secondary harmonics and the progressivity of the decoupling, since the integral over η in equation (5.56) involves all the peaks of the Bessel functions as well as the width of the visibility function $g(\eta)$. This explains why the $C_l^{\rm SW}$ curve has a smoother appearance than the $[\Theta_0 + \psi]^2$ curve: the zeros of the transfer function become minima in the spectrum, and the relative amplitude of the peaks is somewhat different.

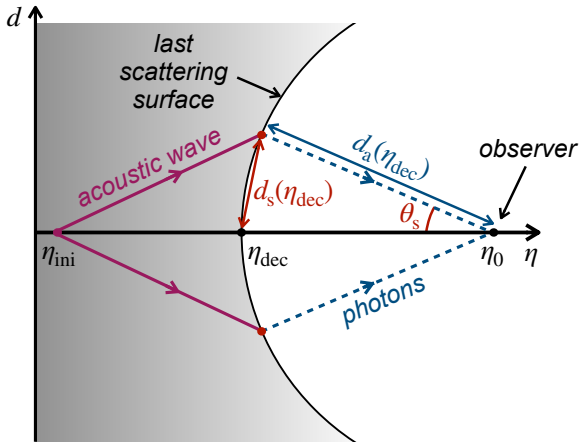

Figure 5.8: Two points on the last scattering surface separated by a distance $2d_s(\eta_{\rm dec})$ are statistically correlated due to an identical contribution of the sound waves emitted at the midpoint in the primordial universe in two opposite directions.

multipole l_d at which we observe an exponential cut-off in the enveloppe of the acoustic peaks, well approximated by $\exp(-2(l/l_{\rm d})^2)$.

We have described the acoustic peaks in Fourier space (as a function of k) and then in multipole space (as a function of l). To get a better intuition, it is useful to think also of their physical origin in real space, sketched in figure 5.8. Imagine two points on the last scattering surface separated by a physical distance $2d_s(\eta_{\rm dec})$. These two points must have received the contribution of sound waves emitted in the primordial universe at the midpoint. Therefore, they both carry information about the initial density fluctuation at the same point: they are statistically correlated. If we define $\theta_{\rm s} \equiv d_{\rm s}(\eta_{\rm dec})/d_{\rm a}(\eta_{\rm dec})$, such a pair of points must be seen today under a separation angle $2\theta_{\rm s}$. If we measured the angular correlation function $\langle \frac{\delta T}{T}(\hat{n}) \frac{\delta T}{T}(\hat{n}') \rangle$ on the last scattering surface, we should see a peak corresponding to this correlation when the two directions $\hat{n}$ and $\hat{n}'$ are separated by such an angle $2\theta_{\rm s}$, that is, for $\hat{n} \cdot \hat{n}' = \cos 2\theta_{\rm s}$. Thus, in multipole space, the spectrum C_l must have a peak at the corresponding mutipole, $l_1 = \pi/\theta_{\rm s}$, which corresponds to the main harmonic of the correlation[11]. There should also be other peaks at $l_N = N\pi/\theta_{\rm s}$, which correspond to the secondary harmonics[12].

[11] When we decompose the map into mutipoles, each a_{lm} corresponds to configurations where an adjacent maximum and minimum are separated by an angle π/l. Therefore, two adjacent maxima or minima are separated by an angle $2\pi/l$. For example, in the case of the quadruple, $l = 2$, two maxima are separated by an angle π. If any pair of points separated by $2\theta_{\rm s}$ is statistically correlated, these two points can typically correspond to two maxima or two minima. The correlation will therefore increase the absolute value of the coefficient $a_{l_1 m}$ with $2\pi/l_1 = 2\theta_{\rm s}$, i.e. $l_1 = \pi/\theta_{\rm s}$. This correlation should therefore produce a peak in the C_l spectrum for this mutipole.

[12] In the multipoles $l_N = N\, l_1$ with $N = 2, 3, ...$, we also find maxima or minima separated by $2\theta_{\rm s}$. Correlation must therefore also contribute to the value of these multipoles, and to other regularly spaced peaks in the C_l spectrum. These are the harmonics in multipole space of the correlation length $2d_{\rm s}$ in real space.

Doppler contribution

Returning to equation (5.60, the second contribution to the $\Theta_l(\eta_0, k)$ transfer function comes from the Doppler term (the physical origin of which is explained in section 5.2.3). The C_l^{Doppler} contribution can be roughly deduced from the transfer function $[\theta_\text{b}/k]$ evaluated at the time of decoupling and with a correspondence[13] $k \longleftrightarrow l = k(\eta_0 - \eta_\text{dec})$. According to the continuity equation (3.24 applied to photons, θ_b is given mainly by the derivative of δ_b, i.e. by $4\Theta_0'$ in the tight coupling approximation. Figure 5.6 shows indeed that $\theta_\text{b}(\eta, k)$ oscillates as a function of η, with a phase shift of $\sim \frac{\pi}{4}$ with respect to Θ_0. We observe again this behaviour for $\theta_\text{b}(\eta_\text{dec}, k)$ as a function of k in figure 5.7 (upper plot). The middle plot shows the square $[\theta_\text{b}/k]^2$ at $\eta = \eta_\text{dec}$, and the bottom plot proves that this square does indeed give the approximate shape[14] of C_l^{Doppler}.

In summary, the Doppler effect does not contribute to the Sachs-Wolfe plateau, but, at the angular scales where acoustic peaks are observed, it contributes to the spectrum, with a phase shift such that the maxima of C_l^{Doppler} almost coincide with the minima of C_l^{SW}. This shown in figure 5.7 (lower plot).

Integrated Sachs–Wolfe (ISW) contribution

We recall that the ISW contribution to the transfer function ($\Theta_l(\eta_0, k)$ is given by an integral over $(\psi' + \phi')$ between the instant of decoupling and the present day (see equation (5.60) and its physical interpretation at the end of the section 5.2.3). Neglecting the role of the anisotropic pressure of photons and electrons, we can write $(\psi' + \phi') \simeq 2\psi'$. However, the upper plot of figure 5.5 shows that ψ is constant with respect to time for most values of k and η after decoupling. Therefore, the integrated Sachs–Wolfe effect only receives contributions from very localised regions in the space $(\eta \geq \eta_\text{dec}, k)$. In figure 5.5, these regions are those where a vertical gradient for ψ is observed after decoupling. It can be seen immediately that there are only two regions that meet these criteria.

The first region is located immediately after decoupling, for modes crossing the Hubble radius. The physical interpretation is simple. We know that during radiation domination, the metric perturbations decrease in the sub-Hubble regime, while during matter domination they remain constant. Decoupling occurs at the beginning of matter domination, but the transition between the two eras is not instantaneous. Consequently, for metric fluctuations, the modes that cross the Hubble radius around decoupling are still subject to damping. These modes are precisely those that contribute to the first peak. The C_l^{ISW} spectrum is therefore non-zero in the vicinity of this peak, as confirmed by the lower plot of Figure 5.7. Since this effect comes from the evolution of the metric fluctuations at times close to decoupling, it is called the early ISW effect.

The second region corresponds to the domination of the cosmological constant. Using Einstein's equations, it can be shown that the metric fluctuations decrease at all scales during this period. However, for modes that have crossed the Hubble radius during radiation domination, the metric fluctuations are already erased. Therefore, the variation of $\psi \simeq \phi$ during the domination of the cosmological constant is only significant for long wavelengths that have crossed the Hubble radius during matter domination. This variation can be seen in the lower left end of the upper plot in Figure 5.5, where the red tone lightens slightly.

[13] Indeed, the derivative $j_l'(x)$ has a main peak slightly shifted with respect to $j_l(x)$, but always close to $l \simeq x$.

[14] The differences arise from the fact that decoupling is not instantaneous, and from convolution with the Bessel function $j_l'(x)$ instead of a Dirac function.

This region generates a late ISW effect that contributes to the smaller mutipoles, as confirmed by figure 5.7. The late ISW lifts the Sachs-Wolfe plateau.

In the lower plot of figure 5.7, it is obvious that the total spectrum is slightly greater than the sum of the individual SW, Doppler, early ISW and late ISW contributions. The difference comes from the SW×Doppler, SW×ISW and Doppler×ISW cross terms, which we will not describe here for the sake of brevity.

Reionisation effect

In order to understand the SW, Doppler and ISW contributions more simply, we have not used the full expression (5.56) of the photon transfer function $\Theta_l(\eta_0, k)$, but its approximate form (5.60), derived from the instantaneous decoupling approximation. Concretely, we have replaced the visibility function $g(\eta)$ by a Dirac function $\delta(\eta - \eta_{\rm dec})$. In doing so, we have not only neglected the progressivity of decoupling, corresponding to the width of the peak of $g(\eta)$ for $\eta \simeq \eta_{\rm dec}$, but also the effect of reionisation, corresponding to the presence of a second peak around $\eta \simeq \eta_{\rm reio}$, visible in figure 5.2 (bottom left). To estimate this second effect, we can adopt a new approximation,

$$g(\eta) = e^{-\tau_{\rm reio}} \delta(\eta - \eta_{\rm dec}) + (1 - e^{-\tau_{\rm reio}}) \delta(\eta - \eta_{\rm reio}) \ , \tag{5.74}$$

always normalized correctly to $\int_0^{\eta_0} g(\eta)\, d\eta = 1$, and respecting the fact that a fraction $e^{-\tau_{\rm reio}}$ of the photons had its last interaction at the time of recombination, and $(1 - e^{-\tau_{\rm reio}})$ at the time of reionisation. The photon transfer function then decomposes into two terms:

- the first peak generates the contribution of recombination, already described in the previous sections, but with the addition of a global normalisation factor $e^{-\tau_{\rm reio}} < 1$. The C_l spectrum is thus renormalized by $e^{-2\tau_{\rm reio}}$. Physically, this corresponds to the fact that the photons that re-interact at the time of reionisation are scattered in random directions: part of the coherence of the anisotropies coming from the last scattering surface is then lost.

- the second peak generates the contribution of reionisation, corresponding to correlations in the temperature distribution of photons that re-scattered at the time of reionisation. This contribution is very different from the one from recombination for two reasons. Firstly, it comes from another virtual last scattering surface, which is closer to us. Secondly, at the time of reionisation, the photon and baryon fluctuations are strongly attenuated on sub-Hubble scales, and only the super-Hubble scales (or the scales at the boundary between the two regimes) contribute to the photon transfer function. The contribution of reionisation is therefore only significant at very large angular scales, of the order of $l < 50$. Further study would show that in the C_l temperature spectrum, the reionisation contribution almost perfectly compensates the suppression of the recombination contribution by a factor $e^{-2\tau_{\rm reio}}$.

Therefore, in the end, the effect of reionisation on the C_l spectrum is equivalent to a multiplication by a progressive step-like function, equal to 1 for small l (approximately $l \leq 14$) and to $e^{-2\tau_{\rm reio}}$ for large l.

Effect of parameters on the temperature spectrum

The dependence of the C_l spectrum on a number of parameters depends of course on the cosmological model considered. In this chapter, for the sake of simplicity, we restrict ourselves exclusively to the minimal ΛCDM model, containing only photons, baryons, electrons, cold dark matter, a cosmological constant, and neutrinos considered as an ultra-relativistic (of negligible mass), whose effects on photon perturbations we will neglect. For the effects of neutrinos (with or without mass) and of the relic density of radiation, the reader may refer to reference [6]. The effects of other parameters, such as the spatial curvature of the Friedmann-Lemaître metric or the parameters describing more complex models for the primordial spectrum, dark matter, dark energy, modified laws of gravity on large scales, etc., are described in the specialised literature.

The total density of ultra-relativistic particles ω_r is not an unknown in the standard cosmological model, it is deduced from the current CMB temperature The minimal ΛCDM model has therefore only six independent parameters. We will choose here the following parameter basis:

$$\{\omega_\mathrm{b}, \omega_\mathrm{m}, \Omega_\Lambda, A_\mathrm{s}, n_\mathrm{s}, \tau_\mathrm{reio}\} \ , \tag{5.75}$$

where ω_b is the density of baryons and ω_m the total density of non-relativistic particles. The first three parameters are sufficient to fully characterise the evolution of the background quantities: the dark matter density is given by $\omega_\mathrm{m} - \omega_\mathrm{b}$, and the reduced Hubble parameter by $h = \sqrt{(\omega_\mathrm{m} + \omega_\mathrm{r})/(1 - \Omega_\Lambda)}$. The parameters A_s and n_s characterise the primordial spectrum $\mathcal{P}_\mathcal{R}$, and the reionisation optical depth τ_reio is sufficient to estimate the impact of reionisation.

The previous sections have proved that the spectrum C_l depends on a certain number of characteristic quantities, whose value can be deduced from the parameters of the basis (5.75):

[C1] The angular scale of the acoustic peaks $\boldsymbol{\theta_\mathrm{s}}$ is given by (see (5.73):

$$\theta_\mathrm{s} = \frac{\pi}{l} = \frac{d_\mathrm{s}(\eta_\mathrm{dec})}{d_\mathrm{a}(\eta_\mathrm{dec})} = \frac{\int_0^{\eta_\mathrm{dec}} c_s(\eta) d\eta}{\int_{\eta_\mathrm{dec}}^{\eta_0} d\eta} \ . \tag{5.76}$$

In the basis (5.75), this angle depends on $\{\omega_\mathrm{b}, \omega_\mathrm{m}, \Omega_\Lambda\}$.

[C2] The angular scale of diffusion damping θ_d obeys:

$$\theta_\mathrm{d} \equiv \frac{\pi}{l} = \frac{\lambda_\mathrm{d}(\eta_\mathrm{dec})}{d_\mathrm{a}(\eta_\mathrm{dec})} = \frac{\left[\int_0^{\eta_\mathrm{dec}} \Gamma_\gamma^{-1}(\eta) d\eta\right]^{1/2}}{\int_{\eta_\mathrm{dec}}^{\eta_0} d\eta} \ . \tag{5.77}$$

Comparing this expression with that of θ_s, we see that the denominator is identical, and therefore always a function of ω_m and Ω_Λ. Since the Thomson scattering rate depends almost only of ω_b, in both cases the numerator is a function of ω_b and ω_m, but with a totally different expression. A variation in one of these parameters therefore has a different impact on the scale of the peaks and on that of the exponential cut-off.

[C3] The baryon-photon ratio at decoupling $\boldsymbol{R|_\mathrm{dec}}$ is given by:

$$R|_\mathrm{dec} = \frac{3}{4} \frac{\bar{\rho}_\mathrm{b}|_\mathrm{dec}}{\bar{\rho}_\gamma|_\mathrm{dec}} = \frac{3}{4} \frac{\omega_\mathrm{b}}{\omega_\gamma} (1 + z_\mathrm{dec}) \ . \tag{5.78}$$

The study of recombination shows that $z_{\rm dec}$ depends very weakly (logarithmically) on the cosmological parameters: thus $R|_{\rm dec}$ depends mainly on ω_b. We saw in section 5.4.2 that this ratio, which determines the impact of gravity forces relative to pressure forces in the photon-electron-baryon fluid at the time of decoupling, controls the asymmetry between the amplitude of the first even and odd peaks.

[C4] The amount of expansion between the time of radiation/matter equality and that of decoupling, $\boldsymbol{a_{\rm dec}/a_{\rm eq}}$, depends on:[15]

$$\frac{a_{\rm dec}}{a_{\rm eq}} = \frac{1+z_{\rm eq}}{1+z_{\rm dec}} = \frac{\omega_{\rm m}}{\omega_{\rm r}(1+z_{\rm dec})} \ . \tag{5.80}$$

With the parameters $\omega_{\rm r}$ and $z_{\rm dec}$ (approximately) fixed, this ratio depends mainly on $\omega_{\rm m}$. It controls two effects. First, the early ISW effect of section 5.4.2 is greater if the perturbations of the metric are not yet fully stabilised at the time of decoupling, and hence, if equality takes place closer to decoupling. Second, there is an interval in k-space such that, if the period between equality and decoupling is long, the modes cross the Hubble radius during matter domination (which starts earlier) and experience a weak gravitational boost, whereas if this period is short, the modes cross the Hubble radius during radiation domination (which ends later) and experience a strong gravitational boost (see section 5.4.2 and figure 5.7). These two effects go in the same direction: a longer period between equality and decoupling (or more precisely, a larger ratio $a_{\rm dec}/a_{\rm eq}$) leads to a decrease in the amplitude of the first and second peaks.

[C5] The amount of expansion during the domination of the cosmological constant $\boldsymbol{a_0/a_\Lambda}$ depends only on Ω_Λ.[16] This ratio determines the magnitude of the late ISW contribution (see section 5.4.2).

So far, we have discussed five distinct effects on the temperature spectrum C_l, governed by only three parameters $\{\omega_{\rm b}, \omega_{\rm m}, \Omega_\Lambda\}$. Thus, we expect that, by measuring the shape of the C_l experimentally, we can infer the value of these three parameters. Besides, there are three other independent effects on the C_l's, controlled by three additional parameters in our basis (5.75):

[C6] The global amplitude of the primordial spectrum $\mathbf{A_s}$ appears as a multiplicative factor in the expression of the temperature spectrum (5.51). An increase in $A_{\rm s}$ therefore increases the overall amplitude of the C_l's.

[C7] The spectral index $\mathbf{n_s}$ controls the overall slope of the C_l, as discussed in section 5.4.2. For instance, an index $n_{\rm s} > 1$ (blue spectrum) amplifies the C_l's at large l (beyond the pivot scale), and attenuates it below.

[C8] The reionisation optical depth $\boldsymbol{\tau_{\rm reio}}$ has an effect explained in the

[15] The redshift $z_{\rm eq}$ of radiation/matter equality is deduced from

$$\rho_r = \rho_{\rm m} \ \Rightarrow \ \omega_{\rm r}\left(\frac{a_0}{a_{\rm eq}}\right)^4 = \omega_{\rm m}\left(\frac{a_0}{a_{\rm eq}}\right)^3 \ \Rightarrow \ 1+z_{\rm eq} = \frac{a_0}{a_{\rm eq}} = \frac{\omega_{\rm m}}{\omega_{\rm r}} \ . \tag{5.79}$$

[16] The redshift z_Λ of matter/Λ equality is deduced from

$$\rho_{\rm m} = \rho_\Lambda \ \Rightarrow \ \Omega_{\rm m}\left(\frac{a_0}{a_\Lambda}\right)^3 = \Omega_\Lambda \ \Rightarrow \ 1+z_\Lambda = \frac{a_0}{a_\Lambda} = \left[\frac{\Omega_\Lambda}{(1-\Omega_\Lambda)}\right]^{1/3}. \tag{5.81}$$

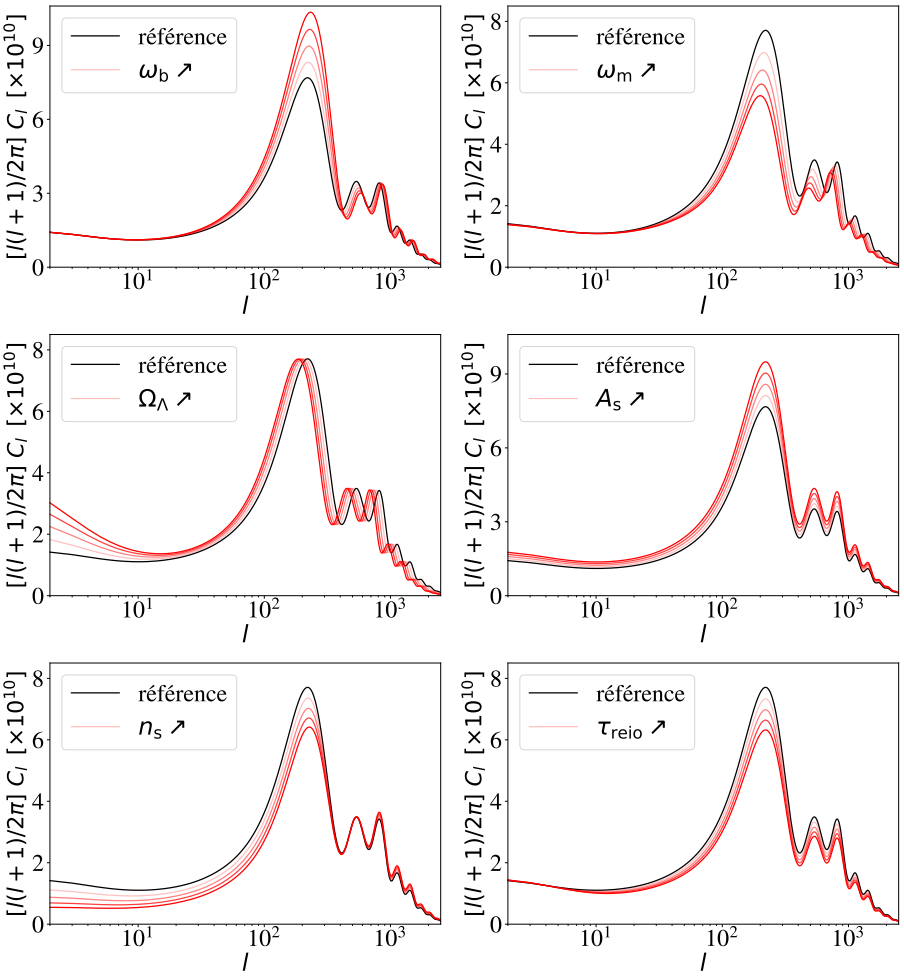

Figure 5.9: Effect on the temperature spectrum C_l of varying one of the ΛCDM parameters in the basis (5.75), all other parameters being fixed. The black spectrum is a reference spectrum. The increasing intensity of red corresponds to increasing values of the free parameter. Obtained with CLASS (`class-code.net`).

section 5.4.2. An increase in $\tau_{\rm reio}$ (with A_s fixed) lowers the normalisation of the C_l's, except for the smallest multipoles $l \leq 14$ which are approximately invariant.

The effect of varying each parameter in the basis (5.75) is summarised in figure 5.9. Each effect can be understood by a careful reasoning based on the previous bullets.

Since in the ΛCDM model, all six parameters have distinct and independent effects, it is theoretically possible to measure them individually from an accurate observation of the temperature map. This conclusion is correct despite the role of cosmic variance, which is important at large angular scales and limits the accuracy with which we measure, for example, the late ISW and reionisation effects.

The case of extended cosmological models is beyond the scope of this chapter.

It is a priori more difficult to constrain models containing more free parameters. Fortunately, many of these additional parameters describe physical ingredients that have other independent effects on the CMB spectrum. This is the case, for example, for the relic radiation density, the total neutrino mass, or several parameters describing non-minimal models of inflation, dark matter and dark energy. The CMB temperature spectrum is therefore a treasure trove of information about the physics that affects – or could possibly affect – the evolution of our universe. This justifies its nickname of Rosetta Stone of the universe.

5.4.3 Overview of CMB polarisation

The CMB can be described at all times as a superposition of electromagnetic waves, each with its own frequency, direction of propagation and polarisation plane. Assuming that the photons were in thermal equilibrium before decoupling, and considering a given point and direction of propagation, the superposition of all frequencies and polarisation planes produces a blackbody spectrum of zero total polarisation, because the polarisation planes are isotropically distributed and their average vanishes.

Around each electron, photons are deflected by Thomson scattering. The conservation of the electric field implies that photons deflected by 90^0 in their own incident plane of polarisation cannot have a transmitted polarisation. Consider a stream of photons coming from a given direction with zero average polarisation. When this stream encounters an electron, the polarisation is transmitted completely for photons deflected by an angle close to zero, but partially for photons deflected by a right angle. Therefore, Thomson scattering is able to induce a net polarisation in some directions. However, as long as the photon-electron-baryon fluid is strongly coupled, each electron perceives a totally isotropic incident photon flux. So all these effects cancel out, and the average polarisation of the transmitted photons in each direction remains zero. On the other hand, as they approach decoupling, the electrons perceive an anisotropic photon flux, which first develops a quadrupole, then an octopole, and so on, as described in section 5.3.1. When the incident photon intensity has a quadrupole component, then the transmitted flux can acquire a non-zero average polarisation in some directions. A net polarisation is thus generated in the vicinity of the last scattering surface at each point and for each direction of propagation.

The CMB photons observed in a given direction originate approximately from a given point $\vec{x}_*$ on the last scattering surface. They therefore have a non-zero mean polarisation, correlated to the quadrupole of the temperature distribution $\Theta_2(\eta_{\rm dec}, \vec{x}_*)$. By placing polarised filters in front of the CMB detectors, it is possible to reconstruct the polarisation map of the CMB. It can be represented in each direction by a stick, whose direction and size indicate the plane and fraction of polarisation of the photons (relative to their intensity). Mathematically, this map forms a spin two field on the last scattering sphere, i.e. a non-oriented vector field. As for a traditional vector field (of spin one), it is possible to decompose this map into two scalar components (of spin zero): an E component describing the curl-free part (as for the electric field), and a B component describing the divergenceless part (as for the magnetic field). Finally, one can in principle obtain three scalar maps: one for the temperature T, one for the mode E, and one for the mode B. For each of the three maps, we can perform a spherical harmonic decomposition and calculate a power spectrum: we thus obtain three spectra C_l^{TT}, C_l^{EE} and C_l^{BB}. Furthermore, we have seen that polarisation is generated by a temperature quadrupole: there

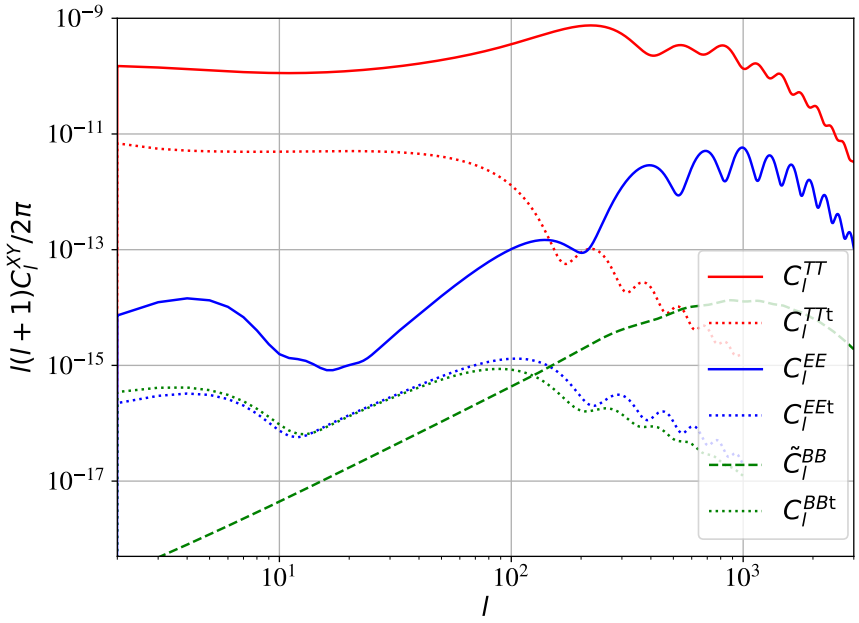

Figure 5.10: *Solid lines*: spectrum of scalar temperature anisotropies $C_l^{(\mathrm{S})TT}$ studied in section 5.4.2, and of the E-mode of scalar polarisation anisotropies $C_l^{(\mathrm{S})EE}$, presented in section 5.4.3. The spectrum of the scalar B-mode, $C_l^{(\mathrm{S})BB}$, is always zero. *Dashed lines*: contribution of the tensor modes to the spectra $C_l^{(\mathrm{T})TT}$, $C_l^{(\mathrm{T})EE}$, $C_l^{(\mathrm{T})BB}$ presented in the section 5.4.4, assuming a tensor-scalar ratio $r = 0.1$ (the overall amplitude of these spectra is proportional to r). *Dashed lines*: contribution to the C_l^{BB} spectrum of a noise generated by gravitational lensing effects, which are not studied in this course. Obtained with CLASS (`class-code.net`).

should therefore be statistical correlations between the temperature map and the polarisation maps, described by cross spectra C_l^{TE} and C_l^{TB}. Since the correlation between temperature and polarisation is only partial, the cross spectra do contain additional information. However, for geometrical (and symmetry) reasons, it is impossible for the scalar modes of cosmological perturbations – which are associated with gradients and an irrotational distribution of matter, see section 3.4 – to generate B modes. In the end, the quantities that can be measured in the presence of only scalar modes – and neglecting some "secondary anisotropies" not studied here – are the T and E maps, as well as the associated spectra C_l^{TT}, C_l^{EE} and C_l^{TE}.

The C_l^{EE} and C_l^{TE} spectra can be predicted theoretically with the same kind of calculation as for C_l^{TT}, but this goes beyond the scope of this lecture (see for example [7]). In the same way that there are transfer functions $\Theta_l(\eta, k)$ for temperature anisotropies, there are transfer functions for polarisation anisotropies. The spectra C_l^{EE} and C_l^{TE} depend on the same cosmological parameters as C_l^{TT}, with a lower sensitivity to some quantities (e.g. to Ω_Λ, because the polarisation transfer function involves no Sachs–Wolfe effect) and a higher sensitivity to others (e.g. to τ_{reio}, because at the time of reionisation some photons are scattered again by electrons in a strongly anisotropic environment, which amplifies the polarisation spectrum at small angular scales). The shape of C_l^{EE} is

shown in figure 5.10. The first peak at $l \sim 5$ comes from re-scattering at the time of reionisation: its amplitude depends directly on τ_{reio}. The other peaks come from acoustic oscillations. Their precise position is slightly different in the temperature and polarisation spectra, because the C_l^{EE} spectrum is not influenced by the Doppler and early ISW effects. We will see in section 5.5 that the C_l^{EE} and C_l^{TE} spectra have been measured and can be compared with theoretical predictions, which further contributes to the measurement of cosmological parameters.

5.4.4 Overview of the role of gravitational waves

We have seen in the section 3.4 that the metric $g_{\mu\nu}$ includes tensor perturbations that correspond to the two degrees of polarisation of gravitational waves in each direction and for each frequency. The chapter 4 shows that during inflation, quantum fluctuations in the metric can generate tensor perturbations that remain imprinted in the metric until the present day. These fluctuations can influence the CMB photons. Indeed, suppose that an observer perceives an isotropic distribution of the photon temperature. If a gravitational wave passes through this location, it produces a distortion of spacetime such that the observer will perceive a quadrupole of temperature Θ_2. Through the Boltzmann hierarchy, this anisotropy will propagate to all other multipoles. Mathematically, this effect is described by a new source term in the Boltzmann equation, involving not only scalar perturbations but also tensor perturbations of the metric.

The corresponding temperature and polarisation anisotropies can be described by spectra $C_l^{(\mathrm{T})XY}$, where $X, Y \in \{T, E, B\}$. For tensor perturbations, there is no symmetry preventing the creation of B modes. These spectra add up to the scalar spectra $C_l^{(\mathrm{S})XY}$ (without statistical correlations between the two sectors) to form the total observable spectra C_l^{XY}.

The global amplitude of the tensor spectra $C_l^{(\mathrm{T})XY}$ depends on the considered inflationary model. It is generally expressed relative to the amplitude of the scalar spectra, thanks to the tensor-to-scalar ratio r defined in the equation (4.88) of chapter 4. Since this parameter is related to the energy scale of inflation, its measurement would provide crucial information to better understand the primordial universe. Figure 5.10 compares the tensor spectra $C_l^{(\mathrm{T})XY}$ with the scalar spectra $C_l^{(\mathrm{S})XY}$ for a hypothetical value of the parameter $r = 0.1$. We will summarise in section 5.5 how the present constraints on the tensor contribution and on the parameter r have been obtained. At the end of section 5.5, we will also explain why a better measurement of the B modes is essential for a future detection of the CMB tensor modes and of the energy scale of inflation.

5.5 Observations of the CMB

Current observations

The CMB anisotropies were first detected by the COBE-DMR experiment, on board the NASA satellite COBE. Since then, many instruments have refined the COBE-DMR measurements considerably. These include ground-based detectors, balloon-borne instruments, and two other satellites: WMAP (NASA) and Planck (ESA). Current data on CMB anisotropies are dominated by results from the Planck satellite, the bulk of which was published in 2018. Planck

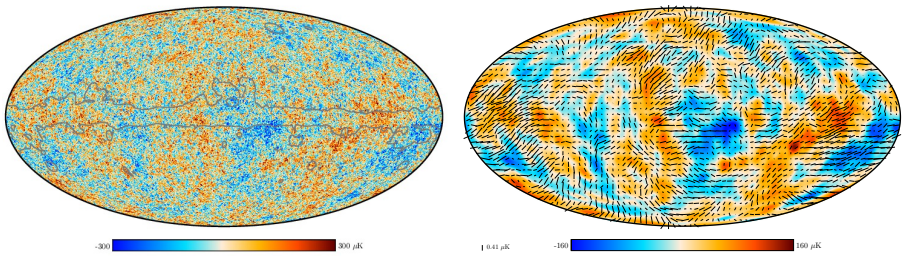

Figure 5.11: All-sky map of temperature anisotropies (left) and polarisation (right) obtained by the Planck collaboration. Extracted from Astron. Astrophys. 641 (2020) A1 (e-Print: 1807.06205 [astro-ph.CO]). Credits: ESA. The equator corresponds to the plane of the Milky Way. The region heavily contaminated by the foregrounds of our galaxy appears in grey. In the polarisation map, the colours show the amplitude of the polarisation, and the sticks the plane of polarisation at each point.

observed the entire sky for four years (2009-2013) and in nine channels (or frequency ranges) covering the interval from 15 GHz to 1000 GHz. However, on large angular scales corresponding to $l \sim 50 - 170$, the Keck and BICEP2 ground-based telescopes have made more accurate measurements of polarisation anisotropies than Planck.

The Planck temperature map reveals a dipole of 3362 μK, which allows us to estimate the velocity of the solar system with respect to the Friedmann reference frame in which the universe is statistically homogeneous and isotropic: $v = 369.8\text{km.s}^{-1}$. After subtracting the monopole and the dipole, we obtain the temperature anisotropy map of figure 5.11 (left). The presence of polarised detectors also gives a polarisation map (figure 5.11, right).

From the maps, it is possible to measure the temperature spectrum C_l^{TT}, the polarisation spectrum C_l^{EE} and C_l^{BB}, and the cross spectrum C_l^{TE}. The constraints obtained by Planck and other experiments on these spectra are shown in figure 5.12. The errors on the C_l^{TT} spectrum are dominated by cosmic variance up to $l \sim 1600$, and then by foreground uncertainties and instrumental noise. This excellent result leaves little room for improvement in the measurement of temperature anisotropies in the future. Planck's errors on the C_l^{EE} spectrum are dominated by cosmic variance only up to $l \sim 800$, while Keck/BICEP2/Planck's errors on the C_l^{BB} spectrum are much larger: a much better measurement of the polarisation is therefore still possible.

Constraints on the minimal ΛCDM model

The most remarkable result of WMAP and Planck is the fact that it is possible to satisfactorily fit all data with the minimal ΛCDM model with six free parameters. Let us first describe the constraints on these six parameters before extending the discussion to tests of the ΛCDM model itself. In the following, unless otherwise stated, all error bars are standard deviations, corresponding to a confidence level of 68%.

The following results are based on a simultaneous fit of the TT, TE, EE spectra (and also, of the spectrum associated to the effect of gravitational lensing of the last scattering surface, which we do not study in this lecture, and which

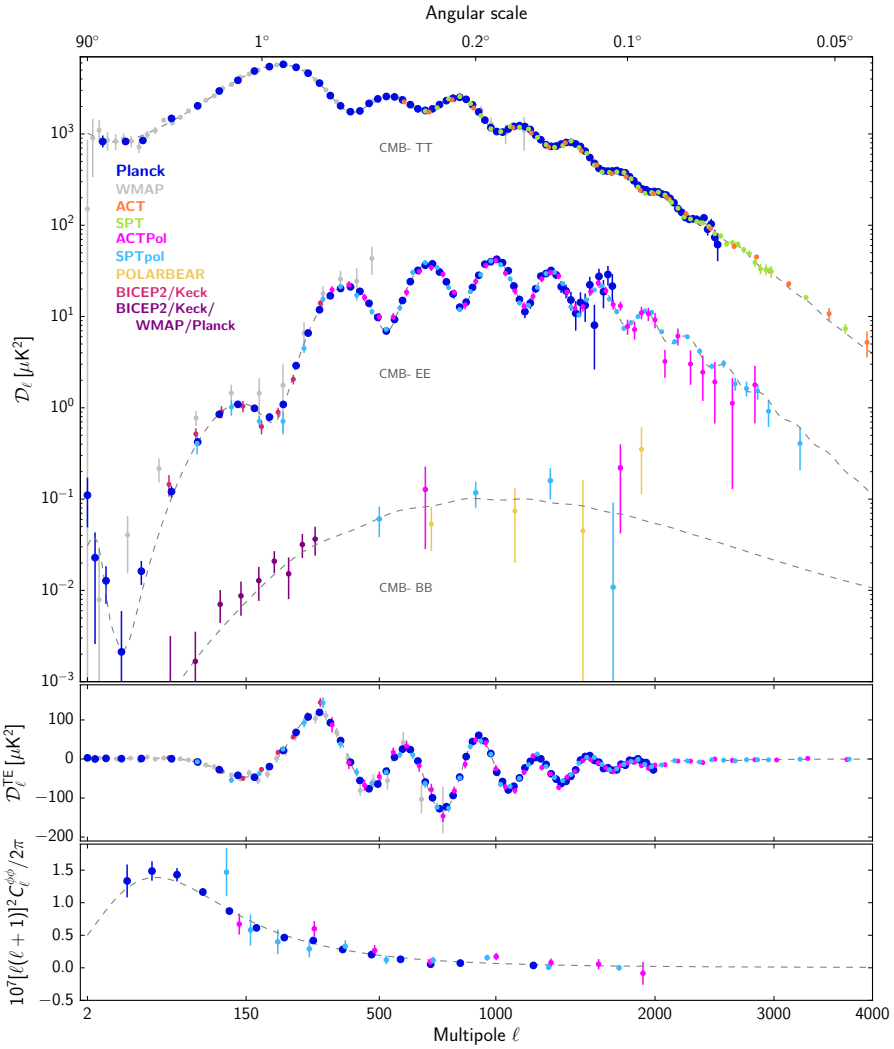

Figure 5.12: Data from Planck, BICEP2/Keck and other experiments on the temperature spectrum C_l^{TT}, on the polarisation spectra C_l^{EE} and C_l^{BB} (top), and on the cross-spectrum C_l^{TE} (middle). (The bottom spectrum has to do with gravitational lensing effects not discussed in this course). Extracted from Astron. Astrophys. 641 (2020) A1 (e-Print: 1807.06205 [astro-ph.CO]). For temperature and polarization, the quantities plotted here are the renormalized spectra $\mathcal{D}_l^{XY}$ related to the spectra C_l^{XY} defined in this chapter by $\mathcal{D}_l^{XY} \equiv l(l+1)C_l^{XY}T_0^2/(2\pi)$.

increases slightly the sensitivity to some parameters).

• **The amplitude and spectral index of the primordial spectrum** are constrained by $\ln(10^{10}A_s) = 3.044 \pm 0.014$ and $n_s = 0.9649 \pm 0.0042$. This confirms the prediction by inflation of a nearly scale-invariant spectrum ($n_s = 1$).

• **The baryon density** is given by $\omega_b = 0.02237 \pm 0.00015$, in remarkable agreement with the value deduced from the observation of the relative density of hydrogen, deuterium and helium in the universe and the standard model of primordial Nucleosynthesis studied in chapter 2, section 2.2.5. This agreement

between two completely different and independent ways of measuring the same parameter is one of the most spectacular achievements of modern cosmology.

• **The density of cold dark matter** is given by $\omega_c = 0.1200 \pm 0.0012$. The CMB alone therefore offers a proof of the existence of dark matter at 100σ. In fact, we have seen in section 5.4.2 that the CMB independently constrains the density of baryons and of total non-relativistic matter, $\omega_m = \omega_b + \omega_c$. Dark matter is therefore required by the CMB to explain a missing mass phenomenon: $\omega_m \gg \omega_b$. This is a conclusion analogous to the historical argument for dark matter from the dynamics of galaxies and clusters.

• **The fractional density of cosmological constant** is $\Omega_\Lambda = 0.6847 \pm 0.0073$. Under the assumption of the ΛCDM model, the CMB alone implies the existence of a non-zero cosmological constant (or equivalent dark energy) at 94σ, and conforms the results of the experiments that measure the acceleration of the expansion of the universe.

• **The reionisation optical depth** is $\tau_{reio} = 0.0544 \pm 0.0073$. This is consistent with the hypothesis of a reionisation around the redshift $z \sim 7$, suggested by the analysis of absorption lines in the quasar spectrum.

Beyond the minimal ΛCDM model

These constraints were obtained after intensive investigations to assess the validity of the Friedmann-Lemaître model in general, and the ΛCDM model in particular.

• A battery of tests has been carried out to check the assumption of a homogeneous and isotropic universe, at least up to the largest scales observed by the COBE, WMAP and Planck satellites, i.e. up to the diameter of our observable universe. To a first approximation, isotropy is well established (some cosmologists are interested in possible anomalies at large scales, but it is very difficult to establish whether they are statistically significant).

• An assumption of the minimal ΛCDM model is that the radiation density is equal to $\omega_r = 4.168 \times 10^{-5}$ (equation (2.63)). This assumption can be relaxed by introducing a free parameter: the effective neutrino number, N_{eff}, whose effects on the CMB are described in [6]. However, the Planck data gives $N_{eff} = 2.99 \pm 0.34$, which is perfectly compatible with the assumption of the minimal model ($N_{eff} \simeq 3$).

• The minimal ΛCDM model also assumes that the primordial spectrum is a power law. Numerous alternatives have been tested without any significant deviation from this assumption being detected.

• The minimal ΛCDM model is based on a Friedmann-Lemaître metric with zero spatial curvature. A non-zero curvature can be parameterised by $\Omega_k = 1 - \Omega_m - \Omega_r - \Omega_\Lambda$. A variation of Ω_k would influence the temperature and polarisation spectra, mainly via θ_s, θ_d and the late ISW effect. However, the position of the acoustic peaks is in excellent agreement with the assumption of zero curvature: Planck gives $\Omega_k = -0.011 \pm 0.006$. The case of a universe with a slight positive curvature remains possible and interesting, but is not required by current observations.

• Tests have been carried out to determine whether dynamical dark energy (i.e., a slowly time-evolving species with w close to -1 but not necessarily exactly equal

to -1 at all times) could explain the data better than a plain cosmological constant, but again without positive results.

• Moreover, the CMB spectra are sensitive to the total neutrino mass summed over the three mass eigenstates, $\sum m_\nu$, mainly due to gravitational lensing of the temperature and polarisation spectra, but also via θ_s, θ_d and early and late ISW effects, as explained in [6]. The Planck data impose $\sum m_\nu < 0.12$ eV.

• In principle, tensor modes (i.e. primordial gravitational waves) could have been detected as a contribution to the temperature spectrum at $l < 100$ (see section 5.4.4 and figure 5.10). The WMAP and Planck data show no such contribution, which rules out the possibility of detecting a non-zero value of the tensor-to-scalar ratio in the future using the temperature spectrum alone. On the other hand, detection is still possible using the polarisation spectra, and in particular with better measurements of C_l^{BB}. Indeed, the spectrum C_l^{BB} only receives a contribution from the tensor modes. In reality, there is a second contribution to C_l^{BB} coming from an effect of gravitational lensing, that we will not treat in this course; but this second contribution can be disentangled and removed up to some extent. The best current constraints come from the combination of the Keck, BICEP2 and Planck experiments. This results in an upper bound $r < 0.05$ which has profound implications for inflationary models.

Chapter 6

Large Scale Structure of the Universe

6.1 Linear matter power spectrum

6.1.1 Definition and range of validity

In chapter 5, the main observable was the CMB temperature map $\frac{\delta T}{T}(\hat{n})$, a two-dimensional function on a sphere. In order to express it as a sum over independent modes that can be predicted by the theory, we performed an expansion in spherical harmonics. Form the theory point of view, the spherical harmonic coefficients a_{lm} should be stochastic variables with a Gaussian distribution of probability, for which we can predict the variance. This led us to the definition of the CMB power spectrum $C_l = \langle |a_{lm}|^2 \rangle$. This spectrum is the most important quantity for CMB physics, since it can both be predicted theoretically and measured from observations.

In this chapter, we are interested in the large scale structure of the universe, described by the density fluctuation of non-relativistic matter (CDM, baryon, and possibly more in some scenarii),

$$\delta_{\mathrm{m}}(t, \vec{x}) = \frac{\delta\rho_{\mathrm{m}}(t, \vec{x})}{\bar{\rho}_{\mathrm{m}}(t)} \ . \tag{6.1}$$

At each time, this three-dimensional function can be expanded in Fourier modes $\delta_{\mathrm{m}}(t, \vec{k})$. As long as cosmological perturbations undergo a linear evolution, from the theory point of view, these Fourier modes should be stochastic variables, independent from each other, and obeying a Gaussian distribution of probability for which we can predict the variance. This variance is just called the matter power spectrum P_{m}. It is a very important quantity since it can both be measured from observation and predicted by the theory. Following the definition of power spectra in 4.3.5, we know that we can define the matter power spectrum P_{m} as

$$P_{\mathrm{m}}(t, k) = \langle |\delta_{\mathrm{m}}(t, \vec{k})|^2 \rangle \tag{6.2}$$

in the case of discretised Fourier modes $\vec{k}$, while in the continuous Fourier picture we would use

$$\langle \delta_{\mathrm{m}}(t, \vec{k}) \delta^*_{\mathrm{m}}(t, \vec{k}') \rangle = P_{\mathrm{m}}(t, k) \ \delta(\vec{k} - \vec{k}') \ . \tag{6.3}$$

We already discussed the fact that the matter power spectrum depends on the wavenumber k and not on the full wave-vector $\vec{k}$ as a consequence of the statistical isotropy of the Universe in the FLRW model.

Applying linear perturbation theory clearly made sense in the case of CMB physics: primordial fluctuations are of the order of 10^{-5}, and until the time of photon decoupling there is no mechanism leading to a significant growth of the fluctuations that are most important for the CMB (the photon temperature anisotropy Θ and the metric fluctuations ϕ and ψ). However, during matter domination, we expect gravitational collapse to enhance the non-relativistic matter fluctuation $\delta_{\rm m}$. Thus, we wonder whether linear perturbation theory can be applied today to the description of the matter power spectrum, and if yes, down to which scale.

To answer this question, we must introduce the notion of smoothed density fluctuations. In the course of this chapter, we will briefly discuss several techniques to measure the density fluctuation $\delta_{\rm m}$ and the matter power spectrum $P_{\rm m}$. They all involve at some point a smoothing over small scales. It means that we never measure the full field $\delta_{\rm m}(t,\vec{x})$ but only the field $\tilde{\delta}_{{\rm m},R}$ smoothed over a given scale R,

$$\tilde{\delta}_{{\rm m},R}(t,\vec{x}) = \int d^3\vec{x}'\, W_R(\vec{x}'-\vec{x})\, \delta_{\rm m}(t,\vec{x}') \ , \tag{6.4}$$

where W_R is a smoothing kernel (also called window function or Wiener filter) over the length R. There are different possible smoothing kernels. A well-known one is the Gaussian smoothing kernel $W_R(\vec{x}) \propto \exp(-\frac{1}{2}\frac{|\vec{x}|^2}{R^2})$. If we have measured a smoothed field $\tilde{\delta}_{{\rm m},R}(t,\vec{x})$, we can measure its variance σ^2, which only depends on time and on the smoothing length R,

$$\sigma_R^2(t) = \langle |\tilde{\delta}_{{\rm m},R}(t,\vec{x})|^2 \rangle_{\vec{x}} \ . \tag{6.5}$$

From the theory point of view, once we have predicted the matter power spectrum, we can predict the variance $\sigma_R^2(t)$ as an integral over the power spectrum,

$$\sigma_R^2(t) = \int \frac{d^3\vec{k}}{(2\pi)^3}\ W_R^2(\vec{k})\ P_{\rm m}(t,k) \ , \tag{6.6}$$

where $W_R(\vec{k})$ is the Fourier transform of $W_R(\vec{x})$.[1] Again, there are different possible expressions for this kernel, but they are all such that $W_R(k)$ is negligible in the limit $kR \ll 1$, to account for the fact that the variance comes from a sum over wavelengths smaller than R, but does not depend on larger wavelengths. We expect that small wavelengths cross the Hubble radius earlier, and thus, that $\delta_{\rm m}(t,k)$ starts to grow earlier for larger k: small scales experience gravitational

[1] Since a convolution in real space gives a product in Fourier space, it is trivial to show that at fixed time t,

$$\tilde{\delta}_{{\rm m},R}(\vec{k}) = W_R(\vec{k})\ \delta_{\rm m}(\vec{k}) \, .$$

The theory point of view consists in replacing averages over $\vec{x}$ by averages over realisations of the stochastic theory: this must be correct in a universe where we assumed statistical homogeneity. Then,

$$\sigma_R^2 = \langle \tilde{\delta}^2_{{\rm m},R}(\vec{x}) \rangle = \int \frac{d^3\vec{k}\, d^3\vec{k}'}{(2\pi)^3} W_R(\vec{k})\ W_R(\vec{k}')\ \langle \delta_{\rm m}(\vec{k})\ \delta^*_{\rm m}(\vec{k}') \rangle e^{-i(\vec{k}-\vec{k}')\vec{x}} \ .$$

After replacing $\langle \delta_{\rm m}(\vec{k})\ \delta^*_{\rm m}(\vec{k}') \rangle$ by $P_{\rm m}(k)\ \delta^{(3)}(\vec{k}'-\vec{k})$ and integrating, one gets Eq. (6.6). Note that one usually chooses isotropic window functions depending only on the modulus of $\vec{x}$ and $\vec{k}$.

collapse earlier than large scales. Therefore, at a given time, it is plausible that small scales entered the regime of non-linear perturbation theory, while large scale are still linear. We can thus define a scale of non-linearity $\lambda_{\rm NR}$ such that $\tilde{\delta}_{\rm m}(t,\vec{x},R)$ smoothed over any scale $R \gg \lambda_{\rm NR}$ is a linearly perturbed field obeying $|\tilde{\delta}_{\rm m}(t,\vec{x},R)| \ll 1$ everywhere. Instead, $\tilde{\delta}_{\rm m}(t,\vec{x},R)$ smoothed over any scale $R \ll \lambda_{\rm NR}$ should be a non-linear field with $|\tilde{\delta}_{\rm m}(t,\vec{x},R)| \sim \mathcal{O}(1)$. This definition of the non-linearity scale can be done directly at the level of the variance $\sigma^2(t,R)$: this variance should be smaller than one for $R \gg \lambda_{\rm NR}$, and $\lambda_{\rm NR}$ can be defined as the largest scale such that $\sigma^2(t,R) \sim 1$ (or 0.1 or a similar number: this is just a matter of definition).

This is the typical way in which the scale of non-linearity is defined (with small differences between different conventions). Today, this scale is typically of the order of $\sim$20 Mpc (to be compared to the typical distance between neighbouring galaxies, $\sim$1 Mpc).

Is it possible to map the large-scale structure of the universe on scales bigger than the non-linearity scale? Actually, it is, because the theoretical limit on the largest scale that we can map is the observable radius of the Universe $R_{\rm obs}$, known to be of the order of $R_H(t_0) = 4000$ Mpc. So, in theory, we can reconstruct a linear density field containing several scales spanning about two decades, from $\lambda_{\rm min} \sim 20 - 40$ Mpc to $\lambda_{\rm max} \sim 4000$ Mpc. In practise, it is very difficult to map the large-scale structure up to $R_{\rm obs}$, but we do have better and better maps covering about one decade of linear scales, up to several hundreds of Mpc.

Moreover, we observe large scale structures in our past-light cone: we map remote objects that emitted light in the past, at some redshift $z > 0$. We can see millions of objects at high redshifts of $z = 1$, 2, 3, and even more. These high redshifts correspond to early times at wich the scale of non-linearity was smaller, and the range of validity of linear perturbation theory was wider.

Thus, being able to predict the linear matter power spectrum is very useful for several reasons:

- we can directly compare the predicted linear matter power spectrum with the power spectrum inferred from large-scale structure maps provided that these maps are smoothed over a scale of the order of 20 Mpc at $z = 0$, and even less at high redshift,

- when using a smaller smoothing scale in order to keep more information, we need to model the non-linear evolution of $\delta_{\rm m}$ and predict a non-linear matter power spectrum. But even for this problem, we need to understand first the linear matter power spectrum, which plays the role of an initial condition.

According to linear perturbation theory, $\delta_{\rm m}(t,\vec{k})$ can be decomposed into a transfer function times an initial curvature perturbation. Using the discrete Fourier convention, we can then express the matter power spectrum as:

$$\begin{aligned} P_{\rm m}(t,k) &= \langle|\delta_{\rm m}(t,\vec{k})|^2\rangle = \delta^2_{\rm m}(t,k)\ \langle|\mathcal{R}(\vec{k})|^2\rangle & (6.7)\\ &= \delta^2_{\rm m}(t,k) P_{\mathcal{R}}(k) = \delta^2_{\rm m}(t,k)\ \frac{2\pi^2}{k^3}\ \mathcal{P}_{\mathcal{R}}(k) & (6.8)\\ &= \delta^2_{\rm m}(t,k)\ \frac{2\pi^2}{k^3}\ A_s \left(\frac{k}{k_*}\right)^{n_s-1} & (6.9)\\ &\propto \delta^2_{\rm m}(t,k)\ A_s\ k^{n_s-4}\ , & (6.10) \end{aligned}$$

where in the last two lines we assumed a standard inflationary model leading to a power-law primordial spectrum. The fluctuation $\delta_{\rm m}$ is defined as a fluctuation in the density field of non-relativistic matter, that is, in the minimal ΛCDM model, of baryons and cold dark matter (denoted respectively with a subscript 'b' or 'c'). Thus, the transfer function $\delta_{\rm m}^2(t,k)$ can be expanded as:

$$\delta_{\rm m} \equiv \frac{\delta\rho_{\rm m}}{\bar\rho_{\rm m}} = \frac{\delta\rho_{\rm c}+\delta\rho_{\rm b}}{\bar\rho_{\rm c}+\bar\rho_{\rm b}} = \frac{\bar\rho_{\rm c}\ \delta_{\rm c}+\bar\rho_{\rm b}\ \delta_{\rm b}}{\bar\rho_{\rm c}+\bar\rho_{\rm b}} \tag{6.11}$$

$$= \frac{\bar\rho_{\rm c,0}\ \delta_{\rm c}+\bar\rho_{\rm b,0}\ \delta_{\rm b}}{\bar\rho_{\rm c,0}+\bar\rho_{\rm b,0}} = \frac{\Omega_{\rm c}\ \delta_{\rm c}+\Omega_{\rm b}\ \delta_{\rm b}}{\Omega_{\rm c}+\Omega_{\rm b}}\ . \tag{6.12}$$

To understand the matter power spectrum, we can follow the following strategy:

1. Understand semi-analytically the individual behaviour of the transfer functions $\delta_{\rm c}(t,k)$ and $\delta_{\rm b}(t,k)$.
2. Infer the behaviour of their weighted sum $\delta_{\rm m}(t,k)$.
3. Multiply $\delta_{\rm m}^2(t,k)$ by the primordial spectrum $P_{\mathcal R}(k)$ to infer the behaviour of the matter power spectrum.

We will follow these steps in the next two subsections.

6.1.2 Theoretical prediction neglecting baryons

For simplicity, we will first work in a Universe containing a negligible baryon fraction, $\Omega_{\rm b} \ll \Omega_{\rm c}$, such that $\delta_{\rm m}(t,k) \simeq \delta_{\rm c}(t,k)$. Thus, we only need to understand the behaviour of the transfer function $\delta_{\rm c}(t,k)$ in a universe dominated first by radiation (photons and neutrinos), then by matter (assumed to be cold dark matter only) and finally by Λ (a background component that does not have spatial fluctuations).

We discussed the evolution of $\delta_{\rm c}$ (or more generally of the total non-relativistic fluctuation $\delta_{\rm m}$) superficially at the beginning of section 5.4.1, and in more in details in one of the last exercise sheets. Let us discuss it explicitely here. By combining the continuity and Euler equation of CDM perturbations, which are really simple since, for non-relativistic matter, one has $w_x = c_{s,x} = c_{a,x} = \sigma_x = 0$, we can write an equation of motion for $\delta_{\rm m}$, valid at all times and in all regimes (super- and sub-Hubble):

$$\delta_{\rm c}'' + \frac{a'}{a}\delta_{\rm c}' = -k^2\psi + 3\phi'' + 3\frac{a'}{a}\phi'\ . \tag{6.13}$$

The physical interpretation is simple. In an expanding universe, the clustering rate depends on the expansion rate: expansion increases distances, weakens gravitational forces, and slows down clustering processes. In the above equation, this is accounted by the second term, the usual Hubble friction term. On the right-hand side, the first term represents gravitational forces, and the last two terms account for dilation and for the local modulation of Hubble friction. There is no term linear in $\delta_{\rm c}$: in general, there should be a term $c_s^2k^2\delta_{\rm c}$ accounting for pressure forces, but CDM is pressureless.

On super-Hubble scales, we have seen that in the Newtonian gauge, adiabatic ICs predict constant density fluctuations $\delta_{\rm c}$. To be precise, we have seen in previous exercises that ϕ and $\delta_{\rm c}$ vary on super-Hubble scales only when the total equation of state of the universe changes (i.e. around the time of radiation/matter equality, and during Λ domination). Instead, they remain constant

on those scale during the radiation and matter dominated regime. When writing the adiabatic ICs during RD, we found the following relation with primordial curvature fluctuations: $\delta_{\rm c}(\eta,\vec{k}) = -\mathcal{R}(\vec{k})$. Thus the transfer function normalised to $\mathcal{R}=1$ simply reads $\delta_{\rm c}(\eta,k) = -1$.

Deep inside the Hubble radius, one can show that the terms with the derivative of the metric are always very small compared to those with the Laplacian of the metric, such that

$$\delta_{\rm c}'' + \frac{a'}{a}\delta_{\rm c}' = -k^2\psi \ . \tag{6.14}$$

During RD, we know that the metric is driven to zero on sub-Hubble scales. The evolution of $\delta_{\rm c}$ is thus governed by

$$\delta_{\rm c}'' + \frac{a'}{a}\delta_{\rm c}' \simeq 0 \ , \tag{6.15}$$

with the general solution

$$\delta_{\rm c} = A + B\ln\eta = \tilde{A} + \tilde{B}\ln a \ , \tag{6.16}$$

where (A,B) or $(\tilde{A},\tilde{B})$ are integration constants (remember that $a \propto \eta$ during RD). Thus, in this regime, CDM fluctuations grow logarithmically, that is, very slowly. The reason is that CDM is not yet a self-gravitating fluid experiencing gravitational collapse. It is only a test fluid in an environment dominated by relativistic particles (photons, neutrinos). The gravitational potential responds to the latter particle, which do not cluster. Thus, CDM can fall in potential wells but cannot contribute to make the potential wells deeper. This is the fundamental reason for which the growth of $\delta_{\rm c}$ is so slow at this stage.

During matter domination, we cannot neglect the term $-k^2\psi$ on the rhs of the evolution equation, but since CDM now dominates gravitational forces, we can replace it in terms of δ_c using the Poisson equation, as already seen in the exercises. We obtain:

$$\delta_{\rm c}'' + \frac{a'}{a}\delta_{\rm c}' - \frac{3}{2}\left(\frac{a'}{a}\right)^2 \delta_{\rm c} = 0 \ , \tag{6.17}$$

where the second term accounts for the expansion effect and the third term for gravity forces in the self-gravitating CDM fluid. Using the fact that $a \propto \eta^2$ during MD we get the general solution

$$\delta_{\rm c} = C\eta^{-3} + D\eta^2 = \tilde{C}a^{-3/2} + \tilde{D}a \ , \tag{6.18}$$

where (C,D) or $(\tilde{C},\tilde{D})$ are integration constants. The C solution is a decaying mode that we can neglect, but the D solution is a growing mode: we find that $\delta_{\rm c}$ grows linearly with the scale factor during MD on sub-Hubble scale, which is a very famous and important result that you should remember to understand all the rest of this chapter. Thus, in this regime, CDM fluctuations grow efficiently, since now CDM is a self-gravitating fluid experiencing gravitational collapse. The gravitational potential responds to CDM. CDM accumulates in potential wells, which makes these wells deeper, so CDM accumulates even more, and so on and so forth. Note that growth rate of $\delta_{\rm c}$ does not depend on k (i.e., $\delta_{\rm c}'/\delta_{\rm c} = a'/a$ is independent of k).

Finally, to account for Λ domination, we just need to write the Poisson equation more carefully, noticing that $\bar{\rho}_{\rm tot}$ is now given by $\bar{\rho}_{\rm c} + \bar{\rho}_\Lambda$. We can write the result as

$$\delta_{\rm c}'' + \frac{a'}{a}\delta_{\rm c}' - \frac{3}{2}\left(\frac{a'}{a}\right)^2 \Omega_{\rm c}(a)\ \delta_{\rm c} = 0 \ , \tag{6.19}$$

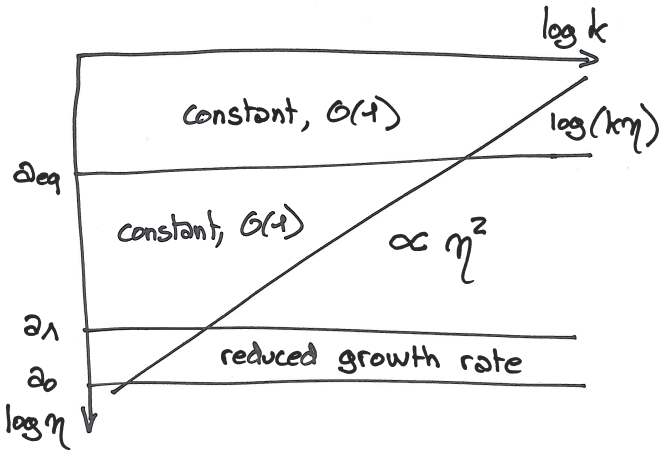

Figure 6.1: Qualitative evolution of the transfer function $\delta_{\rm c}(\eta, k)$ (normalized as usual to $\mathcal{R}(k) = 1$ at initial time) in different regimes: during radiation domination, matter domination, Λ domination, and on super/sub-Hubble scales.

where $\Omega_{\rm c}(a)$ is the fraction of the critical density coming from CDM at a given value of time (or of the scale factor). During Λ domination, the function $a(\eta)$ is more complicated than during MD, and $\Omega_{\rm c}(a)$ decreases from 1 to approximately 0.7 today. With a bit of work, one can show that $\delta_{\rm c}(\eta)$ grows at a smaller rate than during matter domination (i.e. slower than η^2), and that the growth rate still does not depend on k.

In summary, during radiation domination, $\delta_{\rm c}(\eta, k)$ is constant on super-Hubble scales and grows logarithmically on sub-Hubble scales. A more precise calculation would show that up to a numerical factor of order one, $\delta_{\rm c}$ is given on sub-Hubble scales by $\delta_{\rm c}(\eta, k) = \log(k\eta)$. During matter domination, $\delta_{\rm c}(\eta, k)$ is still constant on super-Hubble scales, and grows like η^2 on sub-Hubble scales. Finally, during Λ domination, it grows more slowly. These different behaviors are reported in Fig. 6.1.

This simple discussion is sufficient for understanding the shape of the matter power spectrum at different time. Like in a cartoon, Fig. 6.2 shows this shape at four different times: at some initial time when all relevant modes are super-Hubble; at radiation/matter equality; at matter/Λ equality; and today. Let us comment these plots. We first need to define the comoving wavenumbers corresponding to wavelengths crossing the Hubble radius at the time of radiation/matter equality, of matter/Λ equality, and today:

$$k_{\rm eq} = a_{\rm eq} H_{\rm eq} \ , \qquad k_\Lambda = a_\Lambda H_\Lambda \ , \qquad k_0 = a_0 H_0 \ . \tag{6.20}$$

We can now review the evolution of $P(k)$ with respect to time, following the same steps as in Fig. 6.2.

1. At initial time, if we assume a scale-invariant spectrum with $n_{\rm s} = 1$, we know from (6.10) that $P_{\rm m}(k) \propto k^{-3}$, with an amplitude given by $A_{\rm s}$.

2. During radiation domination, modes grow logarithmically inside the Hubble radius, like $\log(k\eta)$. At equality, super-Hubble modes ($k_{\rm eq}\eta \ll 1$) are still shaped like at initial time, while sub-Hubble modes ($k_{\rm eq}\eta \gg 1$) have

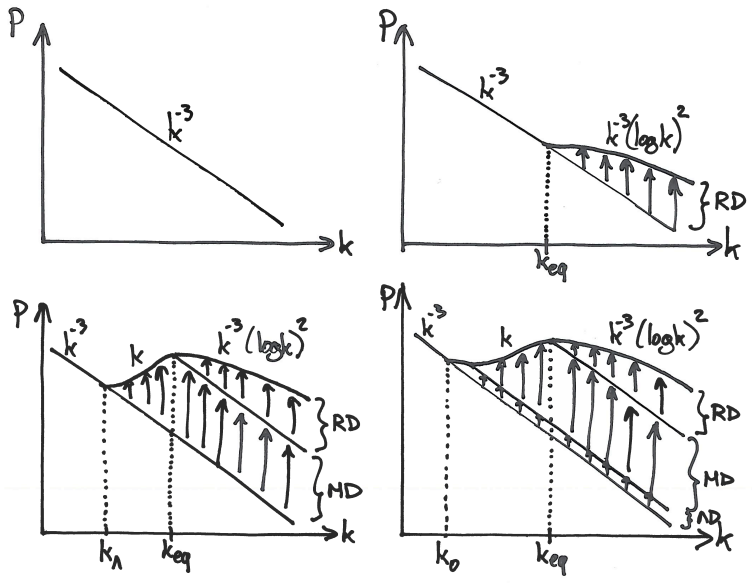

Figure 6.2: Shape of the matter power spectrum $P(k)$ (log-log scale) at four different times: *(upper left)* when initial conditions are imposed (and all wavenumbers are super-Hubble); *(upper right)* at radiation/matter equality (arrows show the logarithmic growth during radiation domination); *(lower left)* at matter/Λ equality (lower set of arrows show the growth during matter domination); *(lower right)* today (lower set of arrows show the growth during Λ domination).

been enhanced by a factor $[\delta_{\rm c}(\eta_{\rm eq},k)/\delta_{\rm c}(\eta_{\rm ini},k)]^2 \simeq [\log(k\eta_{\rm eq})]^2$. The two asymptotes of $P_{\rm m}(k)$ are then given by k^{-3} for $k \ll k_{\rm eq}$ and $k^{-3}[\log(k)]^2$ for $k \gg k_{\rm eq}$.

3. At the end of matter domination, when $a = a_\Lambda$, modes still outside the Hubble radius keep being shaped like at initial time. This concerns all modes with $k \ll k_\Lambda$. Modes $k \gg k_{\rm eq}$ have been amplified during matter domination by a factor $[\delta_{\rm c}(\eta_\Lambda,k)/\delta_{\rm c}(\eta_{\rm eq},k)]^2 \simeq (\eta_\Lambda/\eta_{\rm eq})^4$. This factor does not depend on k and preserves the shape of the power spectrum on those scales. Finally, intermediate modes entering the Hubble scale during matter domination have been amplified by $(\eta_\Lambda/\eta_*)^4$, where η_* is their time of Hubble crossing, given approximately by $\eta_* = 1/k$. Hence they have been amplified by $(k\eta_\Lambda)^4$. Putting all these informations together, we see that the spectrum has three branches, scaling respectively like:

   - $P_{\rm m}(k) \propto k^{-3}$ for $k < k_\Lambda$,
   - $P_{\rm m}(k) \propto k^{-3}k^4 = k$ for $k_\Lambda < k < k_{\rm eq}$,
   - $P_{\rm m}(k) \propto k^{-3}(\log k)^2$ for $k > k_{\rm eq}$.

4. During Λ domination, $\delta_{\rm c}(k,\eta)$ grows more slowly than η^2, but it still grows at the same rate for all sub-Hubble modes. So the shape of the power spectrum today is unaltered by this stage, and given by:

   - $P_{\rm m}(k) \propto k^{-3}$ for $k < k_0$,
   - $P_{\rm m}(k) \propto k^{-3}k^4 = k$ for $k_\Lambda < k < k_{\rm eq}$,
   - $P_{\rm m}(k) \propto k^{-3}(\log k)^2$ for $k > k_{\rm eq}$.

   We do not enter into details for the small range of modes obeying $k_0 < k < k_\Lambda$.

All this discussion was carried under the assumption of a scale-invariant primordial spectrum. If $n_{\rm s} \neq 1$, the above shape should simply be rescaled by $k^{n_{\rm s}-1}$, and the three branches of the power spectrum are given by:

- $P(k) \propto k^{n_{\rm s}-4}$ for $k < k_0$,
- $P(k) \propto k^{n_{\rm s}}$ for $k_\Lambda < k < k_{\rm eq}$,
- $P(k) \propto k^{n_{\rm s}-4}(\log k)^2$ for $k > k_{\rm eq}$.

We will study observational techniques later on. In any case, it is clear that we cannot see modes with $k < k_0$ wich are of the order of the observable radius of the universe or bigger. Thus, we expect observations to return a matter power spectrum with two branches: a large-scale (small-k) branch growing like $k^{n_{\rm s}}$, a broad peak around $k \simeq k_{\rm eq}$, and a small-scale (large-k) branch decreasing like $k^{n_{\rm s}-4}(\log k)^2$. This is exactly what observations tell us.

It is also clear that the matter power spectrum contains information on the cosmological parameters:

- The overall amplitude should depend on A_s (through the primordial spectrum), on the total time that perturbations had to grow, which depends on η_0 or ω_m, and on the duration of Λ domination, which depends on Ω_Λ.
- The overall slope of the spectrum on all scales depends on the tilt n_s.
- The location of the broad peak depends on $k_{\rm eq}$, that is, on $\omega_{\rm m}$.

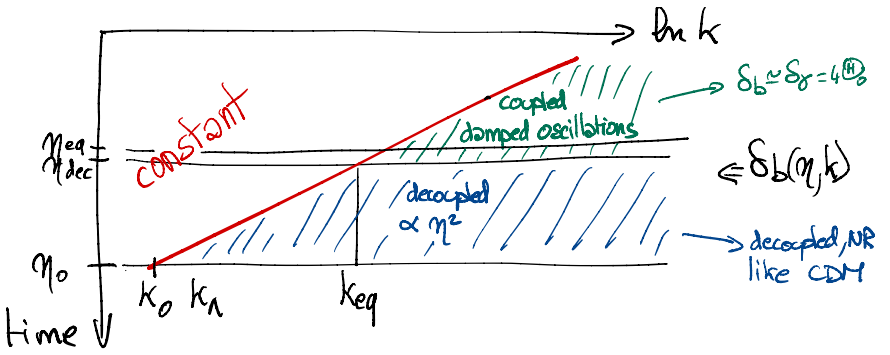

Figure 6.3: Qualitative evolution of the baryon transfer function.

This closes the presentation of the shape of the linear matter power spectrum $P_{\rm m}(\eta_0, k)$ in the limit $\Omega_{\rm b} \ll \Omega_{\rm c}$. The shape of $P_{\rm m}(\eta, k)$ at some slightly earlier time would be the same with just a slightly different overall amplitude. What remains to be seen is the impact of a non-negligible baryon fraction on the power spectrum.

6.1.3 Baryon acoustic oscillations

Let us now take into account the fact that the total non-relativistic matter density includes a baryonic component, with relative fluctuations $\delta_{\rm b}$. We work in the approximation in which baryons and photons decouple instantaneously from each other at the time of decoupling $\eta_{\rm dec}$.[2] The qualitative behaviour of $\delta_{\rm b}$ is summarised in figure 6.3:

- Until decoupling, baryons are tightly coupled to photons, such that

$$\delta_{\rm b} = \frac{3}{4}\delta_\gamma = 3\Theta_0 \ , \tag{6.21}$$

  and we know that Θ_0 experiences damped oscillations within the Hubble radius, with $\Theta_0 \longrightarrow 0$ when $k\eta \longrightarrow \infty$.

- After decoupling, baryons play the role of a decoupled non-relativistic matter component with the same equations of motion as CDM and the same solutions (with a growing mode $\delta_{\rm b} \propto a \propto \eta^2$).

A simple way to explain the role of baryons is to consider that after decoupling, there is no need to distinguish between baryons and CDM: we can treat them as a single effective fluid 'm' obeying to the equation of motion of non-relativistc matter fluctuations (6.19),

$$\delta_{\rm m}'' + \frac{a'}{a}\delta_{\rm m}' - \frac{3}{2}\left(\frac{a'}{a}\right)^2 \Omega_{\rm m}(a)\ \delta_{\rm m} = 0 \ . \tag{6.22}$$

The only difference with respect to the discussion of the previous subsection is that we must take initial conditions for this equation at $\eta_{\rm dec}$ given by the

[2] In a more accurate approach, we would make a subtle distinction between the "photon decoupling time" $\eta_{\rm dec}$ at which the probability that photons interact becomes negligible and the "baryon drag time" $\eta_{\rm drag}$ at which the acceleration of baryons is influenced more by gravitational force stronger than by Thomson interactions. This $\eta_{\rm drag}$ is slightly bigger than $\eta_{\rm dec}$, but this can be ignored in first approximation.

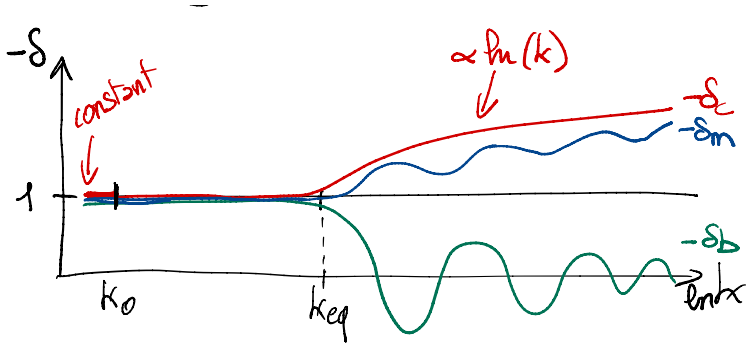

Figure 6.4: Shape of the transfer functions $\delta_{\rm c}$, $\delta_{\rm b}$ and $\delta_{\rm m}$ at the time of photon decoupling.

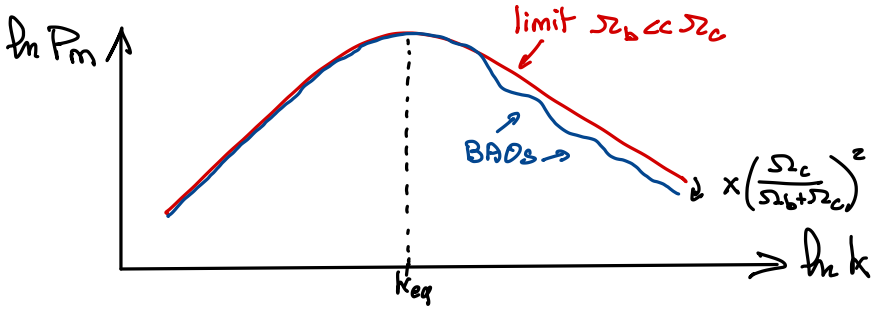

Figure 6.5: Impact of baryons on the shape of the matter power spectrum $P_{\rm m}(\eta_0, k)$.

weighted sum

$$\delta_{\rm m}(\eta_{\rm dec}, k) = \frac{\Omega_{\rm c}}{\Omega_{\rm c}+\Omega_{\rm b}}\,\delta_{\rm c}(\eta_{\rm dec}, k) + \frac{\Omega_{\rm b}}{\Omega_{\rm c}+\Omega_{\rm b}}\,\delta_{\rm b}(\eta_{\rm dec}, k)\ , \tag{6.23}$$

with $\delta_{\rm c}(\eta_{\rm dec}, k) \neq \delta_{\rm b}(\eta_{\rm dec}, k)$. At decoupling, the transfer function $\delta_{\rm c}(\eta_{\rm dec}, k)$ and $\delta_{\rm b}(\eta_{\rm dec}, k)$ are equal to each other on super-Hubble scales $k < \frac{1}{\eta_{\rm dec}}$ as predicted by adiabatic initial conditions. However, on sub-Hubble scales, $\delta_{\rm c}(\eta_{\rm dec}, k)$ has the shape studied in the previous subsection, with a quadratic increase $\delta_{\rm m} \propto k^2$ for $\frac{1}{\eta_{\rm dec}} < k < k_{\rm eq}$, and a logarithmic increase $\delta_{\rm m} \propto \ln(k\eta_{\rm eq})$ for $k > k_{\rm eq}$. Instead, the baryon fluctuations exhibit the same damped oscillations as the photons. This behaviour is summarised in figure 6.4. The middle curve in the plot gives a hint of the shape of the weighted sum $\delta_{\rm m}(\eta_{\rm dec}, k)$.

At the level of the power spectrum, the impact of baryons can be summarised by the fact that the initial condition for $\delta_{\rm m}$ at $\eta = \eta_{\rm dec}$ is no longer given by the upper curve but by the middle curve in figure 6.4. Thus, all what we said on the power spectrum remains true, apart from a reduction of power with superimposed oscillations for $k > k_{\rm eq}$. This is illustrated in figure 6.5. At large k, the oscillations become gradually negligible and the matter power spectrum is just multiplied by the scale-independent factor $\left(\frac{\Omega_{\rm c}}{\Omega_{\rm b}+\Omega_{\rm c}}\right)^2$.

The oscillations are a signature of acoustic oscillations in the early universe (just like the peaks in the CMB spectrum). Given our discussion on acoustic oscillations in the CMB chapter, in first approximation, these oscillation should originate from a factor $\cos[kr_s(\eta_{\rm dec})] \sim \cos(kc_s\eta_{\rm dec})$ in the transfer function of

baryons, where $r_s(\eta_{\rm dec})$ is the comoving sound horizon at decoupling. They are called the baryon acoustic oscillations (BAOs). Like in the CMB, the scale of the first oscillation is set by $k_{\rm BAO} = 1/r_s(\eta_{\rm dec})$, and the other peaks, standing for the harmonics of the first one, are regularly spaced. The intuitive understanding of the sound horizon playing the role of a correlation length seen under a given angle in figure 5.8 applies as well to BAOs as to CMB peaks, but instead of seeing this correlation on the last scattering surface, we see it in the neighbouring distribution of matter, that is, on spheres with a radius much smaller than $(\eta_0 - \eta_{\rm dec})$.

We can now complete our discussion concerning the information on cosmological parameters contained in the linear matter power spectrum. We still assume a minimal ΛCDM model with parameters $\{\omega_{\rm b}, \omega_{\rm m}, \Omega_\Lambda, A_s, n_s\}$. We see that:

- The overall amplitude depends on $\{A_s, \omega_{\rm m}, \Omega_\Lambda\}$ (see previous subsection).
- The overall slope depends on the tilt n_s.
- The location of the broad peak depends on $k_{\rm eq}$, that is, on $\omega_{\rm m}$.
- The step-like suppression at $k \gg k_{\rm eq}$ depends on $\frac{\Omega_{\rm c}}{\Omega_{\rm b}+\Omega_{\rm c}} = 1 - \frac{\omega_b}{\omega_m}$.
- The scale of the BAOs depends on the sound horizon, and thus, like for the CMB, on $\{\omega_{\rm b}, \omega_{\rm m}\}$.

Thus, the measurement of the matter power spectrum contains a lot of information complementary to the one in the CMB spectrum. This is even more true when one considers models beyond ΛCDM, with a possible signature from neutrino masses or non-minimal properties of dark matter or dark energy. The matter power spectrum has the additional advantage that the measurement is performed in a 3D sphere rather than a 2D surface, so one can probe many independent Fourier modes and get a very small theoretical error compared to the cosmic variance of the CMB.

We will see that the full modelling of the matter power spectrum –including non-linear corrections– is tricky. Thus, observers have developed a technique to extract information from the measurement of the scale of the BAOs without actually fitting the full shape of $P_{\rm m}(z,k)$ to the data. Overall, this technique relies on the following steps:

- take a huge galaxy map of galaxies with billions of objects,
- divide it in shells containing all objects with a redshift $z_i - \epsilon < z < z_i + \epsilon$,
- using a smoothing process, build a 2D density map in each shell, $\delta_{{\rm m},i}(\hat{n})$, where $\hat{n}$ is a unit vector standing for two angles on a sphere,
- compute the 2-point correlation function $\xi(\theta) = \langle \delta_{{\rm m},i}(\hat{n}) \delta_{{\rm m},i}(\hat{n}') \rangle$ such that $\cos\theta = \hat{n} \cdot \hat{n}'$.

Given that the matter power spectrum indicates an excess of power on the comoving wavelength $k_{\rm BAO} = 1/r_s(\eta_{\rm dec})$, the 2-point correlation function $\xi(\theta)$ should feature an excess for the angle $\theta_{s,i}$ under which we see this comoving scale, namely,

$$\theta_{s,i} = \frac{d_s(z_i)}{d_a(z_i)} = \frac{a(z_i)\, r_s(\eta_{\rm dec})}{a(z_i)\, r_a(z_i)} \ , \tag{6.24}$$

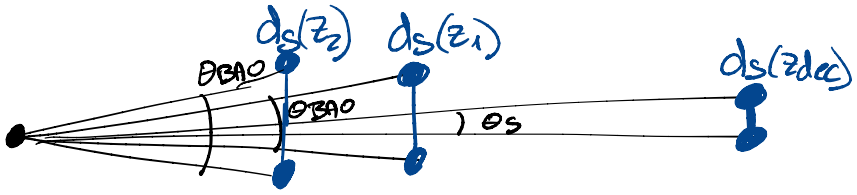

Figure 6.6: We can measure the angle under which we see the sound horizon at many different redshifts using the BAO technique in complement to the CMB spectrum measurement.

where $r_a(z_i) = \int_\eta^{\eta_0} d\tilde{\eta} = \int_z^{z_i} \frac{d\tilde{z}}{a(z)H(z)}$ is the comoving angular diameter distance. Finally, we remember that $r_s(\eta_{\rm dec}) = \int_0^{\eta_{\rm dec}} c_s(\tilde{\eta})\, d\tilde{\eta}$ can be theoretically predicted for each given value of the parameters $\{\omega_{\rm b}, \omega_{\rm m}\}$, and that these parameters can be inferred from the CMB spectrum or other techniques (e.g. $\omega_{\rm b}$ can be inferred from BBN and primordial elements). Thus, each time that we measure an angle $\theta_{s,i}$ we can infer the comoving angular diameter distance to the shell,

$$r_a(z_i) = \frac{r_s(\eta_{\rm dec})|_{\rm CMB}}{\theta_{s,i}|_{\text{galaxy map in shell } i}} \,. \tag{6.25}$$

By placing the measurements in a diagram of r_a versus z_i, we build a Hubble diagram, like when we plot the luminosity distance of supernovae versus redshift.

In summary, on the one hand, using cepheids and supernovae as standard candles, we can reconstruct the function $d_L(z)$ at all redshifts where we see such candles. On the other hand, using the sound horizon (inferred from the CMB spectra or another technique) as a standard ruler, we can reconstruct the function $r_a(z)$ at all redshifts where we see BAOs, plus the redshift $z_{\rm dec}$ of the CMB (see figure 6.6).

This technique has been widely used. Constraints from BAOs are crucial to determine parameters like e.g. the curvature scale Ω_k, because it removes parameter degeneracies from the measurement of the CMB spectrum only. So far, the constraints on cosmological parameter inferred from BAOs have always been very consistent with those inferred from the CMB spectra.

6.2 Observing the matter power spectrum

Mainly two big categories of observations:

- **Galaxy redshift surveys.** Measures the redshift and two angles for each galaxy. Reconstruction of a smoothed map. Its power spectrum in each redshift bin is the galaxy power spectrum $P_{\rm g}(k, z)$. It can be related to the matter power spectrum $P_{\rm m}(k, t)$ after modelling three non-trivial effects: non-linear clustering on the smallest scales (with coupling between different Fourier modes); light-to-mass biais (independent of k on linear scales, k-dependent on non-linear scales); redshift-space distorsions (related to peculiar velocities, which are correlated with density fluctuations).

- **Weak lensing surveys.** Also called " cosmic shear survey". Search for coherent shearing of galaxy shapes (transforming circular images into elliptical images). Allows to reconstruct the sum of the two metric fluctuations, $\phi + \psi$, projected along each line of sight. One can use tomography:

measure the shearing of galaxies in each z-bin, reconstruct a few 2D maps of $\phi(\hat{n}, z) + \psi(\hat{n}, z)$ probing different redhsift ranges, use of the Poisson equation to infer 2D maps of the total matter fluctuation $\delta_{\rm tot.} \simeq \delta_{\rm m}$ in these ranges, and finally infer the matter power spectrum $P_{\rm m}(k, t)$. One still needs to model non-linear clustering on the smallest scales. There is no issue with bias and redshift-space distorsions for these surveys. However, another effect needs to be modelled: intrinsic alignments (from tidal forces).

All in all, these techniques allow to measure $P_{\rm m}(k, t)$ and thus all the parameters of the standard cosmological model excepted $\tau_{\rm reio}$. However, theoretical predictions are not as easy and accurate as in the CMB case due to the effects mentioned above:

- non-linear clustering,
- for galaxy redshift surveys: light-to-mass bias,
- for galaxy redshift surveys: redshift space distorsions,
- for weak lensing surveys: intrinsic alignment of galaxies,
- other effects on very small scale: astrophysical sources like supernovae explosions can alter the clustering of matter on such scales through the effect of winds, shock waves, etc. (generic name: "baryonic feedback").

N-body simulations allow to estimate all these effects, but they are slow and not always very accurate. On mildly non-linear scales, the first four effects can be modelled analytically. This is part of a vast theoretical framework called the "Theory of Large Scale Structure".

Chapter 7

Summary and conclusions

Until 2015: amazing (and even unexpected) level of consensus on the standard cosmological model, thanks to consistent fits to CMB, BAO, supernovae, BBN and light element abundances, weak lensing surveys.

Still some deep questions on the nature of DM and DE. For DM: too many models, no direct detection. For DE: a pure Λ appears as fine-tuned; alternative with dynamical DE are usually also fine-tuned and "ad hoc", with a lack of opportunities to test models independently from cosmological observations.

Since 2015, problems seem to accumulate, although it could be the case that everything is solved in the future with a better modelling of systematics:

- nearby calibrated standard candles (distance ladder) measure H_0 directly and independently from the value inferred indirectly from fits to CMB or BAOs. increasing tension since 10 years (Hubble tension), now approaching 7σ.

- weak lensing survey are giving a power spectrum amplitude a bit lower than that inferred indirectly from fits to the CMB, but this seems to be going away with the most recent data.

- BAO start to prefer a different "late expansion history" compared to CMB, maybe even with dynamical DE, but this is only a 3σ tension.

- there are even a few other minor tensions between some data sets and the standard model.

In the future, these tensions may disappear and the standard cosmological model may get even better confirmed. If this is not the case, we may need to find a slightly different standard model with additional physical ingredients. But all the simplest ones have been investigated and do not provide better fits compared to the standard model. This applies to models with:

- spatial curvature;

- a large GW background from inflation;

- a neutrino mass summed over the three species (the three mass eigenstates) significantly greater than the minimum value compatible with laboratory bounds ($\sum m_\nu > 0.06$ eV);

- dark matter that is not fully cold (because of some large velocity dispersion), fully stable (thus, decaying) or fully decoupled (thus, feebly interacting);
- additional relics from a dark sector, contributing to the budget of relativistic species (increasing $N_{\rm eff}$) or acting as an additional component of DM with slightly different properties;
- non-minimal inflation (with several fields, non-adiabatic initial conditions, a non-power-law primordial power spectrum...);
- large-scale magnetic fields of primordial origin;
- deviations from the Friedmann assumption of homogeneity on very large scales.

Finally, models of dynamical dark energy or with departures from Einstein's gravity theory on very large scales are currently mildly favoured due to the recent BAO data, but this may still go away in the near future; and it seems that these models do not help in solving the Hubble tension.

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] *Physical Foundations of Cosmology*, by Viatcheslav Mukhanov, Cambridge Univ. Press, 2005.
[5] *Cosmology*, by Steven Weinberg, Oxford Univ. Press, 2008.
[6] *Neutrino Cosmology*, by Julien Lesgourgues, Gianpiero Mangano, Gennaro Miele, Sergio Pastor, Cambridge Univ. Press, 2013.
[7] *The cosmic microwave background*, by Ruth Durrer, Cambridge Univ. Press, 2020.
[8] *The young universe*, edited by E. Taillet, Wiley-ISTE 2022.