Cosmology II

Martin Kunz, University of Geneva

June 14, 2005

Contents

Chapter 1

Introduction

This lecture on cosmology emphasises the mathematical methods used to model and understand the evolution of the universe on large scales. It consists of two main parts. In the first part (chapters two to five) we study a universe which is perfectly homogeneous and isotropic. We first define the geometric structure describing such a universe and then discuss the resulting solutions of the Einstein equations of General Relativity. We then employ a kinetic model of matter to study the thermal history of the universe in more detail. The first part then ends with a summary of observational evidence for the model.

The second part starts by discussing several short comings of this standard model, and shows how the inflationary paradigm solves them. We then implement a scalar field model of inflation and find that fluctuations are created on all length scales. We follow the (linear) evolution of these fluctuations and compute their power spectrum today. The final chapter discusses the resulting anisotropies in the cosmic microwave background radiation and compares the theoretical results to observations.

We will always (except where explicitly stated) use "natural" units where $k_B = c = \hbar = 1$. As an example, a temperature of 1 K corresponds to about 8.6×10^{-5} eV and the Hubble constant $H_0 = 100h$ km/s/Mpc is also $h(9.78\,\mathrm{Gyr})^{-1}$.

The initial version of these lecture notes was based on a course by Mar Gil-Bastero, and I would like to thank her for her permission to use her notes. Other important sources include lecture notes by Ruth Durrer and Mark Hindmarsh, and the books *The Early Universe* by Kolb & Turner, *Structure formation in the universe* by Padmanabhan, *Cosmological Inflation and Large Scale Structure* by Liddle & Lyth and *Modern Cosmology* by Dodelson.

I would also like to thank David Parkinson for pointing out several mistakes in earlier version of the notes (there are certainly still many left to be found), and Marcus Ruser for writing up both the exercises and solutions that belong to these lectures.

Chapter 2

The geometrical structure

2.1 The metric

Of the four fundamental forces that we know to exist (the weak and strong force, electromagnetism and gravitation), the first two are effectively only relevant on small length scales. If there was a charge asymmetry in the universe, electromagnetism would thus be the dominant influence on the evolution of the universe. But our observations suggest that the positive and negative charges cancel each other very precisely on large scales, and that there is no net asymmetry. As there are no ”negative” gravitational charges, we can expect that the large-scale behaviour of the universe is governed by the weakest of the four, gravitation. It is therefore the theory of gravity, General Relativity, that we need to use as the basic theory. The basic equations (the *Einstein equations*) connect the curvature and the energy content of the universe, and we will spend the first two chapters studying them. The curvature is derived from the metric, which is the fundamental geometric quantity in General Relativity. It determines the local geometric structure (although it does not constrain the global topology) and in this chapter we will try to find a suitable description. From it we can then derive key concepts like the notion of distances, the expansion of the universe and compute quantities like the age of the universe.

In general, the metric is a covariant, symmetric tensor with two indices, $g_{\mu\nu}$. It depends on space-time and has a priori ten different entries. We will often consider the *line element* or infinitesimal 4-distance,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu. \tag{2.1}$$

Of course this is in general far too complicated to consider. We will try to reduce the number of different components to the minimum possible and to isolate the space-time dependence into a small number of factors. Formally this is done within differential geometry by postulating symmetries and then use the corresponding Killing vectors to impose constraints on the metric. Instead, we will just try to guess the correct result. An actual derivation can be found e.g. in ”Gravitation and Cosmology” by Steven Weinberg.

The fundamental, *cosmological principle* says that no point in the universe is preferred. Thus the universe needs to be *isotropic* (i.e. the same in all directions, rotationally symmetric) and *homogeneous* (the same at all points, symmetric under translations).

If we impose these conditions on space-time, we are led to a steady-state universe which looks the same at all times. One can show that it is unique, and that it is the exponentially expanding de Sitter space. We will see in the next chapter that this solution is not consistent with the existence of matter. Thus we can only impose homogeneity and isotropy on the spatial part of the metric. A theorem of differential geometry tells us that the only 3-spaces which fulfil this requirement are those with (spatially) constant curvature $K(t)$. For any value of K the spaces are uniquely determined (up to isometries), and there are three classes: open ($K < 0$), flat ($K = 0$) and closed ($K > 0$) spaces. The line element is then in spherical coordinates (t, R, θ, ϕ).

$$ds^2 = dt^2 - \left(\frac{dR^2}{1 - KR^2} + R^2 d\Omega\right). \tag{2.2}$$

Both R and K are functions of time (but not of space because of the homogeneity condition). $d\Omega = d\theta^2 + \sin^2\theta d\phi^2$ is the two-dimensional angular volume element. We see as well that the metric is reduced to the Minkowski case with signature $(+, -, -, -)$ if $K = 0$ and if R is independent of time. t is the physical (*proper*) time and R the physical distance. We can rewrite the metric by extracting the time dependence of R into a *scale factor* $a(t)$ by introducing the *comoving distance* r, $dR(t) \equiv a(t)dr$. Then we are left with

$$ds^2 = dt^2 - a(t)^2 \left(\frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega\right). \tag{2.3}$$

If we set up two observers without relative motion at a comoving distance r_0, then they will remain at this distance (hence the name), but their physical distance R will change over time. The curvature parameter κ is related to K by $K = a^2\kappa$. Conventionally, the scale factor is normalised so that κ takes the values $-1, 0$ or 1. Another common normalisation, especially for $\kappa = 0$, is to set the scale factor today to unity, $a(t_0) = a_0 = 1$.

It is possible and often useful to rewrite the metric (2.3) by transforming $r \longrightarrow S_\kappa(r)$ where

$$S_\kappa(r) = \begin{cases} \sin r & \kappa = +1 \\ r & \kappa = 0 \\ \sinh r & \kappa = -1 \end{cases} \tag{2.4}$$

The resulting metric (with the comoving distance called again r) is

$$ds^2 = dt^2 - a^2(t)\left(dr^2 + S_\kappa(r)^2 d\Omega\right). \tag{2.5}$$

The two expressions (2.3) and (2.5) are equivalent and we can use whichever is simpler in a given situation.

We can also introduce the time analogue to the comoving distance, called *conformal time* η and defined by $dt \equiv a(t)d\eta$. Written in these coordinates, the line element (2.3) is

$$ds^2 = a(\eta)^2 \left(d\eta^2 - \left\{\frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega\right\}\right) \tag{2.6}$$

and of course the equivalent expression for (2.5) exists as well. The form given above shows that the metric is conformally related to the Minkowski metric for $\kappa = 0$, i.e. that there is a transformation $g_{\mu\nu} = f\eta_{\mu\nu}$ where f is an arbitrary multiplicative factor (the scale factor in our case). Indeed, it is easy to show that the metric is conformally flat in general.

This metric fully describes the (local) geometry of the Friedmann-Lemaître-Robertson-Walker (FLRW or FRW) model of the universe[1]. We will spend the rest of this first chapter to study some of the properties of the FRW metric.

Remark 1 *The Einstein equations of general relativity link the geometry (the curvature) to the energy (matter / radiation) in the universe. We have so far only considered the former, and we will have to check in the next chapter if we can at all formulate a self-consistent theory!*

2.2 The curvature

The spatial (three-dimensional) curvature is given by ${}^3R = 6K = 6\kappa/a(t)^2$. By construction the curvature is constant on the slices of equal time. Given the three different possibilities for κ we distinguish the three following cases:

1. **closed universe:** The spatial sections are spheres S^3 and ${}^3R > 0$, $\kappa = +1$.
2. **flat (critical) universe:** The spatial sections are flat and ${}^3R = \kappa = 0$.
3. **open universe:** The spatial sections correspond to the hyperbolic space H^3 with ${}^3R < 0$ and $\kappa = -1$.

As an illustration, we study three two-dimensional examples:

1. flat space: We can use as coordinates are either x and y with $dl^2 = dx^2 + dy^2$ or else spherical coordinates r and θ with $dl^2 = dr^2 + r^2 d\theta^2$. The two are related by the coordinate transform $x = r\cos\theta$, $y = r\sin\theta$.

2. closed space: Using the surface of a sphere, we set $R^2 = x^2 + y^2 + z^2$ and so $z = \sqrt{R^2 - r^2}$ where $r^2 = x^2 + y^2$. The line element then becomes

$$dl^2 = dx^2 + dy^2 + dz^2 = dr^2 + r^2 d\theta^2 + \frac{r^2}{R^2 - r^2} dr^2 = \frac{R^2}{R^2 - r^2} dr^2 + r^2 d\theta^2 = R^2 \left(\frac{dr'^2}{1 - r'^2} + r'^2 d\theta^2 \right) \tag{2.7}$$

with $r' = r/R$.

3. open space: We replace in the previous expression $R \to iR$ and end up with

$$dl^2 = R^2 \left(\frac{dr'^2}{1 + r'^2} + r'^2 d\theta^2 \right) \tag{2.8}$$

2.3 The cosmological redshift

Let us look at two objects at a fixed comoving distance d, e.g. two galaxies. As their separation does not change in comoving coordinates, we find that (using that light moves on lines with $ds^2 = 0$) it is the same

[1] Although we will not discuss this in detail, it is important not to forget that the topology is not fixed by general relativity. For example, the flat ($\kappa = 0$) universe is not necessarily infinite but may be in the shape of a finite, flat torus.

at different times,

$$d = \int_{\eta_1}^{\eta_0} d\eta = \int_{\eta_1+\delta\eta_1}^{\eta_0+\delta\eta_0} d\eta. \tag{2.9}$$

Expressed in proper time,

$$d = \int_{t_1}^{t_0} dt/a(t) = \int_{t_1+\delta t_1}^{t_0+\delta t_0} dt/a(t), \tag{2.10}$$

and so $\delta t_0/a(t_0) = \delta t_1/a(t_1)$. The local physical clocks, which also measure the atomic processes through which the light are emitted and so its frequency ν , run thus with the ratio

$$\frac{\delta t_1}{\delta t_0} = \frac{\nu_0}{\nu_1} = \frac{a(t_1)}{a(t_0)} = \frac{\lambda_1}{\lambda_0}. \tag{2.11}$$

Defining the *redshift* as $z \equiv (\lambda_0 - \lambda_1)/\lambda_1$ we have that

$$1 + z = \frac{a(t_0)}{a(t_1)} \tag{2.12}$$

Of course z is only a redshift as long as $a(t_0) > a(t_1)$, i.e. while the universe is expanding. In the opposite case of a collapsing universe it turns into a blueshift.

Although it is tempting to explain the cosmological redshift as a Doppler shift (especially given the Hubble law, see the next section), it is not really correct. Firstly, at very large distances or during periods of rapid expansion (inflation) we would find objects that move away from us faster than the speed of light. Secondly, in General Relativity the presence of matter introduces a gravitational redshift which cannot be separated out. So at best the interpretation as a "recession velocity" can be done in the case of an empty universe and for $z \approx v/c << 1$.

2.4 The Hubble law and the acceleration parameter

Let us consider two observers at fixed comoving distance r_0. They will observe a relative velocity $v_r = d/dtR(t_0) = \dot{a}r_0 = \dot{a}/aR(t_0) \equiv H_0 R$. Thus the relative velocity is proportional to the distance *independently of the position of the observer!* If we lay down a uniform grid in space, then an observer will see all grid points around him recede or move towards him with a velocity proportional to the distance of the grid points. Although an observer might from this conclude that she is in the centre of the universe, this is not true; homogeneity is fully respected. The proportionality factor $H = \dot{a}/a$ is called the *Hubble parameter*, and its value today is denoted by H_0. We will return later to this phenomenon, as the existence of such an uniform expansion is a key observational test of the model.

Because of the Hubble law and the ease with which the frequency shift can be measured, distances are often given in redshift rather than in physical units. One should not forget however, that galaxies or other astrophysical objects generally have peculiar velocities which lead to an additional redshift and cannot always be neglected.

For object at higher redshifts, the universe can expand non-negligibly during the time that the photons need to reach us. In this case, the Hubble constant does not remain constant. Observers often use an additional quantity, the acceleration parameter denoted as q_0, defined by an expansion of the scale factor:

$$(1+z)^{-1} = \frac{a(t)}{a_0} = 1 + H_0(t-t_0) - \frac{1}{2}q_0 H_0^2 (t-t_0)^2 + \dots \tag{2.13}$$

where $q_0 = -\ddot{a}_0/(a_0 H_0^2)$. If $q_0 < 0$ then $\ddot{a}_0 > 0$ and the expansion is accelerating, otherwise it is decelerating.

2.5 Cosmological distances and redshift-distance relation

In order to test the Hubble law, we need to measure distances. There are two basic ways in which this is done: we can either consider the apparent angular size of an object and compare it to its known diameter, or we can measure the apparent luminosity and compare to the known real luminosity of an object.

2.5.1 The metric distance

The spatial volume element of our metric (2.3) today (at time t_0) is $dV = a_0^2 S_\kappa(r)^2 d\Omega dr$. The distance appearing in this expression is sometimes called the metric distance, $d_m(r) = a_0 S_\kappa(r)$. We only need it to define the two following distances.

2.5.2 The luminosity distance

We assume that we know the intrinsic, absolute luminosity L_0 of a certain object, called a *standard candle*. Type Ia supernovae are believed to be approximate standard candles, and they are used to map out cosmological distances. Let us assume that we observe a flux F from the standard candle with absolute luminosity L at a fixed comoving distance r. The definition of the luminosity distance is then

$$F \equiv \frac{L}{4\pi d_L^2} \tag{2.14}$$

The physical surface through which the total flux flows today is $4\pi d_m^2$. But there are several further effects which affect the light:

1. The photons are redshifted $\Rightarrow$ the energy received decreases by $1/(1+z)$.
2. They arrive at a lower rate $\Rightarrow$ the energy received decreases by $1/(1+z)$.
3. The frequency bands are compressed $\Rightarrow$ the energy received increases by $(1+z)$, compensating one of the previous factors, and
4. the photon frequency is shifted.

The flux in the band around frequency ν_0 is then

$$F_\nu(\nu_0) = \frac{L_\nu[(1+z)\nu_0]}{4\pi(1+z)d_m^2}. \tag{2.15}$$

We recover the total flux by integrating over ν. The variable substitution $x = (1+z)\nu$ adds another redshift factor, and the total result turns out to be

$$F = \frac{L}{4\pi d_m^2(1+z)^2}. \tag{2.16}$$

and by comparing with eq. (2.14)

$$d_L = (1+z)d_m. \tag{2.17}$$

2.5.3 The angular diameter distance

Let us now discuss the case where we observe an object with known (physical!) diameter D under an angle δ. We assume again that the object is at a comoving distance r and the photons which we observe today were emitted at time t_1. From this we define the *angular diameter distance* via the relation $D = \delta d_A$. Looking at the configuration in comoving coordinates *today*, we find that $D = a(t_1)S_\kappa(r) = a_0 S_\kappa(r)a(t_1)/a_0\delta = d_m(r)/(1+z)\delta$, as the astrophysical object is held together by forces stronger than gravity and so does not expand with the universe. Eliminating D, the angular diameter distance is

$$d_A = \frac{a_1}{a_0}d_m = \frac{1}{1+z}d_m. \tag{2.18}$$

Comparing the two final expressions, we find that the two distances are not independent. This *reciprocity relation* between the luminosity distance and the angular diameter distance, $d_L = (1+z)^2 d_A$, is more general than presented here and holds in all metric spaces without torsion.

2.5.4 The redshift-distance relation

As an illustration, we express d_L in terms of the Hubble constant and acceleration parameter. This only makes sense at small distances, and so we take $S_\kappa(r) \approx r$. Some longish manipulations, starting from (2.13) lead to the following relations:

$$z = H_0(t_0 - t) + H_0^2(1 - q_0/2)^2(t_0 - t)^2 + \dots \tag{2.19}$$

$$a_0 r = \int_t^{t_0} dt/a(t) = (t_0 - t) + H_0/2(t_0 - t)^2 + \dots \tag{2.20}$$

Solving the first one for $H_0(t_0-t)$ and inserting it into the second one results in $H_0 a_0 r = z-(1+q_0)/2z^2 + \dots$, and when combined with (2.17) gives

$$H_0 d_L \approx H_0 a_0 r(1+z) \approx z + (1-q_0)/2z^2 + \dots \tag{2.21}$$

Hubbles law holds at low z for the luminosity distance, but at higher redshifts corrections from the acceleration make the distance-redshift deviate from a straight line.

2.6 The age of the universe

It is straightforward to compute the age of the universe using the Hubble parameter:

$$t_0 = \int_0^{t_0} dt = \int_0^{a_0} \frac{da}{\dot{a}} = \int_0^{a_0} \frac{da}{aH(a)} = \int_0^{\infty} \frac{dz}{H(z)(1+z)}. \tag{2.22}$$

To derive a precise number, we need to know the time evolution of the scale factor $a(t)$. This is determined by the Einstein equations and thus by the energy content of the universe, which we will consider in the next chapter.

Chapter 3

The dynamics of the universe

3.1 Introducing matter and energy

As mentioned in the last chapter, we are missing a central part of our theory. The metric that we have been using so far,

$$ds^2 = dt^2 - a(t)^2 \left(\frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega \right) \tag{3.1}$$

contains a free function $a(t)$. This should not surprise us. So far we have just been using differential geometry and the cosmological principle. Now we need to use General Relativity in order to determine this function. To this end, we need to solve the Einstein equations,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G_N T_{\mu\nu}. \tag{3.2}$$

$G_{\mu\nu}$ is the Einstein tensor, given by a combination the Ricci tensor $R_{\mu\nu} \equiv R^{\alpha}_{\mu\alpha\nu}$ and the Ricci scalar $R \equiv g^{\mu\nu} R_{\mu\nu}$ which are related to the curvature of space-time, and are determined by the metric:

$$\Gamma^{\alpha}_{\mu\nu} = \frac{1}{2} g^{\alpha\beta} \left(g_{\beta\mu,\nu} + g_{\beta\nu,\mu} - g_{\mu\nu,\beta} \right) \tag{3.3}$$

$$R^{\alpha}_{\beta\mu\nu} = \Gamma^{\alpha}_{\nu\beta,\mu} - \Gamma^{\alpha}_{\mu\beta,\nu} + \Gamma^{\delta}_{\nu\beta} \Gamma^{\alpha}_{\mu\delta} - \Gamma^{\delta}_{\mu\beta} \Gamma^{\alpha}_{\nu\delta}. \tag{3.4}$$

$T_{\mu\nu}$ is the stress-energy tensor or energy-momentum tensor (EM tensor). It is determined by the matter and energy in the universe.

Clearly, if the left hand side of the Einstein equations have to obey the cosmological principle (isotropy and homogeneity), then the same is required for the right hand side. If the matter in the universe is not homogeneous, then we cannot describe the structure of space-time with a metric that is. Thus, in the inertial frame moving with the matter, the energy-momentum tensor needs to have the following form:

$$T^{\mu}_{0} = 0, \quad T^1_1 = T^2_2 = T^3_3, \tag{3.5}$$

and none of the elements can depend on space, only on time. This specific form of the EM tensor arises for ideal or perfect fluids. It is then given by

$$T^{\nu}_{\mu} = \mathrm{diag}\left(\rho(t), -p(t), -p(t), -p(t)\right), \tag{3.6}$$

where $\rho(t)$ is the energy density of the perfect fluid, and $p(t)$ its pressure. In general, if the observer is moving with respect to the rest-frame of the fluid, the EM tensor can be written in explicitly covariant form as

$$T^{\mu\nu} = pg^{\mu\nu} + (\rho + p)U^{\mu}U^{\nu} \tag{3.7}$$

and we recover the previous case when the fluid velocity four-vector is $U^0 = 1$, $U^i = 0$.

The relation between the pressure and the energy density, $p = p(\rho)$ is called the *equation of state*. For (ideal) matter, radiation and vacuum energy it takes a very simple form,

$$p = w\rho, \quad \begin{cases} w = 0 & \text{"pressureless" matter} \\ w = 1/3 & \text{radiation} \\ w = -1 & \text{vacuum energy or cosmological constant} \end{cases} \tag{3.8}$$

For the third case, we have moved the cosmological constant term $\sim \Lambda g_{\mu\nu}$ from the geometric side of the Einstein equation to the matter side and interpreted it as $T^{(\Lambda)}_{\mu\nu}$. Comparing it with eq. (3.6) leads immediately to the equation of state given above.

3.2 Energy conservation

Now we need to solve the Einstein equations. As a first step, we remember that the EM tensor is conserved,

$$T^{\mu\nu}_{\ ;\mu} = 0 = G^{\mu\nu}_{\ ;\mu}. \tag{3.9}$$

The right hand side of this equation follows not only from the Einstein equations, but are also a geometric property in differential geometry and known as *Bianchi identities*. The time-component is

$$T^{\nu}_{0;\nu} = \dot{\rho} + \Gamma^{i}_{i0}(\rho + p) = \dot{\rho} + 3\left(\frac{\dot{a}}{a}\right)(\rho + p) = 0. \tag{3.10}$$

We can rewrite this equation as

$$d(\rho a^3) = -p\, da^3, \tag{3.11}$$

which can be seen as the first law of thermodynamics, as ρa^3 is proportional to the total energy, and a^3 to the volume. We can further write the equation as $d[(\rho + p)a^3] = a^3 dp$, and for the simple equation of state $p = w\rho$ with w constant we find immediately that

$$\rho \propto a(t)^{-3(1+w)} \propto \begin{cases} a(t)^{-3} & \text{for } w = 0 & \text{(matter)} \\ a(t)^{-4} & \text{for } w = 1/3 & \text{(radiation)} \\ \text{const.} & \text{for } w = -1 & \text{(vacuum energy)} \end{cases} \tag{3.12}$$

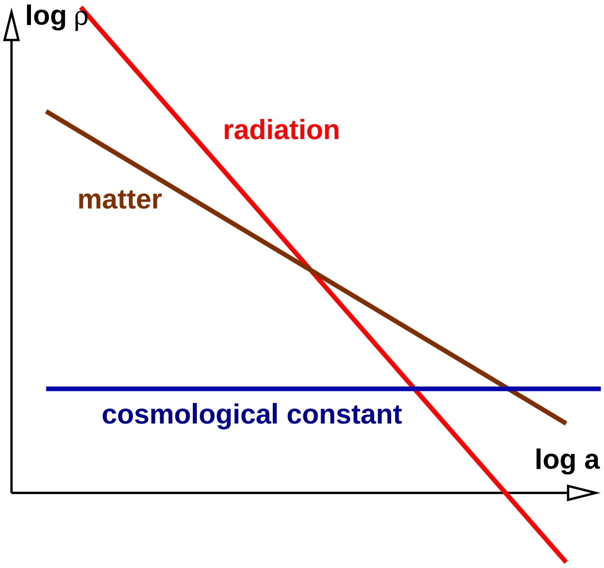

Figure 3.1: Evolution of the energy densities with the scale factor: radiation scales like $1/a^4$ and matter like $1/a^3$, while the cosmological constant has a constant energy density independent of the expansion of the universe.

We can see that the energy density in radiation decreases more rapidly than the one in matter. This can be interpreted as the energy loss through the redshifting of the photons. It also shows that in an expanding universe, radiation will dominate at early times, while matter will become more important later on. If there is a non-vanishing contribution from vacuum energy, it will always start to dominate eventually.

3.3 The Friedmann equations

The non-zero components of the Ricci tensor and scalar are:

$$R_{00} = -3\frac{\ddot{a}}{a} \tag{3.13}$$

$$R_{ij} = -\left[\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{\kappa}{a^2}\right] g_{ij} \tag{3.14}$$

$$R = -6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2}\right] \tag{3.15}$$

The $0-0$ component of the Einstein equations is then

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2} = \frac{8\pi G_N}{3}\rho = \frac{\rho}{3m_p^2}, \tag{3.16}$$

where we introduced the reduced Planck mass, $m_p \equiv 1/(8\pi G_N) \approx 2.410^{-18}\mathrm{GeV}$. The ρ on the right hand side is the sum of all contributions to the energy density in the universe.

The $i - i$ component is

$$2\left(\frac{\ddot{a}}{a}\right) + \left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2} = -8\pi G_N p, \tag{3.17}$$

and if we replace the second and third term on the left hand side by $8\pi G_N \rho/3$ by using the first Friedmann equation, (3.16), we find the second Friedmann equation,

$$\left(\frac{\ddot{a}}{a}\right) = -\frac{4\pi G_N}{3}(\rho + 3p) = -\frac{\rho + 3p}{6m_p^2}. \tag{3.18}$$

The equations (3.11), (3.16) and (3.18) are not independent. Any one of them can be derived from the other two.

3.4 The critical density and the geometry of the universe

Looking at the first Friedmann equation, (3.16), we find that there is a special energy density for which the curvature vanishes. Writing $H(t) \equiv \dot{a}/a$ and using

$$H(t)^2 + \frac{\kappa}{a^2} = \frac{\rho}{3m_p^2} \tag{3.19}$$

we find

$$\frac{\kappa}{a^2 H^2} = \frac{\rho}{3m_p^2 H^2} - 1 \equiv \frac{\rho}{\rho_c} - 1 \equiv \Omega - 1. \tag{3.20}$$

If the total energy density in the universe, ρ, is equal to the *critical* energy density $\rho_c \equiv 3m_p^2 H^2$ then the curvature vanishes and the spatial sections of the universe is flat. Expressed with the *density parameter* $\Omega(t) \equiv \rho/\rho_c$:

$$\begin{aligned}
\Omega(t) > 1 \quad &\Rightarrow \quad \kappa > 0 \Rightarrow \textbf{closed} \text{ universe}\\
\Omega(t) = 1 \quad &\Rightarrow \quad \kappa = 0 \Rightarrow \textbf{flat} \text{ universe}\\
\Omega(t) < 1 \quad &\Rightarrow \quad \kappa < 0 \Rightarrow \textbf{open} \text{ universe}
\end{aligned}$$

We will often split the density parameter into its constituents and write Ω_r for the contribution from radiation, Ω_m for the contribution by matter and Ω_Λ for the vacuum energy contribution. We can also define a "curvature" density parameter, $\Omega_\kappa \equiv -\kappa/(a^2 H^2)$. The Friedmann equation becomes then simply

$$1 = \Omega(t) + \Omega_\kappa(t) = \Omega_r(t) + \Omega_m(t) + \Omega_\Lambda(t) + \Omega_\kappa(t). \tag{3.21}$$

Of course if the universe is flat, i.e. if $\Omega_\kappa = 0$, then $\Omega = 1$ at all times.

There is an additional, useful way to rewrite the Friedmann equation (3.16), by dividing it by H_0^2 and writing out the dependence of the energy density on the scale factor explicitly. This leads to

$$\frac{H^2}{H_0^2} = \frac{\rho_r^{(0)}}{3m_p^2 H_0^2}\left(\frac{a_0}{a}\right)^4 + \frac{\rho_m^{(0)}}{3m_p^2 H_0^2}\left(\frac{a_0}{a}\right)^3 + \frac{\rho_\Lambda^{(0)}}{3m_p^2 H_0^2} - \frac{\kappa}{H_0^2}\frac{1}{a^2}. \tag{3.22}$$

With the critical energy density today given by $3m_p^2 H_0^2 = \rho_c^{(0)}$ and by noting that $\kappa/H_0^2 = -a_0^2 \Omega_\kappa^{(0)}$ we can write

$$H(a)^2 = H_0^2 \left[\Omega_r^{(0)} \left(\frac{a_0}{a}\right)^4 + \Omega_m^{(0)} \left(\frac{a_0}{a}\right)^3 + \Omega_\kappa^{(0)} \left(\frac{a_0}{a}\right)^2 + \Omega_\Lambda^{(0)} \right]. \tag{3.23}$$

This provides an easy way to compute the Hubble constant as a function of the scale factor or the redshift. For the latter, we only need to replace $(a_0/a) \to (1+z)$. We also see immediately that we can only either choose a_0 freely or set $|\kappa| = 1$, as we have the relation $a_0^2 = -\kappa/(H_0^2 \Omega_\kappa^{(0)})$.

The determination of the density parameters today, $\Omega_X^{(0)}$, is one of the main goals of observational cosmology, and we will discuss this in detail in the last chapter. It should be noted that in the literature, the superscript (0) is normally dropped, so that e.g. Ω_m usually denotes the matter density *today* in terms of the critical density *today*. From now on we will follow this notation and drop the (0) superscripts on the density parameters.

3.5 Special cases

In the following sections we use the Friedmann equations to study some special cases and to get a feeling how the universe evolves dynamically.

3.5.1 Acceleration

Using the second Friedmann equation, (3.18), we find that

$$\frac{\ddot{a}}{a} = \frac{-4\pi G_N}{3}(\rho + 3p). \tag{3.24}$$

So if $(\rho + 3p) > 0$, or $w > -1/3$ which is the "normal" case of matter and radiation, then $\ddot{a} < 0$ and the universe is decelerating. In the opposite case, it is accelerating. Specifically, $p/\rho < 0$ is required for acceleration.

The acceleration parameter $q_0 = -\ddot{a}_0/(a_0 H_0^2)$ can also be expressed in terms of the energy momentum tensor,

$$q_0 = \frac{1}{2}\Omega_0(1 + 3w), \tag{3.25}$$

where Ω is the total density parameter, and w the effective equation of state parameter (or it can be written as the sum over all contributions). For the common cases we find

$$\begin{array}{ll} \text{radiation dominated :} & q_0 = \Omega_0 > 0 \\ \text{matter dominated :} & q_0 = \Omega_0/2 > 0 \\ \text{"vacuum" dominated :} & q_0 = -\Omega_0 < 0 \end{array} \tag{3.26}$$

So if we measure $q_0 < 0$ then we need some exotic component with a negative equation of state, $p_X/\rho_X = w_X < -1/3$.

3.5.2 Age of the universe

We derived in the last chapter an expression for the age of the universe, namely

$$t_0 = \int_0^\infty \frac{dz}{H(z)(1+z)}. \tag{3.27}$$

Now we have an explicit expression for $H(z)$. In general, we need to integrate the equation numerically. But let us assume that the universe is matter dominated, i.e. $\Omega_r = \Omega_\Lambda = 0$ and $\Omega_m = \Omega_0$. In this case,

$$H_0 t_0 = \int_0^\infty \frac{dz}{(1+z)^2\sqrt{\Omega_0(1+z) + (1-\Omega_0)}}. \tag{3.28}$$

It is very easy to solve the equation if $\Omega_0 = 1$, where

$$H_0 t_0 = \frac{2}{3}. \tag{3.29}$$

3.5.3 The radiation dominated universe

In the early universe, radiation will eventually dominate over matter, curvature and the cosmological constant, as its energy density scales with the highest negative power of the scale factor, $\rho_r \propto a^{-4}$. For $t \to 0$ we can thus approximate the first Friedmann equation by

$$\left(\frac{\dot a}{a}\right)^2 = H_0^2 \Omega_r \left(\frac{a_0}{a}\right)^4. \tag{3.30}$$

We can rewrite this as $a\dot a = \text{const.}$, or $ada \propto dt$, with the solution

$$a(t) \propto t^{1/2}. \tag{3.31}$$

If we express this in terms of the conformal time $\eta = dt/a$, we find that $a(\eta) \propto \eta$.

3.5.4 The matter dominated universe

Currently the energy density in the radiation is negligible. We only have to take into account the matter, curvature and possibly a cosmological constant. We set $H_0 \Omega_0 a_0^3 = C > 0$ and start by assuming that $\Omega_\Lambda = 0$. The equation to be solved is then

$$\dot a^2 = -\kappa + \frac{C}{a}, \tag{3.32}$$

and we will study the three cases $\kappa = 0, \pm 1$ separately.

1. Euclidian space, $\kappa = 0$: The Friedmann equation is now $\dot a^2 = C/a$ and $a(0) = 0$, which can be solved immediately and gives

$$a(t) \propto t^{2/3}, \tag{3.33}$$

expressed in conformal time we have $a(\eta) \propto \eta^2$. We can also notice that $\dot{a} \to 0$ for $t \to \infty$, so the universe expands forever, but ever more slowly. The age of the universe is easy to compute in this special case, as we computed earlier, we have that $t_0 = 2/(3H_0)$. If we write $H_0 = 100h\text{km/s/Mpc}$ it is $1/H_0 = 9.8 \times 10^9 h^{-1}$ years and so

$$t_0 = \frac{2}{3}\frac{1}{H_0} = 6.5/h\text{Gyr}. \tag{3.34}$$

2. Elliptic model, $\kappa = 1$: The Friedmann equation is now

$$\dot{a}^2 = \frac{C}{a} - 1. \tag{3.35}$$

We see that there is a critical value of the expansion factor, $a_{\max} = C$, where $\dot{a} = 0$ and where the universe changes from expansion to contraction. The universe ends in a "Big Crunch" when $a = 0$ is reached again in finite time. We can analytically solve the equation by setting $a = C \sin^2\theta$, in which case the equation is transformed into

$$2C = \frac{1}{\sin^2\theta}\frac{dt}{d\theta}, \tag{3.36}$$

the solution of which is $t = C(\theta - \cos\theta\sin\theta)$. The "Big Crunch" is reached for $\theta = \pi$ at a time $t = \pi C$. For $\theta \ll 1$ we recover $a \propto t^{3/2}$. As expected, the curvature contribution is subdominant at early times.

3. Hyperbolic model, $\kappa = -1$: In this case we have to solve the equation

$$\dot{a}^2 = \frac{C}{a} + 1. \tag{3.37}$$

In this case, $\dot{a} \geq 0$ at all times, the universe will expand forever. This time we can solve the equation with the substitution $a = C \sinh^2\chi$. The solution is now $t = C(\cosh\chi\sinh\chi - \chi)$, and for $\chi \ll 1$ we recover once more $a \propto t^{2/3}$.

`[figure here]`

3.5.5 Non-zero cosmological constant

If $\Omega_\Lambda = 1$ and $\Omega_r = \Omega_m = \Omega_\kappa = 0$, then the first Friedmann equation becomes

$$\dot{a}^2 = \frac{1}{3}\Lambda a^2 \tag{3.38}$$

which has the solution

$$a(t) = \exp(Ht), \quad H = \sqrt{\frac{\Lambda}{3}}. \tag{3.39}$$

In this model, also known as *steady state* or *de Sitter* model, the universe is completely dominated by a cosmological constant, and the scale factor grows exponentially. Both the acceleration parameter $q = -1$ and the Hubble constant do not change over time. Indeed, all observables are constant, and this universe obeys the perfect cosmological principle of homogeneity and isotropy of space-time.

The model was originally introduced by Fred Hoyle in 1948, but today it is mainly interesting because it may be the end stadium of the evolution of the universe, and especially because the paradigm of *inflation* predicts a phase of (near) exponential expansion in the very early universe. For that reason we will discuss it again in the second part of this lecture.

Meanwhile let's discuss qualitatively the cases where we have also matter and curvature contributions. We introduce again the quantity C, and find the equation

$$\dot{a}^2 = -\kappa + \frac{C}{a} + \frac{\Lambda}{3}a^2 \equiv G(a). \tag{3.40}$$

1. $\kappa = -1$ **:** If $\Lambda > 0$ then $G(a) > 0$ at all times, with a minimum at $a_{\min} = (3C/2\Lambda)^{1/3}$. For small values of a (at early times) eq. (3.40) reduces to the hyperbolic case with vanishing cosmological constant, and $a \propto t^{3/2}$. At late times the cosmological constant dominates and $a \propto \exp(Ht)$. If $\Lambda < 0$ then there exists an a_c with $G(a_c) = 0$. If $a(t=0) = 0$, the universe will expand until a_c and then recollapse again.

2. $\kappa = 0$ **:** This case behaves similar to the previous one, but eq. (3.40) can be solved explicitely. We find

$$a(t)^3 = \frac{3C}{|\Lambda|} \sinh\left(\frac{\sqrt{3\Lambda}}{2}t\right)^2. \tag{3.41}$$

3. $\kappa = 1$ **:** This is a rather complicated case. If we write $P(a) = aG(a)$ then $P(a)$ is a polynomial of degree 3. One of the solutions is real. If Λ is very large, then the scale factor has a minimum in the past. In this case there is no big-bang, and this is called a *bouncing universe*. This case is meanwhile ruled out by observations, as the bounce would happen at a too low redshift. For $\Lambda < 0$, the universe will always recollapse eventually, while for most values of $\Lambda > 0$ it will expand forever, except if C is very large and Λ small, in which case it can also recollapse.

Chapter 4

The thermal history of the universe

Although treating the contents of the universe as perfect fluids allowed us to derive the basic equations governing the expansion of the universe, it is not sufficient to study in detail the behaviour and interactions of realistic particles at high temperature. For this we need to use statistical mechanics.

If the reaction rate is sufficiently high, we can use the well-known equilibrium distributions which we introduce below. In this case, the calculations are straightforward and we only need to understand the evolution of the temperature. This we will do using thermodynamics, where we will see that entropy is conserved. This then allows us to gain a qualitative understanding of many processes in the early universe like the freeze-out of the (light) neutrinos and the decoupling of the photons.

But we cannot treat the detailed mechanism of the freeze-out of particle species in equilibrium. For quantitative results we will need to develop a theory that can handle out-of-equilibrium effects. To this end we will study the *Boltzmann equation*, which will give us a better understanding of the generic freeze-out of (heavy) dark matter particles and of nucleosynthesis.

4.1 Equilibrium distributions: Number densities, energy densities and pressure

For a particle species A (with mass m) in statistical equilibrium, the number density n, energy density ρ and pressure p are given as integrals over the distribution function $f_A(\mathbf{k}, t)$, where $\mathbf{k}$ is the 3-momentum. Different species of particles will be interacting constantly and exchanging energy and momentum. If the rate of these reactions $\Gamma(t) = n\langle\sigma v\rangle$ is much higher than the rate of expansion $H(t)$, then these interactions can produce and maintain thermodynamic equilibrium among the interacting particles with some temperature T. Because in general the interactions have a short range, we may assume that the role of these interactions is just to provide a mechanism for thermalization, and they have nothing to do in deciding the form of the distribution function. Therefore, particles may be treated as an ideal (Bose or Fermi) gas, with the equilibrium distribution function:

$$f_A(\mathbf{k}, t)d^3\mathbf{k} = \frac{g_A}{(2\pi)^3}(\exp[(E_A - \mu_A)/T_A] \pm 1)^{-1} d^3\mathbf{k}\,, \tag{4.1}$$

where g_A is the spin degeneracy factor, μ_A is the chemical potential, T_A is the temperature of this species and $E(\mathbf{k}) = \sqrt{\mathbf{k^2+m^2}}$. The "+" sign corresponds to fermions, and the "-" sign to bosons. If the species is in chemical equilibrium through the reactions $A + B \rightleftharpoons C + D$, then the chemical potentials satisfy: $\mu_A + \mu_B = \mu_C + \mu_D$ (roughly, the chemical potential describes how the free energy changes with the number of particles).

Given the distribution function, we can then calculate the number density n, energy density ρ and pressure p:

$$
\begin{aligned}
n &= \int f(\mathbf{k}) d^3\mathbf{k} = \frac{g}{2\pi^2}\int (\exp[(E-\mu)/T] \pm 1)^{-1} k^2 dk \\
&= \frac{g}{2\pi^2}\int_m^\infty \frac{(E^2-m^2)^{1/2}}{(\exp[(E-\mu)/T] \pm 1)} E dE\,, &(4.2)\\
\rho &= \int E(k) f(\mathbf{k}) d^3\mathbf{k} = \frac{g}{2\pi^2}\int_m^\infty \frac{(E^2-m^2)^{1/2}}{(\exp[(E-\mu)/T] \pm 1)} E^2 dE\,, &(4.3)\\
p &= \int \frac{|k|^2}{3E(k)} f(\mathbf{k}) d^3\mathbf{k} = \frac{g}{2\pi^2}\frac{1}{3}\int_m^\infty \frac{(E^2-m^2)^{3/2}}{(\exp[(E-\mu)/T] \pm 1)} dE\,. &(4.4)
\end{aligned}
$$

We now compute these integrals for two special cases, namely in the relativistic and non-relativistic limits.

(a) Relativistic species: $m \ll T$

In the early universe and for relativistic particles which are created and annihilated continuously, we can in general neglect the chemical potential. We will therefore only consider the case $\mu \ll T$.

$$
n = T^3 \frac{g}{2\pi^2}\int_0^\infty \frac{x^2 dx}{(\exp[x] \pm 1)} \propto T^3\,, \tag{4.5}
$$

where $x = E/T$. Using the Riemann zeta function $\zeta(n) = (1/\Gamma(n)) \int_0^\infty u^{n-1} du/(\exp(u)-1)$, we have for bosons:

$$
n_B = T^3 \frac{g\zeta(3)}{\pi^2}\,, \tag{4.6}
$$

and for fermions, we can use that:

$$
\frac{1}{\exp(x)+1} = \frac{1}{\exp(x)-1} - \frac{2}{\exp(2x)-1}\,, \tag{4.7}
$$

and then:

$$
n_F = T^3 \frac{g\zeta(3)}{\pi^2}(1-\frac{1}{4}) = \frac{3}{4} n_B \propto T^3\,. \tag{4.8}
$$

For the energy density and pressure we get respectively:

$$
\begin{aligned}
\rho_B &= T^4 \frac{g}{2\pi^2}\int_0^\infty \frac{x^3 dx}{(\exp[x] \pm 1)} = T^4 \frac{g}{30}\pi^2\,, &(4.9)\\
\rho_F &= T^4 \frac{g}{30}\pi^2 (1-\frac{1}{8}) = \frac{7}{8}\rho_B\,, &(4.10)
\end{aligned}
$$

and

$$
p = \frac{\rho}{3}\,. \tag{4.11}
$$

There are several points here worthy of note: Firstly, we have derived from statistical mechanics the equation of state for radiation, and indeed we find that $w = p/\rho = 1/3$. Secondly, in eq. (4.10) we derived the Stefan-Boltzmann law, $\rho_\gamma = \sigma_{SB} T^4$. We also showed in the last chapter that the energy density in radiation scales like $\rho_\gamma \propto a^{-4}$. Combined with the above, it follows immediately that the temperature of radiation (and of any relativistic species) scales like

$$T_\gamma \propto 1/a. \tag{4.12}$$

The important consequence is that the universe was much *hotter* when it was smaller. Here we use the temperature of the radiation as the "temperature of the universe", both because it is well defined as the radiation has a thermal spectrum (and thus a unique, well-defined temperature) and because at early times the other particle species interact with the radiation and so share its temperature. We will later discuss what happens when this is no longer true.

(b) Non-relativistic (massive) species, $m \gg T$:

At a given temperature T the main contribution to the (kinetic) energy comes from momenta $k \sim T$. In the non-relativistic case we can therefore expand the expression for the energy, $E = \sqrt{k^2+m^2} = m\sqrt{1+k^2/m^2} \approx m + k^2/(2m)$. Furthermore, the exponential in the denominator is very large, $\exp(m/T) \gg \pm 1$, so that we can drop the ± 1. We then find

$$\begin{aligned} n &\simeq \frac{g}{2\pi^2} T^3 e^{\mu/T} \int_{m/T}^{\infty} e^{-x} (x^2 - m^2/T^2)^{1/2} x dx \\ &= g \left(\frac{mT}{2\pi}\right)^{3/2} e^{-(m-\mu)/T}, && (4.13) \\ \rho &= mn + \frac{3}{2} Tn, && (4.14) \\ p &= nT \ll \rho \quad (p \simeq 0). && (4.15) \end{aligned}$$

We find in the expression for the energy density again the average kinetic energy per particle at a temperature T, $E_{\mathrm{kin}} = \frac{3}{2} k_B T$.

Therefore, in general for relativistic species the number densities go as T^3 and the energy density behaves as T^4, while for massive species they are suppressed by the Boltzmann factor $\exp(-m/T)$. This exponential suppression of the number density means that non-relativistic particles soon drop below the limit where they interact sufficiently often to stay in equilibrium. At this point their interactions will freeze out and they will stop to annihilate. The same happens if there is a different number of particles and anti-particles, where the annihilation will stop when there are no more anti-particles left. In either case, we can invoke an effective chemical potential to enforce the correct number of particles (but we will need the Boltzmann equation to compute $\mu(T)$, or equivalently $n(T)$, in this case).

4.1.1 Several relativistic species: effective number of degrees of freedom

If we have a collection of relativistic species, each of them in equilibrium at different temperatures T_i, we can write the total energy density ρ_R, summing over all the contributions, as:

$$\rho_R = \frac{T_\gamma^4}{2\pi^2} \sum_i g_i \left(\frac{T_i}{T_\gamma}\right)^4 \int_{x_i}^{\infty} \frac{u^3 du}{\exp(u - y_i) \pm 1}, \tag{4.16}$$

where $y_i = \mu_i/T$ and T_γ is the temperature of the photons. Neglecting the chemical potentials ($\mu_i \ll T$), the total energy density in relativistic particles can be written in a compact form as:

$$\rho_R = \frac{T_\gamma^4}{30}\pi^2 g_* \tag{4.17}$$

where g_* is the "effective" number of degrees of freedom, given by:

$$g_* = \sum_{bosons} g_i \left(\frac{T_i}{T_\gamma}\right)^4 + \frac{7}{8}\sum_{fermions} g_i \left(\frac{T_i}{T_\gamma}\right)^4 . \tag{4.18}$$

As the temperature decreases, the effective number of degrees of freedom in radiation will decrease, as massive particles start behaving as non-relativistic when their mass become larger then T:

$T \ll 1\,\mathrm{MeV}$: the only relativistic particles would be the 3 neutrino species (fermions with 2 degrees of freedom each) and the photon (boson, 2 polarisation states). Neutrinos at this temperature are decoupled from the thermal bath and they are slightly colder than the photons, with a temperature $T_\nu = (4/11)^{1/3}T_\gamma$. Therefore:

$$g_* = 2 + \frac{7}{8} \times 3 \times 2 \times \left(\frac{4}{11}\right)^{4/3} \simeq 3.36 . \tag{4.19}$$

$1\,\mathrm{MeV} < T < 100\,\mathrm{MeV}$: Electrons and positrons have a mass of about $0.5\,\mathrm{MeV}$, and so they are now also relativistic. As the difference between neutrino and photon temperature is due to the electron-positron annihilation (see discussion in the section 4.4), we have $T_\nu = T_\gamma$ and so

$$g_* = 2 + \frac{7}{8}(3 \times 2 + 2 \times 2) = 10.75 \tag{4.20}$$

$T > 300$ GeV: this is above the electroweak unification scale, and for the particles in the Standard Model we have $g_* \simeq 106.75$ (exercise!)

When the radiation energy density dominates the expansion[1], the rate of expansion will behave as T^2,

$$H = \frac{\sqrt{\rho_R}}{\sqrt{3}m_P} = 1.66 g_*^{1/2}\frac{T^2}{m_P} , \tag{4.21}$$

and given that $H = 1/(2t)$, then $T \propto t^{-1/2} \propto 1/a$.

4.2 Entropy

We will see later that the neutrinos fall out of thermal equilibrium shortly before the electrons and positrons annihilate. The annihilation energy therefore heats up the photons but not the neutrinos. How can we

[1] Even if matter dominates today, back in time it would have been radiation dominated.

compute the neutrino temperature relative to the (measured) photon temperature after the electron-positron annihilation? We can do it if we find a conserved quantity – in this section we show that entropy is conserved.

Exercise: Show that $\partial p/\partial T = (\rho+p)/T$ for particles in equilibrium (which therefore have the corresponding distribution function) and $\mu = 0$.

According to the second law of thermodynamics, the entropy S of the particles in equilibrium at a temperature T within a volume V satisfies:

$$TdS = d(\rho V) + pdV = d[(\rho + p)V] - Vdp\,, \tag{4.22}$$

Using the result of the exercise, Eq. (4.22) can be recast as:

$$dS = \frac{1}{T}d[(\rho + p)V] - \frac{V}{T^2}(\rho + p)dT = d\left[\frac{V}{T}(\rho + p) + Constant\right]\,. \tag{4.23}$$

Therefore, except for an additive constant, we have in equilibrium:

$$S = \frac{V}{T}(\rho + p)\,. \tag{4.24}$$

We now need to show that the entropy is conserved. To this end we first observe that energy conservation, $\dot{\rho} + 3H(\rho + p) = 0$, leads for a volume $V = a^3$ to

$$\frac{d}{dt}[V(\rho + p)] = V\frac{dp}{dt}, \tag{4.25}$$

and therefore

$$\frac{dS}{dt} = \frac{1}{T}\frac{d[(\rho + p)V]}{dt} - \frac{(\rho + p)V}{T^2}\frac{dT}{dt} = 0, \tag{4.26}$$

where we also used the result of the exercise again.

This result states the conservation of the entropy in a volume $V = a^3$, a fact which which will be essential in computing the temperature of the photons before and after electron-positron annihilation. The entropy density is then defined as:

$$s = \frac{S}{a^3} = \frac{\rho + p}{T}\,. \tag{4.27}$$

For a relativistic species A in thermal equilibrium at a temperature T_A, we have seen that:

$$\rho_A = g_A^{\text{eff}}\frac{\pi^2}{30}T_A^4 = 3p_A\,, \tag{4.28}$$

where $g_a^{\text{eff}} = g_A$ for bosons, and $g_a^{\text{eff}} = 7g_A/8$ for fermions. The total entropy density s, summing over all possible contributions is given by:

$$\begin{aligned} s &= \sum_{A=F,B} s_A(T_A) = \frac{\pi^2}{30}(1 + \frac{1}{3})T_\gamma^3\left[\sum_{A=B} g_A\left(\frac{T_A}{T_\gamma}\right)^3 + \frac{7}{8}\sum_{A=F} g_A\left(\frac{T_A}{T_\gamma}\right)^3\right] \\ &= \frac{2\pi^2}{45}g_{*S}T_\gamma^3\,, \end{aligned} \tag{4.29}$$

where T_γ is the temperature of the photons, and we have defined the effective number of degrees of freedom in entropy:

$$g_{*S} = \sum_{A=B} g_A \left(\frac{T_A}{T_\gamma}\right)^3 + \frac{7}{8} \sum_{A=F} g_A \left(\frac{T_A}{T_\gamma}\right)^3 . \tag{4.30}$$

Notice that g_{*S} is equal to g_* only when all the relativistic species are in equilibrium at the $same$ temperature.

Using the conservation law Eq. (4.22), we have then:

$$d(g_{*S}a^3T^3) = 0 \quad \Rightarrow \quad T \propto g_{*S}^{-1/3} a^{-1} . \tag{4.31}$$

As long as g_{*S}= constant, then the T of the thermal bath decreases as the inverse of the scale factor, $T \propto a^{-1}$, and $s \propto T^3 \propto a^{-3}$. We will show below that we can define an effective temperature for particles that have decoupled, and that this temperature *always* scales like $T_{\text{eff}} \propto 1/a$ because of the redshift of the particle energy. Equivalently, we can say that the entropy of every decoupled particle species is conserved separately.

To any species of particles which is not being created or destroyed, $n \propto a^{-3}$, we can assign a conserved number $Y \propto na^3$. Since $s \propto a^{-3}$, we can define this number to be $Y \equiv n/s$. It follows that, in thermal equilibrium for a relativistic species A ($m \ll T$, $\mu \ll T$):

$$Y_A = \frac{45}{2\pi^4}\zeta(3)\frac{g_A}{g_{*S}} , \tag{4.32}$$

whereas for a non-relativistic species ($m \gg T$), we have:

$$Y_A = \frac{45}{4\sqrt{2\pi^7}}\frac{g_A}{g_{*S}} \left(\frac{m}{T}\right)^{3/2} e^{-(m-\mu)/T} . \tag{4.33}$$

Here T is generally dominated by the relativistic species, as they contribute by far the biggest part to the entropy. It is then the effective chemical potential $\mu(T)$ of the non-relativistic particles which has to ensure that Y_A remains constant and which fixes the final abundance of those particles.

4.3 Decoupling

At any time, the Universe will contain a blackbody distribution of photons with some temperature T_γ. If a species interacts (directly or indirectly) with the photons with a rate $\Gamma_{A\gamma}$ high enough ($\Gamma_{A\gamma} \gg H$), then these particles will have the same temperature as the photons: $T_A = T_\gamma$. Any set of particles species which are interacting among themselves at a high enough rate will share the same temperature. But in general this may not be the case and the Universe could be populated by different species each with its own temperature. As a rule of thumb, a species A will maintain an equilibrium distribution when $\Gamma_A \gg H$, and it will decouple from the thermal bath when the interaction rate drops below the rate of expansion, $\Gamma_A \ll H$.

The rate of interaction can be expressed as:

$$\Gamma_A \equiv n_T \langle \sigma_X v \rangle , \tag{4.34}$$

where n_T is the number density of the target particles, v is the relative velocity, and σ_X is the interaction cross section. $\langle \sigma_X v \rangle$ denotes an average value of this combination (σ_X usually depends on the energy). We should note that the interaction rates between particles A and B are not symmetric, e.g. if $n_A \gg n_B$ then also $\Gamma_B \gg \Gamma_A$. It is then possible that particle B is still in thermal equilibrium with A, while A has already decoupled from B.

As an example, let us look at the self-interaction of relativistic particles A, with $n_A \propto T^3$ and an energy and momentum $E_A \simeq p_A \sim T^4$. If the interaction with coupling constant α_X is mediated by a relativistic particle (either massless, or with a mass $m_X < T$), then $\sigma_X \propto \alpha_X^2 T^{-2}$, and we have:

$$\Gamma_A = n_A(T)\langle\sigma_X v\rangle_T \sim \alpha_X^2 T\,. \tag{4.35}$$

In a radiation dominated Universe, $H \propto T^2/m_P$, and then $\Gamma_A > H$ when $T < \alpha_X^2 m_P$. Hence, we do not expect the particles to be in equilibrium beyond the Planck scale, but they will be equilibrated at a lower temperature.

On the other hand, if the interactions are mediated by a non-relativistic particle ($m_X > T$), we have $\sigma_X \propto \alpha_X^2 T^2/m_X^4$, and

$$\Gamma_A = n_A(T)\langle\sigma_X v\rangle_T \sim \alpha_X^2 \frac{T^5}{m_X^4}\,. \tag{4.36}$$

In this case (a typical example constitutes the weak interaction between the neutrinos), the particles will fall out of equilibrium when the temperature falls below $T \sim (m_X^4/(m_P\alpha_X^2))^{1/3}$.

As long as $\Gamma_A \gg H$ and the interactions can maintain equilibrium, the distribution function f_A maintain the form of the equilibrium distribution. Once the species A is completely decoupled ($\Gamma_A \ll H(t)$) the particles will be travelling along geodesics. The distribution function is conserved during the free propagation. Consider for example dN (freely propagating) particles in a proper volume dV, with momentum in the range $(k, k + d^3k)$, at some time t. The phase space distribution function $f(x,k,t)$ is defined by $dN = f(x,k,t)dVd^3k$. At a later time $t+\delta t$, the proper volume would have increased by a factor $(a(t+\delta t)/a(t))^3$, while the volume in momentum would be redshifted by $(a(t+\delta t)/a(t))^{-3}$, and then the phase volume dVd^3k does not change during the free propagation. Since the number of particles dN is also conserved, it follows that f is conserved.

This allows us to obtain the function $f_{dec}(k,t)$ after decoupling from the known $f_{eq}(k,t)$ before decoupling. Let us assume that decoupling occurs instantaneously at some time $t = t_D$, when the temperature was T_D and the scale factor a_D. For $t < t_D$, $f_{eq}(k,t)$ is given by Eq. (4.1). At some later time $t < t_D$, and because of the redshift in momentum, all particles with momentum k at time t, had momentum $k(a/a_D)$ at time t_D. Therefore:

$$f_{dec}(k,t) = f_{eq}(k\frac{a}{a_D}, t_D)\,, \quad t > t_D\,. \tag{4.37}$$

If decoupling occurs when the species is ultra-relativistic ($T_D \gg m$), then:

$$f_{dec}(k,t) = \frac{g}{(2\pi)^3}\left[\exp(\frac{ka}{T_D a_D}) \pm 1\right]^{-1}\,. \tag{4.38}$$

This has the same form as the distribution function f_{eq} for a relativistic species with temperature

$$T(t) = T_D\frac{a_D}{a(t)}\,. \tag{4.39}$$

The "temperature" in the distribution function falls strictly as a^{-1}; the entropy of these particles, $S_A = s_A a^3$ is conserved separately. For the species which are still in thermal equilibrium, $T \propto g_{*S}^{-1/3} a^{-1}$ falls more slowly.

The number density of these decoupled particles is given by:

$$n_A = g_A^{\text{eff}} \left(\frac{\zeta(3) T_D^3}{\pi^2} \right) \left(\frac{a_D}{a} \right)^3 . \tag{4.40}$$

This number density will be comparable to the number density of photons at any given time. That is, any such decoupled species will continue to exist in our universe today as a relic background with number density comparable to the number density of photons.

Notice that if a massive particle decouples when it is relativistic $T_D \gg m$, then the distribution function is "frozen" in the form of the distribution function f_{eq} of massless particles. These particles will become non-relativistic when the temperature of the thermal bath drops below their mass, such that their energy will now be $E \simeq m$. The distribution function and number density of the particles will still be given by the frozen-in form corresponding to relativistic particles, but the energy density will be that of non-relativistic particles $\rho \simeq nm$.

If the particles decouple when they are already non-relativistic, then

$$f_{dec}(k) = f_{eq}(k \frac{a}{a_D}, T_D) \simeq \frac{g}{(2\pi)^3} \exp[-\frac{(m-\mu)}{T_D}] \exp[-\frac{k^2}{2mT_D} (\frac{a}{a_D})^2] . \tag{4.41}$$

The distribution function has the same form as that of a non-relativistic Maxwell-Boltzmann gas with a temperature $T = T_D (a_D/a)^2$, and chemical potential $\mu(t) = m + (\mu_D - m)(T/T_D)$.

4.4 The thermal history from neutrino decoupling onwards

4.4.1 Neutrino decoupling

At around $T \simeq 10^{12}\text{K} \simeq O(100)$ MeV, the energy density of the universe is essentially given by that of the relativistic particles $e^{\pm}$, ν, $\bar{\nu}$ and photons. Since they are in equilibrium with the same temperature, the effective number of degrees of freedom is $g_* = 10.75$, and the rate of expansion in this radiation-dominated epoch is given by:

$$H(T) = \frac{\rho_R^{1/2}}{\sqrt{3} m_P} \simeq 5.44 \frac{T^2}{m_P} . \tag{4.42}$$

Neutrinos are kept in equilibrium via weak interaction processes ($\bar{\nu}\nu \leftrightarrow e^+ e^-$,...), with a cross section given by:

$$\sigma_F \simeq G_F^2 E^2 \simeq G_F^2 T^2 , \tag{4.43}$$

where G_F is the Fermi constant ($G_F = \pi \alpha_W / (\sqrt{2} m_W^2) = 1.1664 \times 10^{-5}\,\text{GeV}^{-2}$). The interaction rate per (massless) neutrino is:

$$\Gamma_F = n \langle \sigma_F v \rangle \simeq 1.3 G_F^2 T^5 , \tag{4.44}$$

and

$$\frac{\Gamma_F}{H(T)} \simeq 0.24 T^3 G_F^2 m_P \simeq \left(\frac{T}{1 \text{ MeV}}\right)^3 . \tag{4.45}$$

Therefore neutrinos decouple from the rest of the matter at a temperature around $T_D \simeq 1$ MeV. Below 1 MeV, the neutrino temperature scales as a^{-1}. Shortly after neutrino decoupling, the temperature drops below the mass of the electron ($T < 0.5$ MeV), and the entropy in the electron-positron pairs is transferred to the photons, but not to the neutrinos. We have then:

$$g_*(T_D > T > m_e) = 2 + \frac{7}{8}4 = \frac{11}{2}, \qquad g_*(T < m_e) = 2 . \tag{4.46}$$

The conservation of the entropy $S = g_{*S}(aT)^3$ for the particles which are in equilibrium with radiation, shows that $g_{*S}(T_\gamma a)^3 = g_*(T_\gamma a)^3$ remains constant during expansion. Because g_* decreases after $T < m_e$, the value of $(aT_\gamma)^3$ will be larger after $e\bar{e}$ annihilation than its value before:

$$\frac{(aT_\gamma)^3_{after}}{(aT_\gamma)^3_{before}} = \frac{(g_*)_{before}}{(g_*)_{after}} = \frac{11}{4} . \tag{4.47}$$

Neutrinos do not participate in this process and their entropy is separately conserved. But before $e\bar{e}$ annihilation began, photons and neutrinos had the same temperature. Therefore:

$$(aT_\gamma)_{after} = \left(\frac{11}{4}\right)^{1/3} (aT_\gamma)_{before} = \left(\frac{11}{4}\right)^{1/3} (aT_\nu)_{before} = \left(\frac{11}{4}\right)^{1/3} (aT_\nu)_{after} . \tag{4.48}$$

The temperature of the photons is larger than that of the neutrinos today by a factor $(11/4)^{1/3}$. And because $T_\gamma \neq T_\nu$, today $g_* \neq g_{*S}$, with $g_* \simeq 3.36$ and $g_{*S} \simeq 3.91$.

4.4.2 Matter-radiation equality

The total energy density today is $\rho_0 = 1.88 \times 10^{-29} \Omega_0 h^2 \,\text{g cm}^{-3}$, and the energy density of the thermal bath at the temperature $T_0 \simeq 2.73\,K = 2.35 \times 10^{-4}\,\text{eV}$ is $\rho_R = g_*(\pi^2/30) T_0^4 = 8.09 \times 10^{-34}\,\text{g cm}^{-3}$. Using the fact that $\rho_R/\rho_M \propto a_0/a = 1+z$, it follows that the redshift, temperature and time of equal matter and radiation energy densities are given by:

$$\begin{aligned} 1 + z_{EQ} &= 2.32 \times 10^4 \Omega_0 h^2 , & (4.49)\\ T_{EQ} &= T_0(1 + z_{EQ}) = 5.5 \Omega_0 h^2 \text{eV} , & (4.50)\\ t_{EQ} &= \frac{2}{3} H_0^{-1} \Omega_0^{-1/2} (1 + z_{EQ})^{-3/2} = 1.4 \times 10^3 (\Omega_0 h^2)^{-2} \text{yr} . & (4.51) \end{aligned}$$

4.4.3 Photon decoupling and recombination

Matter and radiation (photon and electrons) were in good thermal contact because of rapid interactions between the photons and electrons, through Compton scattering $e\gamma \leftrightarrow e\gamma$. However, eventually the density of free electrons became too low to maintain thermal contact (electrons and protons combine to form neutral hydrogen), and matter and radiation decoupled. This occurs when $\Gamma_\gamma \simeq H$, with the interaction rate of the photons given by:

$$\Gamma_\gamma = n_e \sigma_T \,, \tag{4.52}$$

where n_e is the number density of free electrons, and $\sigma_T = 8\pi\alpha_{em}^2/(3m_e^2) = 6.65 \times 10^{-25}\,\mathrm{cm}^2$ the Thomson cross section (scattering of classical electromagnetic radiation by a free e^-).

The equilibrium abundance of free electrons is determined by the Saha equation. Let n_H, n_p and n_e denote the number density of hydrogen, free protons and free electrons respectively. We are going to assume that all the baryons in the Universe are in the form of protons. Because of the charge neutrality of the Universe $n_p = n_e$, and baryon conservation gives $n_B = n_p + n_H = n_e + n_H$. In thermal equilibrium, with $T < m_i$, for $i = e, p, H$, we have

$$n_i = g_i \left(\frac{m_i T}{2\pi}\right)^{3/2} \exp[\frac{\mu_i - m_i}{T}] \,. \tag{4.53}$$

Due to the reactions $p+e \to H+\gamma$, the chemical potentials satisfy: $\mu_p + \mu_e = \mu_H$; and $m_H = m_p + m_e - B$, where B is the binding energy of hydrogen $B = 13.6$ eV. Then, in order to remove the dependency on the chemical potentials, n_H can be expressed in terms of n_p and n_e as:

$$n_H = \frac{g_H}{g_p g_e} n_p n_e \left(\frac{m_e T}{2\pi}\right)^{-3/2} \exp[\frac{B}{T}] \,, \tag{4.54}$$

with $g_p = g_e = 2$ and $g_H = 4$. The fractional ionisation is:

$$X_e \equiv \frac{n_p}{n_B} = \frac{n_e}{n_e + n_H} \,, \tag{4.55}$$

It is useful to study $(1 - X_e)/X_e^2 = n_H n_B / n_e^2$, as this depends only on $n_B = \eta n_\gamma$,

$$\frac{1 - X_e^{eq}}{(X_e^{eq})^2} = \frac{4\sqrt{2}\zeta(3)}{\sqrt{\pi}} \frac{n_B}{n_\gamma} \left(\frac{T}{m_e}\right)^{3/2} \exp[\frac{B}{T}] \,. \tag{4.56}$$

This is the Saha equation for the equilibrium ionisation fraction.

The baryon-to-photon ratio is of the order $\eta = n_B/n_\gamma \simeq 10^{-10}$, and the temperature is related to the redshift by $T = (1+z)T_0$. If we define recombination as the point when 90% of the electrons have combined with protons, $X_e^{eq} = 0.1$, from Eq. (4.56) we obtain that this occurs at a redshift, temperature and time given by:

$$\begin{aligned} z_{rec} &\simeq 1300 \,, & (4.57)\\ T_{rec} &\simeq 0.308\,\mathrm{eV} \,, & (4.58)\\ t_{rec} &\simeq 4.4 \times 10^{12} (\Omega_0 h^2)^{-1/2}\, s = 1.4 \times 10^5 (\Omega_0 h^2)^{-1/2}\,\mathrm{yr} \,. & (4.59) \end{aligned}$$

Now, using that the density of free electrons is given by $n_e = X_e n_B = X_e \eta n_\gamma$, and the Thomson cross-section $\sigma_T = 6.65 \times 10^{-25}\,\mathrm{cm}^2$, we can find the photon interaction rate Γ_γ and compare to the Hubble rate for a matter-dominated universe, $H = H_0 \Omega_0 (1+z)^{3/2}$. Decoupling of photons and matter occurs at a redshift, temperature and time given by:

$$\begin{aligned} z_D &\simeq 1200 - 1100 \,, & (4.60)\\ T_D &\simeq 0.26\,\mathrm{eV} \,, & (4.61)\\ t_D &\simeq 5.64 \times 10^{12} (\Omega_) h^2)^{-1/2}\, s = 1.8 \times 10^5 (\Omega_0 h^2)^{-1/2}\,\mathrm{yr} \,. & (4.62) \end{aligned}$$

It should be stressed that three distinct events take place in the universe at around $T \simeq 0.3 - 0.2\,\mathrm{eV}$. First, most of the protons and electrons combine to form hydrogen. Then, the process of recombination stops, leaving a small fraction of free electrons and protons, when the interaction rate for $p + e \leftrightarrow H + \gamma$ drops below H. Finally, the photon mean-free path, given by Γ_γ^{-1} becomes larger than H^{-1}, and photons decouple from matter. This is the origin of the cosmic microwave background radiation (CMB).

Naively we would have expected recombination to take place at the ionisation energy of hydrogen, $T \approx B =$ 13.6 eV. But the large number of photons per baryon ($\eta \ll 1$) means that even at a lower temperature there are enough photons with a higher energy (in the tail of the distribution function) to ionise neutral atoms. We will find a similar effect during nucleosynthesis.

4.5 Non-equilibrium processes: The Boltzmann equation

We have seen qualitatively that, when the interaction rate of a particle species A, Γ_A, is much larger than H, then they are kept in thermal equilibrium; when Γ_A drops below H, they are decoupled. However, we have so far always used the *equilibrium* form of the one-particle distribution function $f(x, p, t)$. A more accurate description of the decoupling process needs to follow the *evolution* of the distribution function. This is given by the Boltzmann equation, which can be written in a very compact form as:

$$\hat{L}[f_A] = \frac{1}{2}\hat{C}_A[f]\,, \tag{4.63}$$

where $\hat{L}$ is the Liouville operator, and $\hat{C}_A$ the collision operator.

4.5.1 The Liouville operator

In the non-relativistic limit the Liouville operator is just the total time derivative. For a particle species of mass m subject to a force F, this corresponds to

$$\hat{L}_{NR} = \frac{d}{dt} = \frac{\partial}{\partial t} + \frac{\partial x}{\partial t}\frac{\partial}{\partial x} + \frac{\partial p}{\partial t}\frac{\partial}{\partial p} = \frac{\partial}{\partial t} + \bar{v}\nabla_x + \frac{F}{m}\nabla_v\,. \tag{4.64}$$

And the covariant, relativistic generalisation of $\hat{L}$ is:

$$\hat{L} = p^\alpha \frac{\partial}{\partial x^\alpha} - \Gamma^\alpha_{\beta\gamma} p^\beta p^\gamma \frac{\partial}{\partial p^\alpha}\,. \tag{4.65}$$

Here $p_\mu = (E, \bar{p})$ and $p_\mu p^\mu = m^2$. In the absence of collisions, $\hat{L}[f] = 0$ just means that particles move on geodesics. Indeed, we have used the geodesic equation,

$$\frac{\partial p^\alpha}{\partial t} + \Gamma^\alpha_{\beta\gamma} p^\beta p^\gamma = 0, \tag{4.66}$$

to derive the covariant form of the Liouville operator.

For the *isotropic* and *homogeneous* FRW model, the distribution function f_A depends only on the modulus of $\bar{p}$ (i.e., the energy), such that $f_A = f_A(E, t)$. So only the $\alpha = 0$ components do not vanish, and we find

$$\hat{L}[f_A] = \left[E\frac{d}{dt} - \frac{\dot{a}}{a}\bar{p}^2\frac{\partial}{\partial E}\right] f_A(E, t)\,. \tag{4.67}$$

In the following we will drop the bar again, $p = |\bar{p}|$. We can write the Boltzmann eq. in terms of the number density

$$n_A = \frac{g_A}{2\pi^2} \int dp p^2 f_A(E,t) \,, \tag{4.68}$$

by integrating Eq. (4.67) over the momentum,

$$\frac{g_A}{2\pi^2} \int dp p^2 \frac{\hat{L}[f_A]}{E} = \frac{dn_A}{dt} - H \frac{g_A}{2\pi^2} \int dp \frac{p^4}{E} \frac{\partial f_A}{\partial E} = \dot{n}_A + 3Hn_A \,. \tag{4.69}$$

In the last step we used integration by parts. Notice that in the absence of interactions, the Boltzmann equation would reduce to:

$$\dot{n}_A + 3Hn_A = 0 \,, \tag{4.70}$$

and this is what we expect in a FRW (isotropic and homogeneous) model: without interactions, the number density of a particle species changes just with the expansion of the universe, $n \propto a^{-3}$.

4.5.2 The Collision operator

The collision operator roughly gives the change in the number of particles A due to interactions. That is, given a space volume d^3pd^3x, it would give the number of particles "in" minus the number "out". For a process of the kind $A + X \leftrightarrow Y$, where X and Y denote collectively any set of other particles, we have that:

$$\begin{aligned} \frac{g_A}{(2\pi)^3} \int d^3p_A \frac{\hat{C}_A[f]}{2E_A} &= -\int d\Pi_A d\Pi_X d\Pi_Y (2\pi)^4 \delta^{(4)}(p_A + p_X - p_Y) \\ &\quad \times \left[|\mathcal{M}_{A+X\to Y}|^2 f_A f_X (1 \pm f_Y) - |\mathcal{M}_{Y\to A+X}|^2 f_Y (1 \pm f_A)(1 \pm f_X) \right] \end{aligned} \tag{4.71}$$

where $d\Pi_i = (g_i/(2\pi)^3) d^3p_i/(2E_i)$. The delta-function ensures energy-momentum conservation. The factors $1 \pm f$ are blocking (fermions, $-$ sign) and stimulated emission (bosons, $+$ sign) factors. $|\mathcal{M}|_{if}$ is the matrix element for the process considered. The latter is related to the differential cross-section by:

$$d\sigma = \frac{1}{2E_A 2E_X |v_A - v_X|} d\Pi_Y |\mathcal{M}_{A+X\to Y}|^2 (2\pi)^4 \delta^{(4)}(p_A + p_X - p_Y) \,. \tag{4.72}$$

We are not going to use the collision operator in its full generality. Instead, we are mainly interested in the details of the decoupling for $(E-\mu)/T \gg 1$, as nothing really happens if the particles decouple when they are still relativistic. We will also assume that the interactions are sufficient to keep the kinetic equilibrium, so that only μ is a free function. I.e. we assume that the particles are completely decoupled when they drop out of kinetic equilibrium (and their distribution function is then frozen). In this way we only need to solve an ordinary differential equation for $\mu(T)$ (or $n(T)$) rather than the full integro-differential Boltzmann equation.

In this scenario we can again neglect the ± 1 in the denominator and the distribution function is $f \propto \exp[-(E-\mu)/T] \ll 1$ and therefore $1 \pm f \approx 1$. We furthermore assume that the interactions are CP conserving,

$$|\mathcal{M}_{Y\to A+X}|^2 = |\mathcal{M}_{A+X\to Y}|^2 = |\mathcal{M}|^2 \,. \tag{4.73}$$

The term inside the square brackets above now is (using energy conservation)

$$[\ldots] \to |\mathcal{M}|^2 e^{-(E_A+E_X)/T}\left(e^{(\mu_A+\mu_X)/T} - e^{\mu_Y/T}\right). \tag{4.74}$$

Writing $n_i^{(eq)}$ for the distribution function in chemical equilibrium ($\mu_i = 0$) we can re-express μ_i with the number densities, $\exp(\mu_i/T) = n_i/n_i^{(eq)}$. The square bracket has now become

$$[\ldots] \to |\mathcal{M}|^2 e^{-(E_A+E_X)/T}\left(\frac{n_A\, n_X}{n_A^{(eq)} n_X^{(eq)}} - \frac{n_Y}{n_Y^{(eq)}}\right) \tag{4.75}$$

The thermal average of the cross section is given by the momentum integral over the matrix element weighted by a thermal factor,

$$\langle\sigma v\rangle = \frac{1}{n_A^{(eq)} n_X^{(eq)}}\int d\Pi_A d\Pi_X d\Pi_Y (2\pi)^4 \delta^{(4)}(p_A + p_X - p_Y)|\mathcal{M}|^2 e^{-(E_A+E_X)/T}, \tag{4.76}$$

and we finally end up with the Boltzmann equation in its desired form,

$$\dot{n}_A + 3Hn_A = -n_A^{(eq)} n_X^{(eq)}\langle\sigma v\rangle\left(\frac{n_A\, n_X}{n_A^{(eq)} n_X^{(eq)}} - \frac{n_Y}{n_Y^{(eq)}}\right). \tag{4.77}$$

Note that if the particles in the final state, Y, are in full equilibrium then we can rewrite the Boltzmann equation as

$$\dot{n}_A + 3Hn_A = \langle\sigma v\rangle\left(n_A^{(eq)} n_X^{(eq)} - n_A\, n_X\right), \tag{4.78}$$

independent of the final state. Also, the left hand side is roughly of the order of $n_A/t \approx n_A H$ while the right hand side is of the order of $n_A n_X\langle\sigma v\rangle$. If interactions are very frequent compared to the expansion rate of the universe, i.e. $H \ll n_X\langle\sigma v\rangle$ then Boltzmann equation requires the terms in brackets on the right to cancel nearly,

$$\frac{n_A\, n_X}{n_A^{(eq)} n_X^{(eq)}} \approx \frac{n_Y}{n_Y^{(eq)}}. \tag{4.79}$$

We have already seen this equation as the Saha equation in recombination, and it will also be used for the nuclear statistical equilibrium when we discuss big-bang nucleosynthesis later on. But first we use our new tool to study the freeze out of heavy particles in more detail.

4.6 Relic abundance of massive stable particles

If the species A is stable, then the dominant process which can change the number of particles in a comoving volume are the annihilation and inverse annihilation
$A + \bar{A} \leftrightarrow Y + \bar{Y}$. We will also assume that there is no asymmetry between particles and antiparticles, and that Y and $\bar{Y}$ are in thermal equilibrium, i. e.,

$$n_A = n_{\bar{A}}, \tag{4.80}$$

$$n_Y = n_{\bar{Y}} = n_Y^{(eq)} \Rightarrow n_Y^{(eq)} n_{\bar{Y}}^{(eq)} = n_A^{(eq)} n_{\bar{A}}^{(eq)}. \tag{4.81}$$

The Boltzmann equation, Eq. (4.77), reduces to:

$$\dot{n}_A + 3Hn_A = -\langle\sigma_A v\rangle[n_A^2 - (n_A^{(eq)})^2]\,. \tag{4.82}$$

We can rewrite this equation as the evolution equation of the number of species $Y_A = n_A/s$, or abundance, in a comoving volume. As in a passively evolving system both $s \propto a^{-3}$ and $n \propto a^{-3}$, we expect the rate of change of Y_A to describe the deviation from collisionless behaviour. Indeed

$$s\dot{Y}_A = \dot{n}_A + 3Hn_A\,. \tag{4.83}$$

And instead of computing the evolution with respect to time, let us introduce the dimensionless variable $x = m/T$. In a radiation dominated universe, we have:

$$H(T) = \frac{1}{2t} = 1.66 g_*^{1/2}\frac{T^2}{m_P} = H(x=1)x^{-2}\,, \tag{4.84}$$

such that:

$$\dot{Y}_A = \frac{H(x=1)}{x}Y_A'\,, \tag{4.85}$$

where "prime" denotes derivative with respect to x. The equilibrium form of Y is given by

$$Y^{(eq)} = 0.278\frac{g}{g_{*S}} \qquad \text{if } x \ll 1 \tag{4.86}$$

$$Y^{(eq)} = 0.145\frac{g}{g_{*S}}x^{3/2}e^{-x} \qquad \text{if } x \gg 1 \tag{4.87}$$

The Boltzmann equations reads now:

$$Y_A' = -\langle\sigma_A v\rangle\frac{xs}{H(x=1)}\left[Y_A^2 - (Y_A^{(eq)})^2\right]\,, \tag{4.88}$$

or:

$$\frac{x}{Y_A^{(eq)}}Y_A' = -\frac{\Gamma_A}{H(x)}\left[\left(\frac{Y_A}{Y_A^{(eq)}}\right)^2 - 1\right]\,. \tag{4.89}$$

Here is our "rule of thumb" in the above equation: Γ/H describes the "efficiency" of the annihilations. When $\Gamma_A \gg H(x)$, interactions are fast enough for the species A to thermalise, and $Y_A \to Y_A^{(eq)}$. When $\Gamma_A \ll H(x)$ the right hand side turns itself off and the relative change in a comoving volume is negligible, $|\Delta Y_A|/Y_A \ll 1$, and the abundance Y_A "freezes in" to the value $Y_A^{(eq)}(x_f)$, where $x_f = m_A/T_f$ and T_f is the "freeze out" temperature.

4.6.1 Hot relics

If the species decouples when still relativistic, say $x_f < 3$, then at that time $Y_A^{(eq)}$ is constant and the final, asymptotic value of Y_A is quite insensitive to the details of freeze-out:

$$Y_A(x\to\infty) = Y_A^{(eq)}(x_f) = 0.278\frac{g_A^{eff}}{g_{*S}(x_f)}\,. \tag{4.90}$$

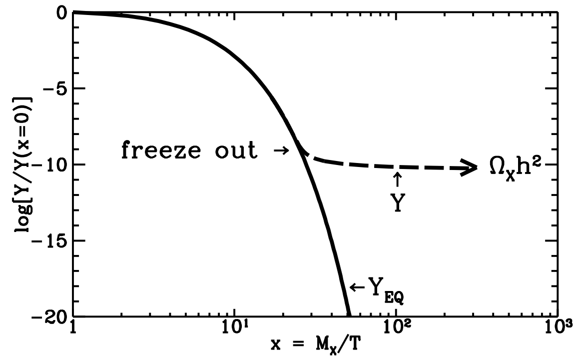

Figure 4.1: $Y(x)$ for massive stable particles. For $T \gg M$ ($x \ll 1$) Y is at its equilibrium value, but decays exponentially as the temperature becomes lower. At freeze-out the particle stops interacting and Y asymptotes to a constant value.

Assuming that there is no entropy production since A decouples, its abundance, number density and relic mass density today are just:

$$Y_{A,0} = Y_\infty\,, \qquad n_{A,0} = s_0 Y_\infty\,, \qquad \Omega_{A,0} = \frac{n_{A,0} m_A}{\rho_c} = \frac{s_0 Y_\infty m_A}{3H_0^2 m_P^2}\,. \tag{4.91}$$

We can then use a limits on Ω_A to place limits on m_A.

4.6.2 Cold relics

If the species decouples when $m \gg T$ ($x_f > 3$), then the abundance in equilibrium is suppressed by the Boltzmann factor $Y_A^{(eq)} \simeq \exp(-m_A/T)$, and solving the Boltzmann equation is more involved (it usually requires numerical integration). But in order to study the general behaviour of the solution to the Boltzmann equation, let us parametrise the temperature dependence of the average cross-section by:

$$\langle \sigma_A v \rangle \equiv \sigma_0 \left(\frac{T}{m_A}\right)^n = \sigma_0 x^{-n}\,, \tag{4.92}$$

such that:

$$\begin{aligned} Y_A' &= -\langle \sigma_A v \rangle \frac{xs}{H(m)} \left[Y_A^2 - (Y_A^{(eq)})^2\right] \\ &= -\frac{\sigma_0 s(m)}{H(m)} x^{-n-2} \left[Y_A^2 - (Y_A^{(eq)})^2\right] = \frac{\lambda}{x^{n+2}} \left[Y_A^2 - (Y_A^{(eq)})^2\right]\,. \end{aligned} \tag{4.93}$$

The equation for $\Delta_A = Y_A - Y_A^{(eq)}$, the departure from equilibrium, is then:

$$\Delta_A' = -\frac{\lambda}{x^{n+2}} \Delta_A (\Delta_A + 2Y_A^{(eq)}) - (Y_A^{(eq)})'\,. \tag{4.94}$$

At early times, Y_A tracks $Y_A^{(eq)}$, such that $\Delta'_A \simeq 0$, and the equation above we have:

$$\Delta \simeq -\frac{(Y_A^{(eq)})'}{\lambda(\Delta_A + 2Y_A^{(eq)})} x^{n+2} \simeq \frac{x^{n+2}}{2\lambda}\,. \tag{4.95}$$

At late times, $Y_A^{(eq)} \to 0$ due to the Boltzmann suppression factor, and then $\Delta_A \simeq Y_A$, so:

$$\Delta'_A \simeq -\frac{\lambda}{x^{n+2}}\Delta_A^2 \Rightarrow Y_{A,\infty} = \Delta_{A,\infty} \simeq \frac{n+1}{\lambda} x_f^{n+1}\,, \tag{4.96}$$

that is:

$$Y_{A,\infty} \simeq \frac{H(m)}{s(m)\sigma_0}(n+1)x_f^{n+1} \simeq C\frac{g_*^{1/2}}{g_{*S}}\frac{(n+1)x_f^{n+1}}{m_P m_A \sigma_0}\,, \tag{4.97}$$

where $C \simeq 3.79$ from the numerical integration. The important point to notice is that the relic abundance is inversely proportional to the annihilation cross-section σ_0, and the mass. The lighter, weakly interacting particles survive with a larger abundance than the heavier, strongly interacting ones. Moreover, the relic density today would be:

$$\rho_A = m_A n_{A,0} = m_A s_0 Y_{A,\infty} \propto \sigma_0^{-1}\,, \tag{4.98}$$

$$\Omega_{A,0} = \frac{s_0 Y_{A,\infty} m_A}{3H_0^2 m_P^2} \propto \sigma_0^{-1}\,, \tag{4.99}$$

and it only depends on the interaction but not on the mass. We find that weakly interacting, stable particles can survive as a cold relics and have the right abundance today to be candidates for "dark matter".

Check, using Eq. (4.99), that indeed if we demand $\Omega_A \sim O(1)$ then $\sigma_0 \sim 10^{-37} cm^2 \sim G_F^2$, i. e., of the order of the cross-section for weak interactions. Some values that can be useful: $T_0 = 2.73$ K, $H_0 = 100h\, kms^{-1}Mpc^{-1}$, $1\,Mpc = 3\times 10^{22} m$, $m_P = 1.2\times 10^{19}\,GeV$, $k_B = 8.62\times 10^{-5}\,eVK^{-1}$.

4.7 Big-Bang Nucleosynthesis

Or how to produce light elements (nuclei) from free protons and neutrons in an expanding Universe. At temperatures larger or of the order of 1 MeV, protons and neutrons are free and they are kept in equilibrium by the weak interactions which are faster than the expansion rate. However, as the temperature drops and weak processes are not able to maintain equilibrium between neutrons and protons, other nuclear species are going to be thermodynamically favoured. That is, neutrons and protons start producing light nuclei: deuterium (D), tritium (T), helium (^{3}He, ^{4}He), ^{7}Be and ^{7}Li, through the following sequence of two-body reactions:

$$\begin{array}{lll}
p(n,\gamma)D & & \\
D(p,\gamma)^3He & D(D,n)^3He & D(D,p)T \\
T(D,n)^4He & T(^4He,\gamma)^7Li & \\
^3He(n,p)T & ^3He(D,p)^4He & ^3He(^4He,\gamma)^7Be \\
^7Li(p,^4He)^4He & ^7Be(n,p)^7Li &
\end{array}$$

In this way BBN (big-bang nucleosynthesis) predicts the *primordial* abundances of these light elements.

Here we only sketch the physics involved in the calculation. The detail calculation of these abundances is done by solving numerically Boltzmann equations taking into account among other things the precise nuclear interactions rates. Our discussion proceeds in three steps. First we will discuss what happens if we assume equilibrium throughout. This will give us an overview and we will learn at what energy BBN starts. In a second step we compute the neutron abundance which determines how many heavier nuclei can be created, and at the end we describe how nucleosynthesis proceeds in reality.

The binding energy of the light nuclei is of the order of a few MeV. But, as with recombination, we expect the low number of baryons compared to the number of photons to delay nucleosynthesis somewhat. We therefore start out with the following particle inventory:

- relativistic particles in equilibrium: electrons, positrons, photons
- relativistic particles that have decoupled: neutrinos
- non-relativistic particles: neutrons and protons

To understand the general way in which nucleosynthesis will progress, we begin by studying the equilibrium case:

4.7.1 Nuclear statistical equilibrium (NSE)

Given a nuclear species, with mass number $A = n + p$ and charge $Z = p$, where $n =$ number of neutrons and p= number of protons in the nucleus, its (kinetic) equilibrium number density is given by:

$$n_A = g_a \left(\frac{m_A T}{2\pi}\right)^{3/2} \exp(\frac{\mu_A - m_A}{T}) . \tag{4.100}$$

If nuclear reactions occur at a rate higher than the expansion rate H, then we obtain chemical equilibrium, and the chemical potential for the nucleus is related to that of protons and neutrons through:

$$\mu_A = Z\mu_p + (A - Z)\mu_n . \tag{4.101}$$

Using Eq. (4.101), we can re-express the number densities in Eq. (4.100) in terms of the number densities of protons and neutrons:

$$n_A = g_A \frac{A^{3/2}}{2^A} \left(\frac{m_N T}{2\pi}\right)^{3(1-A)/2} n_p^Z n_n^{A-Z} exp(\frac{B_A}{T}) , \tag{4.102}$$

where $B_A = Zm_p + (A - Z)m_n - m_A$ is the binding energy of the nucleus, and in the prefactor in Eq. (4.102) we have approximate the nucleon mass $m_N \simeq m_p \simeq m_n \simeq m_A/A$. In order to get rid of the effect of the expansion on the number densities, $n_i \propto a^{-3}$, it is useful to define the mass fraction X_A as the number density of the species per total nucleon density $n_N = n_p + n_n + \sum(An_A)$, normalised such that $\sum_A X_A = 1$,

$$X_A = \frac{n_A A}{n_N}$$

$$
\begin{aligned}
&= g_A \frac{A^{5/2}}{2^A}\left(\frac{m_N T}{2\pi}\right)^{3(1-A)/2} \frac{n_p^Z n_n^{A-Z}}{n_N} \exp(\frac{B_A}{T}) \\
&= g_A \frac{A^{5/2}}{2^A}\left(\frac{m_N T}{2\pi}\right)^{3(1-A)/2} X_p^Z X_n^{A-Z} n_\gamma^{A-1} \eta^{A-1} \exp(\frac{B_A}{T}) \\
&= g_A C A^{5/2}\left(\frac{T}{m_N}\right)^{3(A-1)/2} X_p^Z X_n^{A-Z} \eta^{A-1} \exp(\frac{B_A}{T}), \qquad (4.103)
\end{aligned}
$$

where in the last line we have used $n_\gamma = 2\zeta(3)T^3/\pi^2$, $C = \zeta(3)^{A-1}\pi^{(1-A)/2}2^{(3A-5)/2}$, and

$$\eta \simeq 2.68 \times 10^{-8}\Omega_B h^2 = 10^{-10}\eta_{10} \qquad (4.104)$$

is the baryon-to-photon ratio. Given that the entropy density s is proportional to the photon number density n_γ, a very small value of η means a large entropy in the Universe today per baryon.

At this point we are therefore able to build a system of equilibrium equations,

$$
\begin{aligned}
X_n/X_p &= \exp(-Q/T) &(4.105)\\
X_2 &= \bar{C}(T/m_N)^{3/2} X_p X_n \eta \exp(B_2/T) &(4.106)\\
X_3 &= \ldots &(4.107)\\
\ldots & &(4.108)\\
1 &= X_n + X_p + X_2 + X_3 + \ldots &(4.109)
\end{aligned}
$$

A species becomes thermodynamically favoured when $X_A \simeq 1$, which depends on $\eta^{A-1}\exp(B_A/T)$. The binding energies for the four first light elements are (2=deuterium, 3=tritium):

$$
\begin{aligned}
B_2 &= 2.22\,\text{MeV}, \quad B_3 = 6.92\,\text{MeV}, \\
B_{^3He} &= 7.72\,\text{MeV}, \quad B_{^4He} = 28.3\,\text{MeV}.
\end{aligned}
$$

Notice that:

- The equilibrium abundances of all bound nuclei are negligible as long as free nucleons are in equilibrium due to the small η factor.
- The binding energies are all of the order of MeV, so their mass fractions are suppressed at least until the temperature is of the order of MeV. Indeed, it has to be lower than MeV to compensate the "suppression" given by the η^{A-1} factors.
- The largest binding energy is that of ^{4}He, so almost all the neutrons will end in ^{4}He nuclei. And $X_{^4He}$ becomes of order one at a temperature $T_{^4He} \simeq 0.28$ MeV.
- However, in order to produce ^{4}He we need first the deuterium, which has the smallest binding energy. Then, $X_2 \simeq 1$ at a lower temperature $T_2 \simeq 0.07$ MeV. Only then can the nuclear reactions proceed beyond deuterium, and nucleosynthesis starts.

4.7.2 Neutron to proton ratio

As we have seen before, neutrinos decouple at around $T \sim 1$ MeV; until then the interaction rate for weak reactions is higher than the expansion rate, and neutrons and protons are kept in equilibrium through the weak reactions:

$$\begin{aligned} n &\leftrightarrow p + e^- + \bar{\nu}_e \,, \\ \nu_e + n &\leftrightarrow p + e^- \,, \\ e^+ + n &\leftrightarrow p + \bar{\nu}_e \,. \end{aligned}$$

The low temperature of nucleosynthesis allows us to treat this part of the analysis first and use the resulting neutron abundance as input for the calculation of the light element abundances.

In chemical equilibrium, we have: $\mu_n + \mu_\nu = \mu_p + \mu_e$, and the ratio of neutrons to protons in equilibrium is given by:

$$\frac{n}{p} = \frac{n_n}{n_p} = \frac{X_n}{X_p} = \exp(\frac{\mu_n - \mu_p}{T} - \frac{m_n - m_p}{T}) = \exp(-\frac{Q}{T})\exp(\frac{\mu_e - \mu_\nu}{T}) \,, \tag{4.110}$$

where $Q = m_n - m_p = 1.293$ MeV. Moreover, given that the universe is electrically neutral we have $\mu_e/T \simeq n_e/n_\gamma = n_p/n_\gamma \simeq n_N/n_\gamma \simeq 10^{-10}$ (exercise!), and therefore this factor is negligible. We don't know the chemical potential of the neutrinos, so standard BBN calculations assume that $\mu_\nu = 0$. (It is possible to turn the argument around and use the observations to constrain μ_ν). We thus have

$$\frac{n_n^{(eq)}}{n_p^{(eq)}} = \exp(-\frac{Q}{T}) \,. \tag{4.111}$$

We can get a handle on how the interactions freeze out with the Boltzmann equation. We are interested in an interaction of the form $n + l \leftrightarrow p + l'$ where l and l' are leptons which we take to be in equilibrium. The equation then is

$$a^{-3}\frac{d(n_n a^3)}{dt} = n_l^{(eq)}\langle\sigma v\rangle \left(n_p \frac{n_n^{(eq)}}{n_p^{(eq)}} - n_n \right) . \tag{4.112}$$

We rewrite $n_l^{(eq)}\langle\sigma v\rangle$ as neutron-proton conversion rate λ_{np} which we need to get from particle physics. We also introduce the neutron fraction $X_n = n_n/n_N$ where $n_N = n_n + n_p$ is the nucleon density which scales like a^{-3}, and $X_p = 1 - X_n$. Expressed in this variable the Boltzmann equation is

$$\frac{dX_n}{dt} = \lambda_{np}\left\{(1 - X_n)e^{-Q/T} - X_n\right\} . \tag{4.113}$$

λ_{np} and T are functions of time, so that it is better to replace time by temperature, or more precisely by the "standard" variable $x = Q/T$, as we have done many times before. We find $dx/dt = -x\dot{T}/T \approx xH(x)$ and $H(x) \propto \sqrt{\rho} \propto T^2 \propto x^{-2}$, and the equation is now

$$\frac{dX_n}{dx} = \frac{x\lambda_{np}}{H(x=1)}\left(e^{-x} - X_n(1 + e^{-x})\right) . \tag{4.114}$$

$H(x=1) \approx 1/\mathrm{sec}$ so that the age of the universe at this point is of the order of 1 second. In the book of Dodelson, the neutron-proton conversion rate is parameterised as

$$\lambda_{np} = \frac{255}{\tau_n x^5}\left(12 + 6x + x^2\right), \tag{4.115}$$

where $\tau_n = 886.7$ seconds is the half-life of free neutrons. We also see that $\lambda_{np}(x=1) = 5.5/\mathrm{sec}$ which is larger than H but not by much, so the neutrons will start to fall out of equilibrium close to $T = Q$. Numerically integrating the equation leads to a freeze out temperature of $T_f = 0.5$ MeV and $X_n \approx 0.15$. This is indeed well above the temperature where nucleosynthesis starts, and so we were justified to treat the two problems separately. However, we have so far neglected the finite lifetime of the neutrons, so that we need to add in a radioactive decay factor $\exp(-t/\tau_n)$, which decreases the neutron abundance to about $X_n \approx 0.11$ until the start of nucleosynthesis.

4.7.3 The primordial abundances of light elements

The synthesis of the light elements thus proceeds roughly as follows:

1. At $T \approx 10$ MeV all nuclear processes are in equilibrium, $X_n, X_p \approx 1/2$, $X_{A\geq 2} \ll 1$.
2. At $T \approx 1$ MeV the processes keeping neutrons and protons in equilibrium freeze out, $X_n \approx 0.15, X_p \approx 0.85$. NSE is still a good approximation, and $\eta \ll 1$ ensures that $X_{A\geq 2} \ll 1$.
3. At $T \approx 0.1$ MeV n/p falls to about $1/8$ because of neutron decay. The abundance of helium, $X_{^4He}$, falls below its equilibrium value (which would be about 1) because of the delay in forming deuterium. NSE is now no longer valid. When ^4He is finally produced, the temperature (and so the kinetic energy of the nucleons) is too low for particles to break through the Coulomb barrier of the even heavier elements (metals) like carbon and oxygen. This together with the absence of strongly bound stable nuclei with $A = 5 \ldots 8$ means that nearly all neutrons end up in helium.

As we say before, ^{4}He cannot be synthesised until enough deuterium has formed, and effectively delays the start of nucleosynthesis until $T_{NS} \simeq 65\,KeV$. Because of that, and due to neutron decay, the neutron-to-proton fraction is slightly smaller than that at freeze-out. Assuming that all the neutrons are bound in ^{4}He, we can estimate its primordial abundance as

$$X_{^4He} = 4\frac{n_4}{n_N} \simeq 4\frac{n_n/2}{n_n + n_p} \simeq 2\frac{(n_n/n_p)_{NS}}{1 + (n_n/n_p)_{NS}} \simeq 2\frac{1/8}{1 + 1/8} \simeq 0.22\,. \tag{4.116}$$

Some deuterium and ^{3}He is left unburnt (of the order of 10^{-5}), and a little ^{7}Li and ^{7}Be is synthesised. All the remaining protons will end up as hydrogen. A more precise calculation for its abundance would predict $Y_4 = 0.2262 + 0.0135\ln(\eta_{10})$, in very good agreement with our approximate result.

Predictions:

The primordial abundances of light elements depends on the following parameters:

- *Neutron life-time*: which is used to express the rate of the weak interactions. Nowadays, this is pretty well-known. If τ_n were larger, for example, then we would have a smaller interaction rate, which would imply a larger freeze-out temperature and larger neutron-to-proton ratio, and therefore a larger ^{4}He abundance.

- *Baryon-to-photon ratio η*: The mass fraction for the nuclear species is proportional to this parameter, so the smaller η is, the smaller X_A. A higher η means that the mass fractions for the deuterium, tritium and ^{3}He build up earlier, so that the synthesis of ^{4}He is more efficient, we get more helium and less is left of the lighter nuclei. Until recently, this parameter was obtained by comparing BBN predictions with observations. Today, there is another independent way to measure it through the spectrum of anisotropies in the CMB.

- *Effective number of relativistic degrees of freedom*: or equivalently, the total energy density at the time of BBN. With 3 generations of neutrino species, we have $g_* = 10.75$ at $T \sim 1$ MeV. If there are extra relativistic degrees of freedom (like extra neutrinos), they will contribute:

$$g_* = 10.75 + \sum_B g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_F g_i \left(\frac{T_i}{T}\right)^4 . \tag{4.117}$$

  An extra relativistic degree of freedom means more energy density in radiation, and a larger expansion rate. Therefore, freeze-out occurs earlier, the neutron-to-proton ratio is larger, and so it will be the final primordial helium abundance. Observations constraint the extra number of neutrino families to be less than one.

- *Neutrino degeneracy (or neutrino chemical potentials)*: Standard BBN calculations assume that the chemical potentials for the neutrinos vanish. But strictly speaking their values are unknown. Non-vanishing chemical potentials of ν_μ and ν_τ will have the same effect than an increase in g_*, as they only increases the energy density of relativistic matter at BBN. The predictions are more sensitive to the electron neutrino chemical potential, as this will also enter when computing the neutron-to-proton ratio at freeze-out:

$$\left(\frac{n}{p}\right)_{eq} = \exp\left(-\frac{Q + \mu_{\nu_e}}{T}\right) . \tag{4.118}$$

  If $\mu_{\nu_e} > 0$, a non-zero value will decrease the neutron-to-proton ratio at freeze out, which means less ^{4}He is synthesised.

Chapter 5

Comparison with observations

This chapter discusses the three main observational "pillars" of the cosmological standard model. The first one is the observation of the relation between redshift and distance predicted by the FRW metric and the Friedmann equations, as well as the age of the universe deduced from the expansion history. The existence of a cosmic background radiation constitutes the second key observation, and the third one is the agreement between theoretical predictions and the observations of the abundance of light elements from nucleosynthesis.

5.1 Testing the expansion history

In the chapter on the FRW metric, we found the prediction that for nearby galaxies $v = H_0 r$ (the *Hubble law*). The verification of this law and the measurement of the proportionality constant (the *Hubble constant* H_0) has been one of the key aims of cosmology. Figure 5.1 from xxx shows the original data of Hubble replotted, together with a more recent compilation of data.

The distance ladder was built up over several stages, using first parallax measurements, which allow for a direct determination of distances to remote objects. In this way it is possible to determine a relationship between the pulsation rate and absolute luminosity of the variable Cepheid-type stars. Observations of these stars in other galaxies then allowed to add in secondary distance determinations, especially using type-Ia supernova (SN-Ia). The Hubble Key Project was dedicated this project and came up with $H_0 = 72 \pm 8$ km/s/Mpc.

As SN-Ia are able to probe the luminosity distance out to $z > 1$, we should start to see a deviation from the purely linear relationship between redshift and distance. We found that for a universe filled with different perfect fluids, the luminosity distance is given by

$$d_L(z) = a_0 S_\kappa(1+z)\left(\int_0^z \frac{dz'}{H(z')}\right) \tag{5.1}$$

and the Hubble parameter by

$$H(z)^2 = H_0^2\left[\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_\Lambda + (1-\Omega_r-\Omega_m-\Omega_\Lambda)(1+z)^2\right]. \tag{5.2}$$

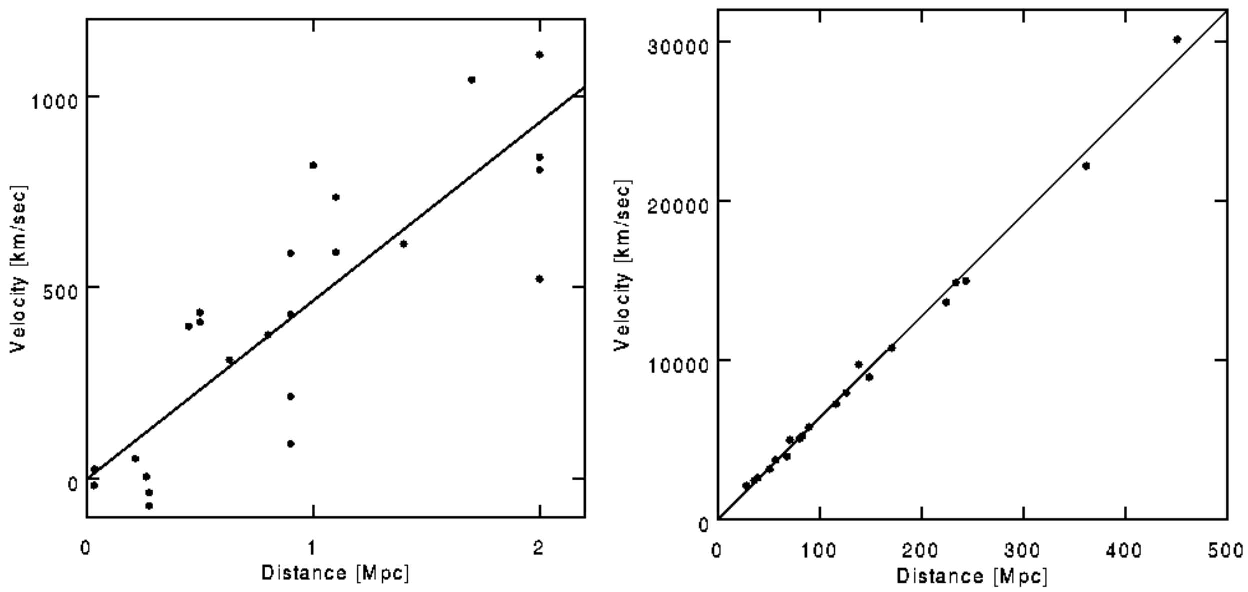

Figure 5.1: Hubble diagrams, showing the relationship between recessional velocities of distant galaxies and their distances. The left plot shows the original data of Hubble (1929), the right plot more recent data by Riess et al (1996). Notice the difference in scale.

Figure 5.2 shows the luminosity distance as measured by the SN-Ia.

We've discussed earlier that the energy density in radiation is much less than the one in matter for $z < 1000$, and so we can neglect Ω_r. If we are able to measure the luminosity distance well enough, then we can put limits on the amount of matter and the cosmological constant in the universe today (see fig. 5.2). For a flat universe, we find that $\Omega_m \approx 0.3$ and $\Omega_\Lambda = 1 - \Omega_m \approx 0.7$. As we don't know the origin of the cosmological constant, we often call it *dark energy*. We will expand on these measurements in the discussion of the anisotropies in the CMB in part II.

The expansion history also determines the age of the universe. Roughly, the age is given by $H_0^{-1} = 9.78/h$ Gyr, which for $h = 0.7$ points at an age of about 14 Gyr. Using the more precise formula (3.27) together with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ confirms this age. On the other hand, a universe with $h = 1$, $\Omega_m = 1$ and no cosmological constant has an expansion age of only $2/3 \cdot 9.78$ Gyr.

The lifetime of a star can be estimated as Mc^2/L where M is the mass of the star and L its luminosity. A somewhat more accurate description is given by

$$t \approx q \frac{0.007 Mc^2}{L} \tag{5.3}$$

where q is the fraction of the mass involved in H burning, and the factor 0.007 is the fraction of the hydrogen rest mass released during the fusion process. For the sun, we find a maximal age of about 10 Gyr. The age of the oldest stars (in globular clusters) is estimated to about 14 ± 2 Gyr. This clearly rules out a high density, high H_0 universe.

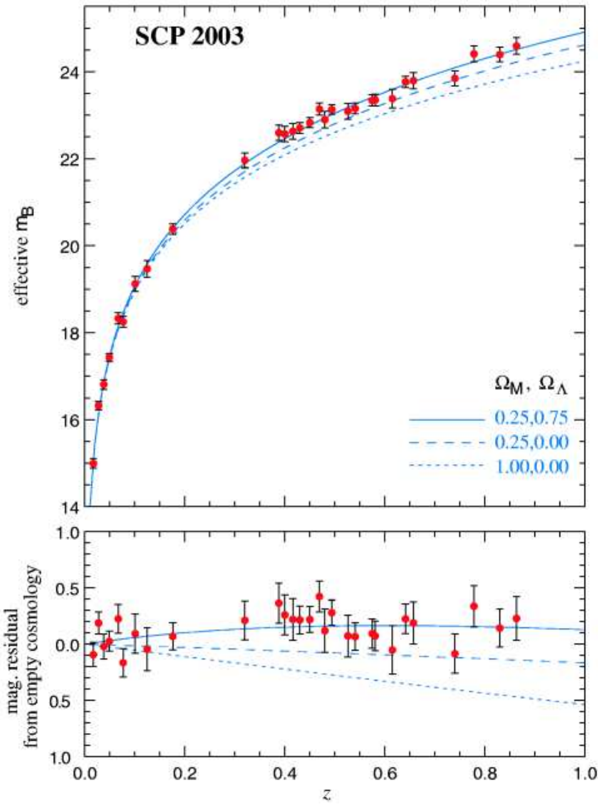

Figure 5.2: Distance-redshift diagram from the Supernova Cosmology Project. The distance is given in magnitudes, $m_B = -5 * \log d_L + 25$. The lower panel shows the difference in magnitudes of the three fiducial models to an empty universe, making it easier to compare them to the data.

Other ways to estimate astrophysical ages include ratios of long-lived radioactive isotopes, e.g. of uranium and thorium, with half-lifes of the order of Gyr. They allow to estimate the age of meteorites in our solar system to be about 4.5 Gyr.

5.2 The cosmic microwave background radiation

After the electrons and the nucleons recombine to form mostly neutral hydrogen and helium, the photons propagate freely through the now-transparent universe. As the universe expands, they slowly cool. Let us follow Gamov and estimate the temperature of the CMB today. We start by noting that early on, the universe is dominated by the radiation, and thus $\rho = \sigma_{SB} T^4$. Furthermore, as $a \propto 1/T$ it is $\dot{a}/a = -\dot{T}/T$. The Friedmann equation expressed with the temperature is thus

$$\left(\frac{\dot{T}}{T}\right)^2 = \frac{8\pi G \sigma_{SB} T^4}{3} \tag{5.4}$$

which is easily solved, leading to

$$t = \left(\frac{3}{32\pi G \sigma_{SB} T^4}\right)^{1/2}. \tag{5.5}$$

So at a temperature of $T_1 = 10^9$ K the age of the universe is about 200 s. Gamov then proceeds to use the abundance of matter in his argument, we will instead use the age of the universe computed above. The age of the universe today is about $H_0^{-1} = 14 \times 10^9$ years, or 4×10^{17} seconds. So $t_0/t_1 \approx 2 \times 10^{15}$. The expansion factor scales as $a \propto t^{1/2}$ during radiation domination, and as $a \propto t^{2/3}$ during matter domination. We can thus expect that the universe has expanded by something like $a_0/a_1 \approx 10^8 - 10^9$, and so that the resulting CMB temperature lies between 1K and 10K.

Indeed, Penzias and Wilson detected the radiation in 1964 and received the Nobel prize in 1978. The first precise observation had to wait for the COBE satellite in 1990, which proved that the spectrum has indeed a perfect blackbody form, and a temperature of $T_0 = 2.725 \pm 0.001$ K.

Among other things, COBE provided a strong limit on the chemical potential of the photons, $|\mu_\gamma| < 9 \times 10^{-5}$. At temperatures of $T < 1$ keV only Thomson scattering is active, which does not change the number of photons. If therefore any process dumps extra, non-thermalised energy into the photons, they cannot reach full equilibrium, but only kinetic equilibrium with a non-zero μ_γ. The COBE result therefore limits the amount of radiation created late in the thermal history of the universe.

The left panel of figure 5.3 shows that the spectrum is a thermal spectrum to a very high precision. It is the best black body known to mankind. This alone results immediately in some constraints: Thomson scattering preserves the number of photons. If any source dumps extra radiation into the universe at late times, when only Thomson scattering is active (i.e. $T \lesssim m_e$), then we will in general end up with the wrong number of photons for a given temperature. Thermalisation will remain incomplete, naively we expect to end up with a non-zero chemical potential μ, as this is the quantity linked to enforcing a given number of particles. A rigorous computation would confirm this result. Such an energy release does therefore distort the CMB spectrum and leads to a non-zero photon chemical potential μ in the distribution function (4.1). The fact that we don't observe this, places strong upper limits on the release of radiation (for example from the decay of

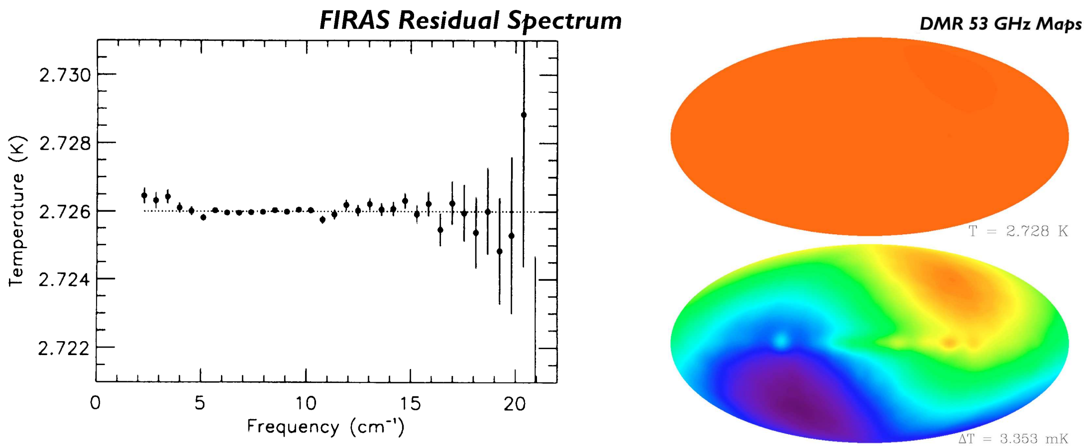

Figure 5.3: The left figure shows the black-body temperature of the CMB as a function of frequency, as measured by the FIRAS instrument on the COBE satellite. The right figure shows how homogeneous the CMB is, the strongest deviation is the velocity-induced dipole at the level of mK (from COBE/DMR).

massive particles) at temperatures below about 1keV. The COBE/FIRAS limits for the chemical potential of the CMB photon distribution are $|\mu| < 9 \times 10^{-5}$ at 95% confidence level (CL).

If the release is so late that no equilibrium is attained at all, or if photons are scattered after recombination by hot electrons, then this will produce additional distortions. Specifically the latter case, which we will encounter again towards the end of the lecture, is parametrised by the *Compton* y parameter. The COBE/FIRAS limits for this are $|y| < 15 \times 10^{-6}$ at 95% CL for a distribution function of the form

$$f_y(E,T) \propto \frac{E^2}{e^x - 1}\left(1 + yx\frac{1}{1-e^{-x}}\left\{x\coth(x/2) - 4\right\}\right). \tag{5.6}$$

where $x \equiv E/T$.

Starting with COBE, a whole series of experiments have also measured the tiny fluctuations in the CMB (at the level of about $10^{-5}T_0$). These fluctuations will play a central role in the second part of this lecture series, where we will spend a lot of time discussing their possible origin and evolution.

5.3 Abundances of the light elements and BBN

The predicted abundances of light elements in standard BBN (no extra neutrinos, no chemical potentials) as a function of η are plotted in Fig. 5.4 (from astro-ph/9903300). In order to compare the predictions to observations, we need to infer the primordial abundance of these elements. This is a difficult procedure, the two most important observables are:

• *Deuterium*: The deuterium can be easily burnt in starts to produce ^{3}He, so that its present abundance provides only a lower bound to the primordial one. The present limit on D/H is of the order of a few times 10^{-5}.

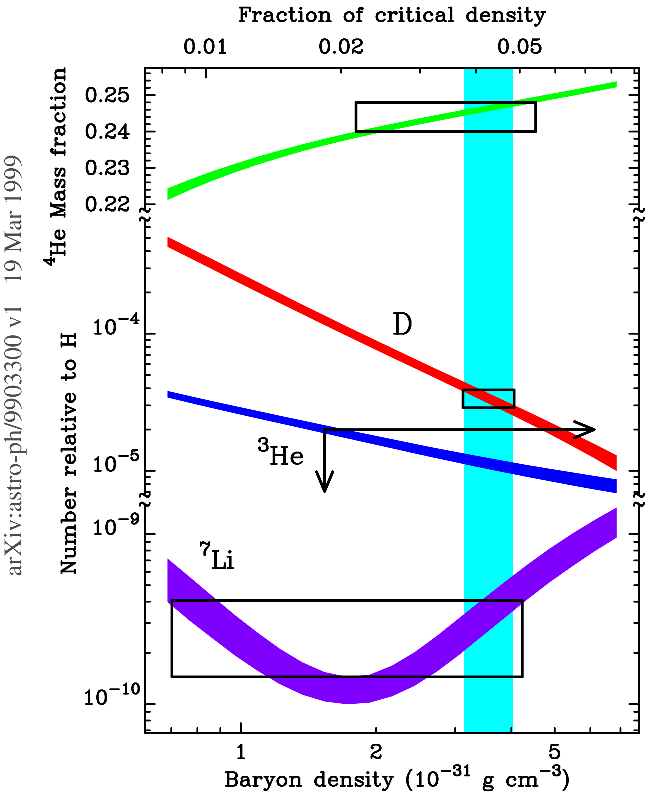

Figure 5.4: This figure shows the prediction and data from big-bang nucleosynthesis calculations. The agreement between data and theory over many orders of magnitudes is remarkable. (figure from astro-ph/9903300).

• *Helium-4*: This is very strongly bound nuclei, so it is not destroyed by stellar processes, but produced in stars together with other elements. The observed abundance should be larger than the primordial one. But in order to extract the value of the primordial abundance from the observations, the best place to look are systems of very low metalicity ("metals" in astrophysics are all the elements from oxygen onwards), and then try to extrapolate the value to zero metalicity. The observations give the range (A. D. Dolgov, hep-ph/0201107): $0.223 < Y_P < 0.245$.

We can use one of the abundances to measure the baryon to photon ratio $\eta = n_B/n_\gamma$. The other measurements then provide a highly non-trivial consistency check if we understand the physics at these energies. They also constrain the number of additional light neutrinos or any other contribution to the energy densities at very high redshifts.

We find that $\Omega_b \approx 0.05$. This is consistent with independent measurements at a redshift of $z = 1100$ from the anisotropies in the CMB.

We conclude that only about 5% of the energy density today is in the form of "normal" matter (i.e. baryons). If indeed $\Omega_m \approx 0.3$ then there has to be *dark matter*. Dark matter is non-baryonic but behaves like dust in the sense that it has a very small pressure and can cluster. Finally, if the universe is flat then we don't know what 95% of the energy in the universe is.

5.4 Conclusions

In the course of these lectures, we have discussed a cosmological model based on a hot beginning and a homogeneous and isotropic universe. We have seen in the last chapter that these initial assumption are supported by observations.

The isotropy is probed directly by the COBE (and now WMAP) measurements of the anisotropies in the cosmic microwave background radiation. The homogeneity is harder to assess – galaxy redshift surveys so far fail to see a clear turnover to homogeneity on large scales. On the other hand, if the universe is isotropic for two different observers, then it is also homogeneous. Hence, if the universe were isotropic but inhomogeneous, we would have to live very close to its "centre". This would seem very much a finetuning issue.

If we accept homogeneity and isotropy then the metric of the universe is uniquely given by the FLRW form, eq. (2.3). There are only two degrees freedom left, a free function called scale factor, $a(t)$, and the curvature κ. We can use standard candles at high redshifts to measure the recent time evolution of the scale factor. We find that the universe is expanding.

This we can understand by assuming that the universe is filled with perfect fluids, described by a given (constant) equation of state, $p = w\rho$. This does indeed seem a surprisingly good description. In this case, we follow the evolution of the universe backwards in time, we conclude that the universe must have been much hotter in the past that it is now.

This explains (and is required by) the existence of the CMB, which is the leftover radiation from the hot initial stage. Indeed, we find that it shows the time when the universe had cooled sufficiently so that the electrons and atom-cores combined to form neutral atoms.

If we move to higher and higher temperature, there will be a moment when the radiation field is strong enough to break apart the elements themselves, so that only free protons and neutrons are left. Conversely, we can predict the abundance of the different elements today, and we see that this again agrees with observations.

This standard model is therefore very compelling, and we know of no other model that is even remotely as successful. But there are still many questions left open:

- What sets the initial conditions?
- What created the fluctuations from which all the structure in the universe formed subsequently?
- What are the dark matter and the dark energy?
- How did the universe begin?

.

In the second part, we will present a possible answer to the first two points, and discuss how the fluctuations evolve.

Chapter 6

Inflation

We have seen that the cosmological standard model is very successful in explaining many observations like the cosmic microwave background radiation (CMB) and the element abundances. But there are a few questions left. Some of these are:

1. **horizon problem:** Why is the universe isotropic and homogeneous?
2. **flatness problem:** Why is $\Omega_{\rm tot} \approx 1$ today?
3. **monopole problem:** Why aren't there any?
4. **origin of fluctuations:** Where do the initial fluctuations come from?

In the following, we discuss the first three problems in turn, and we examine them specifically in a universe dominated by a perfect fluid with equation of state $w = p/\rho$.

Ex: Show that in this case and for a flat universe the scale factor evolves as

$$a(t) \propto \begin{cases} t^{2/(3+3w)} & w \neq -1 \\ e^{Ht} & w = -1 \end{cases} \tag{6.1}$$

6.1 Problems of the standard model

6.1.1 The horizon problem

We know from CMB observations that the universe at recombination was homogeneous to better than 10^{-4}. Why is that a problem? Let us investigate the distance travelled by a photon from the big bang to decoupling and then from decoupling to us.

A photon follows a null trajectory, $ds^2 = 0$. We choose its direction to be radial and find (writing η for the conformal time defined by $dt^2 = a^2(\eta)d\eta^2$):

$$d\eta^2 = dr^2 \tag{6.2}$$

and hence $\Delta r = \Delta\eta$. The comoving distance travelled since the big bang is therefore

$$d_p(\eta) = r(\eta) = \eta. \tag{6.3}$$

Using proper time t we find

$$d_p(t) = \eta(t) = \int d\eta = \int_0^t \frac{dt'}{a(t')}. \tag{6.4}$$

So the comoving region which has been in causal contact is $2d_p$ across, hence a patch of radius $d_p(t_{rec})$ was in causal contact at recombination. The region which we observe when we look at the CMB has a comoving size of

$$d_o = \int_{t_{rec}}^{t_0} \frac{dt'}{a(t')}. \tag{6.5}$$

Let us assume as an illustration that the universe was always matter dominated. In this case we find

$$\begin{aligned} d_p(t_{rec}) &\approx 3t_0^{2/3}t_{rec}^{1/3} & (6.6)\\ d_o &\approx 3t_0\left(1-(t_{rec}/t_0)^{1/3}\right) \approx 3t_0 & (6.7)\end{aligned}$$

and hence $d_0 \gg d_p(t_{rec})$.

How can we solve this problem? Let us assume that at a very early time t_i a different fluid dominated the energy density in the universe. Clearly, eq. (6.4) diverges if the exponent of t is larger or equal than 1, or $w \leq -1/3$. In this case, very large distances can have been in causal contact. We can get a different angle on what is happening by noting that $H(a) \propto a^{(3+3w)/2}$, while physical scales redshift as $\lambda(a) \propto 1/a$. The Hubble-length $1/H$ grows therefore slower than the physical scales for $w < -1/3$ so that a region of the universe that was *already* in causal contact in the past is red-shifted outside the horizon. Thus, after an inflationary phase, a region much larger than $1/H$ was in causal contact in the past.

6.1.2 The flatness problem

We can write the curvature radius of a FRW universe as

$$R_k^2 = \frac{a^2}{|k|} = \left|\frac{8\pi G\rho}{3} - H^2\right|^{-1} = \frac{1}{H^2|\Omega-1|}. \tag{6.8}$$

Today, the energy density of the universe is close to the critical density. Equivalently, the curvature radius is larger than the Hubble radius,

$$R_k^2(t_0) > \frac{1}{H_0^2}. \tag{6.9}$$

This might not seem extraordinary, but let us consider how the relative energy density Ω evolves as a function of time:

$$\Omega(t) \equiv \frac{\rho(t)}{\rho_c(t)} = \frac{8\pi G\rho}{3\left(\frac{\dot{a}}{a}\right)^2} = \frac{\rho}{\rho - \frac{3k}{8\pi G a^2}} \equiv \frac{1}{1-x(t)}. \tag{6.10}$$

Therefore $|x(t)| \propto 1/(\rho a^2) \propto a^{1+3w}$. For matter and radiation x grows therefore with time. As it has to be small today, it had to be even smaller in the past for the standard model of cosmology. We find

$$|\Omega(t) - 1| \approx x(t) = x_0 \left(\frac{a_{eq}}{a_0}\right) \left(\frac{a(t)}{a_{eq}}\right)^2 \approx 10^{-4} \left(\frac{1\text{eV}}{T}\right)^2, \tag{6.11}$$

and hence at the Planck energy of $\approx 10^{19}$GeV we find that $x(t_P) \approx 10^{-60}$ or smaller. This means that at the Planck time, the curvature contribution to the energy density has to be at least 60 orders of magnitude smaller than the contribution from the radiation. For "normal" matter, $\Omega = 1$ is an unstable fixpoint of the time evolution of the universe.

How can this problem be avoided? We have seen that $|x(a)| \propto a^{1+3w}$. So if the universe was at some point in the past dominated by a fluid with $w < -1/3$, then $|x|$ was growing smaller instead of larger (in this case $\Omega = 1$ has turned into a stable fixpoint). If this phase persisted for long enough, then the universe became arbitrarily close to being flat. Through this argument, we can later on derive a minimum duration for inflation.

6.1.3 The monopole problem

Phase transitions are often associated with symmetry breaking. In a supersymmetric model, all running coupling constants seem to reach the same magnitude at about 10^{16}GeV, suggesting a unification of the gauge groups $U(1) \times SU(2) \times SU(3) \to G$. If this is indeed the case, and if G is a simple group, then a general theorem states that monopoles are formed when G is broken into subgroups containing $U(1)$. This of course is what happens in GUT (grand unification) theories when the universe cools below the critical temperature T_c at which the unification happens.

The characteristic mass of the monopoles is the critical temperature of the phase transition, $m_M \approx T_c \sim 10^{16}$GeV. We expect generically to form about one monopole per Hubble volume, so their number density is

$$n_M(T_c) \sim H(T_c)^3 \sim g_*^{3/2} \frac{T_c^6}{m_P^3}. \tag{6.12}$$

Using the entropy density $s \sim g_* T_c^3$ we find

$$\frac{n_M}{s} \sim \left(\frac{T_c}{m_P}\right)^3. \tag{6.13}$$

As the monopoles are "hidden" from each other by their associated gauge field, they freeze out very rapidly and no monopole-antimonopole annihilation occurs. Their contribution to the total energy density in the universe is about

$$\rho_M(t_0) = \frac{m_M}{m_b} \frac{n_M}{s} \frac{s}{n_b} m_b n_b \tag{6.14}$$

and therefore $\Omega_M \sim 10^{16} \cdot 10^{-12} \cdot 10^{10} \cdot 0.05 \sim 10^{12}$.

As we can see, their number density is not too different from the one of normal matter as inferred from big-bang nucleosynthesis ($n_b/s \approx 10^{-10}$), but their mass is 10^{16} times larger than the typical proton mass!

This problem can be avoided by postulating that there is no unification of the gauge groups to a simple group. There is another solution as well, namely (surprise) a rapid expansion of the universe. In this case all the matter is diluted as $1/a^3$, so a sufficiently large increase in the scale factor can get rid of all the monopoles. And all the other matter and radiation. But the energy in the expansion is released at the end of the "inflationary" phase (this is called reheating). Afterwards, the thermal history can proceed as in the standard model. But if $T_c > T_{\rm reheat} \gg m_x$ where x is the normal kind of matter, then the monopoles are no longer a problem.

Formally, reheating is not an adiabatic process and does not conserve the entropy density. Hence if the universe cools like $T \sim 1/a$ while expanding some e^{70} fold (70 "e-foldings"), and is then reheated back to its original temperature, the entropy density grows like $s \sim T^3$ and hence by an enormous amount. The ratio n_M/s and therefore Ω_M become negligibly small.

6.2 Inflation with a scalar field

We have seen therefore that we need something with a negative pressure, $p < -\rho/3$. A cosmological constant would solve the problem, but then inflation never ends and the "normal" evolution of the standard model never starts. We need a dynamical mechanism to create the negative pressure.

Indeed, there is such a thing, namely a scalar field. Scalar fields represent spin-0 particles in field theories. Mathematically, it is just a function of space and time, usually written $\phi(x,t)$. We will continue to assume our universe to be homogeneous, so that the field depends only on time, $\phi(t)$. At any rate, if the universe expands very rapidly, any inhomogeneities will be smoothed out.

6.2.1 equation of state and equation of motion

The Lagrange functional of a scalar field is given by

$$\mathcal{L}_\phi = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi), \tag{6.15}$$

and we could obtain the energy momentum tensor through a variational approach. Instead, we will use it (or rather the energy density and the pressure) as our starting point:

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{6.16}$$

$$p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi). \tag{6.17}$$

We can think of the time derivative as a kinetic energy, while the term with the potential corresponds to a potential energy. As we don't know which particle is responsible for inflation (it is usually called the "inflaton"), we do not know the form of the potential. A few typical examples would be

$$V(\phi) = \frac{1}{2}m^2\phi^2 \qquad \text{massive scalar field} \tag{6.18}$$

$$V(\phi) = \frac{1}{2}\lambda(\phi^2-\sigma^2)^2 \qquad \mathrm{Higgs-type\ potential} \tag{6.19}$$

$$V(\phi) = \frac{1}{2}\lambda\phi^4 \qquad \mathrm{self-interacting\ scalar\ field} \tag{6.20}$$

A scalar field tries to minimise its energy. That means that it tries to move to the position with the least potential energy (the ground state), but in a way that does not lead to a too large contribution from the kinetic term. We could obtain the equation of motion of the field again from the Lagrange functional with a variational principle, but given the energy density and the pressure we can also obtain it from the energy-momentum conservation equation, $\dot{\rho} = -3H(\rho + p)$:

$$\ddot{\phi}\dot{\phi} + V'(\phi)\dot{\phi} = -3H\dot{\phi}^2, \tag{6.21}$$

or

$$\ddot{\phi} + 3H\dot{\phi} = -V'(\phi). \tag{6.22}$$

A prime denotes a derivative with respect to ϕ. The term containing H is the friction term due to the expansion of the universe. The Friedmann equation becomes

$$H^2 = \frac{8\pi}{3m_P^2}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right). \tag{6.23}$$

We have seen that to solve the problems of the cosmological standard model, we need to achieve $p < -\rho/3$. The potential energy has therefore to dominate the kinetic energy, and we will neglect $\dot{\phi}$ in the Friedmann equation (6.23). Since the time derivative should also remain small for some time, we will neglect the first term in the equation of motion. These approximations are called the *slow-roll approximations*. They are not strictly required for inflation to take place, but they are consistent and natural. The basic equations are therefore

$$H^2 = \frac{8\pi}{3m_P^2}V \tag{6.24}$$

$$3H\dot{\phi} = -V'. \tag{6.25}$$

An inflationary model is sometimes parameterised by expanding the potential in a Taylor series, i.e. in higher and higher derivatives of V. The first two so called *slow-roll parameters* are

$$\epsilon(\phi) \equiv \frac{m_P^2}{16\pi}\left(\frac{V'}{V}\right)^2, \tag{6.26}$$

$$\eta(\phi) \equiv \frac{m_P^2}{8\pi}\frac{V''}{V}. \tag{6.27}$$

The first one measures the slope of the potential and the second one the curvature. They can be related to the perturbations produced during inflation. In order for slow-roll to be self-consistent we need that

$$\epsilon \ll 1 \quad \text{and} \quad |\eta| \ll 1. \tag{6.28}$$

It is also possible to express the equation of state with the help of ϵ. For that we notice that in the slow-roll approximation $\dot{\phi}^2 = 2/3\epsilon V$ and therefore (to first order in ϵ,

$$p = \left(\frac{2}{3}\epsilon - 1\right)\rho. \tag{6.29}$$

Our old condition $w < -1/3$ translates therefore into $\epsilon < 1$.

Let us look at a few "typical" inflation scenarios:

6.2.2 scalar field inflation models

old inflation

In the old inflationary scenarios (the original inflation, proposed by Guth in 1981), the scalar field starts trapped in a "false vacuum" state, that is in a local minimum of the potential, with $V(\phi) > 0$. In this case, the inflaton field will undergo a first order phase transition by tunneling to the real vacuum (the ground state). Old inflation has been shown not to work, since it has what is called a "graceful exit" problem: The tunneling process creates bubbles of true vacuum, but they expand too slowly to ever fill the whole volume (which goes on inflating), and remain much smaller than the horizon today.

new inflation

The graceful exit problem can be cured by postulating a second order phase transition (e.g. in a double-well potential). In this case, the field within a fluctuation domain is very homogeneous ($\nabla\phi \sim 0$), and if the "top" of the potential is sufficiently flat, the field will start to roll slowly ($\dot{\phi} \sim 0$). During this "slow-roll" phase, the universe inflates. Later on, the field accelerates, leading to a break-down of inflation. It finally reaches the true vacuum and starts to oscillate. The coupling to other field then leads to the creation of particles and to reheating.

The main problem with new inflation that the flatness of the potential around $\phi = 0$ requires small couplings in the (effective) potential. In this case, the field is not in thermal equilibrium at early times, and so there is no reason for it to be near the origin, even if the effective potential has its true minimum there at high temperatures. This problem can actually be turned into a virtue and then leads to chaotic inflation.

chaotic inflation

If ϕ starts out large, homogeneous and static, it will begin inflating the universe. If the potential is flat enough, then the usual Hubble damping term will keep the field moving slowly and thus allow inflation to persist for a long time. In general, there are no further constraints on the potential (i.e. it can be of the form $m^2\phi^2$).

In general, chaotic inflation can go on forever in certain parts of the universe, and end in others. It thus leads to "islands" of flat FRW universes separated by inflationary gulfs.

hybrid inflation

In general (especially for models motivated by string theory) there are many scalar fields, possibly with couplings between them. In this case a second field can be used to end inflation and there is no graceful exit problem. It is also possible that other fields contribute strongly to the perturbations generated in inflation (*curvaton* scenario). However, these models can become arbitrarily complex and we will discuss only single-field models in these lectures.

6.2.3 The amount of inflation

The amount by which the universe inflates is measured as the number of *e-foldings*, which is the (natural) logarithm of the ratio of the scale factor before and after inflation,

$$N = \log\left(\frac{a(t_f)}{a(t_i)}\right). \tag{6.30}$$

Using $\log a = \int da/a$ we find

$$N = \int_{a_i}^{a_f} \frac{da}{a} = \int_{t_i}^{t_f} \frac{\dot{a}}{a} dt = \int_{t_i}^{t_f} H dt. \tag{6.31}$$

We can rewrite this result with the help of the slow-roll equations (6.24) and (6.25),

$$H dt = H \frac{dt}{d\phi} d\phi = H \frac{d\phi}{\dot{\phi}} = -\frac{3H^2 d\phi}{V'} = -\frac{8\pi}{m_P^2} \frac{V}{V'} d\phi. \tag{6.32}$$

so we finally find

$$N = -\frac{8\pi}{m_P^2} \int_{\phi_i}^{\phi_f} \frac{V}{V'} d\phi. \tag{6.33}$$

We have shown in the discussion of the flatness problem that we need to achieve $|1 - \Omega| < 10^{-60}$ and that this quantity scales like $1/a^2$ if we assume that $p = -\rho$. To solve the flatness problem, which requires an expansion of the universe by roughly a factor of 10^{30}, necessitates therefore about 70 e-foldings of inflation, which is not a problem for most models.

6.2.4 Example: chaotic inflation with a quadratic potential

The probably simplest example of inflation involves just a non-interacting massive scalar field with a potential $V(\phi) = 1/2 m^2 \phi^2$. In this case, the slow roll equations are

$$\begin{aligned} 3H\dot{\phi} + m^2\phi &= 0 && (6.34)\\ H^2 &= \frac{4\pi m^2}{3 m_P^2} \phi^2. && (6.35) \end{aligned}$$

The slow-roll parameters, eq. (6.28), are

$$\epsilon = \eta = \frac{m_P^2}{4\pi\phi^2}, \tag{6.36}$$

so inflation breaks down ($\epsilon \to 1$) at $|\phi_f| \approx m_P/\sqrt{4\pi}$. If the field starts out initially at ϕ_i then the total amount of inflation will be

$$N = 2\pi \frac{\phi_i^2}{m_P^2} - \frac{1}{2}. \tag{6.37}$$

We therefore require $\phi_i \gg m_P$. On the other hand, we can ensure that $V \ll m_p^4$ by choosing the mass small enough. The slow-roll equations can also be solved, resulting in

$$\phi(t) = \phi_i - \frac{m m_P}{\sqrt{12\pi}} t \tag{6.38}$$

$$a(t) = a_i \exp\left\{\sqrt{\frac{4\pi}{3}}\frac{m}{m_P}\left(\phi_i t - \frac{m m_P}{\sqrt{48\pi}} t^2\right)\right\}. \tag{6.39}$$

6.2.5 Reheating after inflation

During the inflationary period, the particles are redshifted away and the temperature drops rapidly to very low values. All the energy density is now stored in the inflaton field. We need to convert this energy back into particles and radiation if the universe is to continue with the normal big-bang evolution.

Once slow-roll inflation breaks down, the inflaton accelerates, reaches the minimum of the potential and starts to oscillate around it. At this point, the coupling to other fields becomes important. Originally, this decay was modeled by an additional dissipative term of the form $\Gamma\dot{\phi}$ in the evolution equation. As long as $\Gamma < H$ (i.e. during inflation) we can neglect the coupling and no particles are produced. As inflation ends, H decreases and the coupling leads to a relatively slow decay of the inflaton. Let us assume that inflation ends at a time t_f, and that the coupling (and thus the decay of the inflaton field) becomes important at a time $t_d > t_f$. We also assume that inflation ends as the inflaton field starts to oscillate around the minimum of the potential, in which case the pressure averaged over many oscillations vanishes, $\langle p \rangle = 0$. After the end of inflation the Friedmann equation is therefore

$$H^2 = \frac{8\pi}{3m_P^2}\rho_f\left(\frac{a_f}{a}\right)^3 \tag{6.40}$$

where we wrote $a_f \equiv a(t_f)$, etc. H is therefore decreasing, and at t_d we have $H = \Gamma$. For simplicity we assume that all the energy of the inflaton field is instantly transferred into radiation,

$$\rho_\phi(t_d) = \rho_\gamma = \frac{\pi^2}{30} g_* T_{\mathrm{RH}}^4 = \Gamma^2 m_P^2 \frac{3}{8\pi} \tag{6.41}$$

The final reheating temperature is therefore proportional to $T_{\mathrm{RH}} \propto \sqrt{\Gamma m_P}$. We remember that in the chaotic inflation example, $H_f \sim \sqrt{m/m_P}\phi_f \sim \sqrt{m m_P}$. We also required that $m \ll m_P$ to ensure that inflation takes place well below the Planck energy. In this case $\Gamma < H_f \ll m_P$ and so also $T_{\mathrm{RH}} \ll m_P$. More accurate calculations generally find relatively low reheat temperatures, less than maybe 10^{10} GeV.

Alternatively, $\Gamma \sim H$ already during inflation. This scenario is called *hot inflation*. In this case, radiation is produced all the time and there is the possibility of a smooth transition from inflation into the radiation dominated era.

About 10 years ago it was realised that this is not the full story. Let us look at our standard chaotic inflation example, with a coupling of the form $-\frac{1}{2}g^2\phi^2\chi^2$ to another light scalar field χ. At the end of inflation, when the inflaton field oscillates about the minimum of the potential, the equation of motion of the χ field becomes

$$\ddot{\chi}_k + 3H\dot{\chi}_k + \left(\frac{k^2}{a^2} + g^2\Phi^2\sin^2(m_\phi t)\right)\chi_k = 0, \tag{6.42}$$

where Φ is the (slowly decaying) amplitude of the fluctuations. This is the equation of an oscillator with a time-varying mass or frequency $\Omega(t) = k^2/a^2 + g^2\Phi^2 \sin^2(m_\phi t)$. Such a time-varying mass can lead to instabilities and parametric resonances. Indeed, one can rewrite eq. (6.42) in the form of the Mathieu equation, for which we know that there are parameter regions where χ grows exponentially. In this case, energy can be transferred very efficiently from the inflaton to the matter field. An explosive reheating ensues, and the final energy density in matter and radiation will be similar to the one in the inflaton field at the end of inflation. This can lead to significantly higher reheating temperatures and to non-thermal distribution of the energy (since the resonant k modes receive most of the energy). This version of reheating is commonly called *preheating*, and there seems to be still some controversy about the relative efficiency of the preheating and the normal (slower) reheating stage.

At any rate, the particles created during reheating are then supposed to decay, the particles and radiation start to interact and thermalise. The evolution of the universe re-enters the one of the standard big-bang model which we have discussed before.

6.3 The creation of initial fluctuations with inflation

Inflation can do more for us than solve the horizon-, flatness- and monopole problem. It creates also fluctuations which are necessary to create the large scale structure (galaxies, clusters, etc) visible in the universe. The discussion presented here is only very cursory and omits many details and potential problems. It is mainly intended to introduce the basic results.

The basic picture is that quantum fluctuations of the field are stretched beyond the horizon, where they become classical fluctuations. So we need to study the inflaton field at the quantum level to calculate the fluctuations. There are several possible approaches, the one which I will present here follows the book of Liddle & Lyth and a lecture by Mark Hindmarsh.

6.3.1 The quantisation of the free scalar field

The equation of motion of a free scalar field $\phi(\mathbf{x}, t)$ in flat, non-expanding space-time is

$$\ddot{\phi} - \nabla^2\phi + m^2\phi = 0 \tag{6.43}$$

and we now include the spatial derivatives. This equation is called the *Klein-Gordon equation*. To solve it we change into Fourier space, i.e. we apply the Fourier transform

$$\phi(\mathbf{x}, t) = \int \frac{d^3k}{(2\pi)^3}\phi(\mathbf{k}, t)e^{i\mathbf{k}\mathbf{x}}. \tag{6.44}$$

Then the Klein-Gordon equation is simply

$$\ddot{\phi}(\mathbf{k}, t) + E(k)^2\phi(\mathbf{k}, t) = 0 \tag{6.45}$$

with $E(k) = \sqrt{k^2 + m^2}$. This is just the equation of a harmonic oscillator, and its solutions are $\phi(\mathbf{k}, t) = \exp(\pm iE(k)t)$. The general solution is a superposition of these plane waves. Transformed back into "real" space we find

$$\phi(\mathbf{x},t) = \int \frac{d^3k}{(2\pi)^3} \frac{1}{\mathcal{N}(\mathbf{k})} \left[a(\mathbf{k}) e^{i(\mathbf{kx}-E(k)t)} + a(\mathbf{k})^* e^{-i(\mathbf{kx}-E(k)t)} \right]. \quad (6.46)$$

The normalisation $\mathcal{N}(\mathbf{k})$ is essentially arbitrary; it serves to determine the normalisation of the expansion coefficients $a(\mathbf{k})$. Typically one chooses either $\mathcal{N}(\mathbf{k}) = 1/(2E(k))$ or $\mathcal{N}(\mathbf{k}) = 1/\sqrt{2E(k)}$. Here we will use the former.

Going back to the Lagrange density, eq. (6.15),

$$\mathcal{L}_\phi = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2, \quad (6.47)$$

(here for the free massive scalar field), we can define the conjugate momentum

$$\pi(x) \equiv \frac{\partial \mathcal{L}}{\partial \dot{\phi}(x)} = \dot{\phi}(x). \quad (6.48)$$

In quantising the field, we promote both the field and the conjugate momentum to operators. We then impose the *equal time canonical commutation relations*

$$\begin{aligned} [\phi(t,\mathbf{x}), \pi(t,\mathbf{x}')] &= i\hbar\delta^3(\mathbf{x}-\mathbf{x}') && (6.49) \\ [\phi(t,\mathbf{x}), \phi(t,\mathbf{x}')] &= 0 && (6.50) \\ [\pi(t,\mathbf{x}), \pi(t,\mathbf{x}')] &= 0 && (6.51) \end{aligned}$$

The quantum version of the field will satisfy the same equations as the classical field, and we also find eq. (6.46), except that we have to interpret the expansion coefficients $a(\mathbf{k})$ and $a^*(\mathbf{k})$ now as operators $\hat{a}(\mathbf{k})$ and $\hat{a}^\dagger(\mathbf{k})$. In defining

$$\phi(t,\mathbf{k}) = \int d^3x \phi(t,\mathbf{x}) e^{-i\mathbf{kx}} \quad (6.52)$$

we can write as in (6.46)

$$\phi(t,\mathbf{k}) = \frac{1}{E(k)} \left(w(\mathbf{k},t)\hat{a}(\mathbf{k}) + w(\mathbf{k},t)^*\hat{a}^\dagger(-\mathbf{k}) \right), \quad (6.53)$$

where

$$w(\mathbf{k}) = e^{-iE(k)t}. \quad (6.54)$$

We also find that the operators $\hat{a}$ and $\hat{a}^\dagger$ have to satisfy the canonical commutation relations,

$$[\hat{a}(\mathbf{k}), \hat{a}^\dagger(\mathbf{k}')] = 2E(k)(2\pi)^3\delta(\mathbf{k}-\mathbf{k}'), \qquad [\hat{a}(\mathbf{k}), \hat{a}(\mathbf{k}')] = 0. \quad (6.55)$$

The additional factor of $E(k)$ in the commutation relations is due to our choice of normalisation $\mathcal{N}(k)$.

We now construct the standard space of state vectors, the *Fock space*, for the scalar field. To this end we define a ground state or vacuum $|0>$ as the state annihilated by all the operators $\hat{a}(\mathbf{k})$ (called *annihilation operators*,

$$\hat{a}(\mathbf{k})|0> = 0 \quad \forall\mathbf{k}. \quad (6.56)$$

The *creation operator* $\hat{a}^\dagger(\mathbf{k})$ is used to construct excited states,

$$|\mathbf{k}> = \hat{a}^\dagger(\mathbf{k})|0> . \tag{6.57}$$

We define that the vacuum state has unit norm, $<0|0> = 1$, and can then compute the normalisation of all other states, eg.

$$<\mathbf{k}|\mathbf{k}'> = <0|\hat{a}(\mathbf{k}')\hat{a}^\dagger(\mathbf{k})|0> = <0|[\hat{a}(\mathbf{k}'), \hat{a}^\dagger(\mathbf{k})]|0> = 2E(k)(2\pi)^3\delta(\mathbf{k}-\mathbf{k}'). \tag{6.58}$$

We interpret $|\mathbf{k}>$ as a single-particle state with momentum $\mathbf{k}$. We can also define a number operator. To this end we introduce a set of operators $\hat{n}(\mathbf{k}) = \hat{a}^\dagger(\mathbf{k})\hat{a}(\mathbf{k})$ for which we find

$$\hat{n}(\mathbf{k})|\mathbf{k}'> = 2E(k)(2\pi)^3\delta(\mathbf{k}-\mathbf{k}')|\mathbf{k}'> . \tag{6.59}$$

The number operator is then

$$N = \int \frac{d^3k}{(2\pi)^3}\frac{\hat{n}(\mathbf{k})}{2E(k)}, \tag{6.60}$$

and it returns the number of particles in a given state.

6.3.2 Classical fluctuations in expanding space-time

The expansion of space-time modifies the equation of motion (6.43). To compute the correct expression, we return to the action

$$S = \int d^4x\sqrt{-g}\left(\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + V(\phi)\right). \tag{6.61}$$

We set $g = \det g_{\mu\nu}$, and $d^4x\sqrt{-g}$ is the invariant measure to ensure that the action is a scalar under coordinate transforms.

For a flat FRW metric, we are dealing therefore with the action

$$S = \int d^4x a^3\left(\frac{1}{2}\dot{\phi}^2 - \frac{1}{2a^2}(\nabla\phi)^2 - V(\phi)\right), \tag{6.62}$$

where the gradient is in comoving coordinates. The corresponding Euler-Lagrange equations ($\delta S/\delta\phi = 0$) give us the equation of motion

$$\ddot{\phi} + 3H\dot{\phi} - \frac{1}{a^2}\nabla^2\phi + V'(\phi) = 0. \tag{6.63}$$

This equation neglects the perturbations of space-time itself due to the fluctuations in the inflaton field. But these perturbations are second order are therefore negligible if the universe is nearly homogeneous.

We now split the field into the homogeneous part $\phi(t)$ and a perturbation:

$$\phi(\mathbf{x},t) = \phi(t) + \delta\phi(\mathbf{x},t). \tag{6.64}$$

We insert this expression into the equation of motion. The potential is expanded in a series, $V'(\phi+\delta\phi) = V'(\phi) + \delta\phi V''(\phi) + O(\delta\phi^2)$, and we neglect all higher order terms in $\delta\phi$. If we keep only the homogeneous terms we recover the equation of motion (6.21), as of course we must. For the fluctuations we obtain

$$(\delta\phi)^{\cdot\cdot} + 3H(\delta\phi)^{\cdot} - \frac{1}{a^2}\nabla^2\delta\phi + V''(\phi)\delta\phi = 0 \tag{6.65}$$

We also rewrite $V''(\phi) = 3\eta H^2$. During the time which is interesting for us (some 60 e-foldings before the end of inflation), $\eta \ll 1$. For a massive, free scalar field, $V''(\phi) = m$.

In order to solve this equation, we need to perform a series of variable transforms. We start by introducing conformal time. As η is now reserved for a slow-roll parameter, we will call the conformal time τ in this section. For a universe expanding as a power law, $a \propto t^p$ we have that $\tau = \int dt/a(t) \propto t^{1-p}/(1-p)$ and $a \propto \tau^{p/(1-p)}$. Correspondingly,

$$\tau = \frac{p}{1-p}\frac{1}{aH} \to \frac{-1}{aH} \text{ for } a \propto e^{Ht}, \tag{6.66}$$

as exponential expansion corresponds to the limit $p \to \infty$. We also notice that $-\infty < \tau < 0$ for $p > 1$ (see exercises).

As in the case of the free field, we change into Fourier space to simplify the gradient terms. Following our discussion of the free field, we choose the Fourier expansion to be

$$\delta\phi(t, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3 E(k)} \left(a(\mathbf{k}) w(\mathbf{k}, t) e^{i\mathbf{kx}} + a(\mathbf{k})^* w(\mathbf{k}, t)^* e^{-i\mathbf{kx}} \right) \tag{6.67}$$

where $E(k) = \sqrt{k^2 + 3\eta H^2}$. Inserting this expansion into the differential equation for $\delta\phi$ we find

$$\ddot{w}(\mathbf{k}, t) + 3H\dot{w}(\mathbf{k}, t) + \left(\frac{k^2}{a^2} + 3\eta H^2 \right) w(\mathbf{k}, t) = 0. \tag{6.68}$$

We are normally interested in what happens on length scales of the order of the horizon, as this is where fluctuations become classical and freeze in during inflation. This scale corresponds to $k/a = H$. We see that we can therefore neglect the term proportional to $\eta \ll 1$ in the equation for w – alternatively we could compute corrections to the final spectrum. We also introduce a new variable $u(\mathbf{k}, t) = w(\mathbf{k}, t)a(t)$, for which the equation of motion is

$$\frac{d}{d\tau^2} u(\mathbf{k}, \tau) + \left(k^2 - \frac{2}{\tau^2} \right) u(\mathbf{k}, \tau) = 0. \tag{6.69}$$

It is easy to check that the following is a solution:

$$u(\mathbf{k}, \tau) = \frac{k\tau - i}{k\tau} e^{-ik\tau} \quad \Rightarrow \quad w(\mathbf{k}, \tau) = -H \frac{k\tau - i}{k} e^{-ik\tau}. \tag{6.70}$$

The second independent solution is its complex conjugate $w(\mathbf{k}, t)^*$.

6.3.3 The vacuum fluctuations of the inflaton field

In classical physics, the presence of fluctuations depends on the initial conditions. $\delta\phi = 0$ is perfectly good solution. But in quantum physics, the existence of the canonical commutation relations precludes this solution, and we expect to find fluctuations.

We now move through the same motions as in the case of a free field in flat, non-expanding space-time. We define the canonical momentum

$$\pi = \frac{\partial \mathcal{L}}{\partial \dot{\phi}} = a^3 \dot{\phi} \tag{6.71}$$

and impose the commutation relations

$$[\hat{\phi}(t,\mathbf{x}),\hat{\pi}(t,\mathbf{x}')] = i\delta^3(\mathbf{x}-\mathbf{x}'). \tag{6.72}$$

Expressed in Fourier space, we find again

$$[\hat{a}(\mathbf{k}),\hat{a}^\dagger(\mathbf{k}')] = 2E(k)\delta^3(\mathbf{k}-\mathbf{k}')(2\pi)^3\,, \quad [\hat{a}(\mathbf{k}),\hat{a}(\mathbf{k}')] = [\hat{a}^\dagger(\mathbf{k}),\hat{a}^\dagger(\mathbf{k}')] = 0 \tag{6.73}$$

We can now consider the mean square fluctuations of the field in the vacuum state, averaged over space,

$$\sigma^2 = \int d^3x < 0|\delta\phi^2(t,\mathbf{x})|0 > . \tag{6.74}$$

Substituting the expansion of the field in terms of the operators, eq. (6.67), and by using the commutation relations, we find

$$\sigma^2 = \int \frac{d^3k}{(2\pi)^3 2k}|w(t,\mathbf{k})|^2 = \int \frac{dk}{k}\left(\frac{H}{2\pi}\right)^2(1+k^2\tau^2). \tag{6.75}$$

By defining the power spectrum of the fluctuations to be

$$\mathcal{P}_{\delta\phi}(k) \equiv \frac{k^3}{2\pi^2}|w(k)|^2 \tag{6.76}$$

we find the result that

$$\mathcal{P}_{\delta\phi}(k) = \left(\frac{H}{2\pi}\right)^2(1+k^2\tau^2) \to \left(\frac{H}{2\pi}\right)^2 \tag{6.77}$$

as $\tau \to 0$ during inflation. We see that in an exponentially expanding universe, a massless scalar field fluctuates on all length scales with an amplitude $H/(2\pi)$.This spectrum is called *scale invariant*, since it doesn't depend on k.

It is also worthwhile to note that the phases of $\phi_{\mathbf{k}}$ are random and that inflation therefore generates (nearly) Gaussian perturbations.

6.3.4 The spectrum of the primordial curvature perturbation

In general, we do not use the perturbation spectrum of $\delta\phi$ directly, but the fluctuations induced in the metric, or more specifically the curvature. The main advantage is that the curvature perturbation remains constant outside the horizon and that it remains present after the scalar field has decayed and thus disappeared, fluctuations and all.

We start by transforming the fluctuations in the field itself into fluctuations in its density $\delta = \delta\rho/\bar{\rho}$:

$$\rho(\phi) = \rho(\bar{\phi}+\delta\phi) = \rho(\bar{\phi}) + \left.\frac{\partial\rho}{\partial\phi}\right|_{\phi=\bar{\phi}}\delta\phi + \ldots = \bar{\rho}+\delta\rho \tag{6.78}$$

and

$$\left.\frac{\partial\rho}{\partial\phi}\right|_{\phi=\bar{\phi}} = \frac{\partial\bar{\rho}}{\partial t}\frac{\partial t}{\partial\bar{\phi}} = \frac{\dot{\bar{\rho}}}{\dot{\bar{\phi}}} \tag{6.79}$$

Together with the continuity equation $\dot{\bar{\rho}} = -3H(1+w)\bar{\rho}$ we find finally

$$\delta_k = -\frac{3H}{\dot{\phi}}(1+w)\delta\phi_k. \tag{6.80}$$

What perturbations does this induce in the metric? We parametrise the metric perturbations as

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + a^2 h_{\mu\nu} \quad |h_{\mu\nu}| \ll 1, \tag{6.81}$$

and

$$h_{\mu\nu}dx^\mu dx^\nu = 2\Psi dt^2 - 2\Phi\gamma_{ij}dx^i dx^j. \tag{6.82}$$

γ_{ij} is the spatial part of the metric. Φ and Ψ are called *Bardeen potentials* and they represent the scalar degrees of freedom of the metric perturbations (as opposed to vector and tensor perturbations). For perfect fluids $\Psi = -\Phi$ corresponds to the Newtonian potential (at least for distances much less than the Hubble radius). Indeed, in coordinate systems moving with the fluid, the Poisson equation holds for the perturbations:

$$4\pi G\delta\rho = \nabla^2\Psi. \tag{6.83}$$

In Fourier space:

$$-\left(\frac{k}{a}\right)^2\Psi = 4\pi G\delta\rho_k = 4\pi G\bar{\rho}\delta_k = \frac{3}{2}H^2\delta_k. \tag{6.84}$$

At horizon crossing where $k/a = H$ we find therefore

$$\Phi|_{k=aH} = \frac{3}{2}\left(\frac{aH}{k}\right)\delta_k \to \frac{3}{2}\delta_k|_{k=aH}. \tag{6.85}$$

Putting it all together, the perturbation spectrum can now be written as

$$\mathcal{P}_\Phi|_{k=aH} = \frac{9}{4}\mathcal{P}_\delta|_{k=aH} = \frac{9^2}{4}(1+w)^2\frac{H^2}{\dot{\phi}^2}\left(\frac{H}{2\pi}\right)^2\Bigg|_{k=aH} \tag{6.86}$$

We often use instead the curvature perturbation defined as $\Phi = 9/2(1+w)\mathcal{R}$ or

$$\mathcal{R}_{\mathbf{k}} = -\frac{H}{\dot{\phi}}\delta\phi_{\mathbf{k}} \tag{6.87}$$

The spectrum of $\mathcal{R}$ is then

$$P_{\mathcal{R}}(k) = \left(\frac{H}{\dot{\phi}}\right)^2\left(\frac{H}{2\pi}\right)^2 \tag{6.88}$$

where the right hand side is again evaluated at horizon crossing. Using the slow-roll equations $3H\dot{\phi} = -V'$ and $3H^2m_P^2 = 8\pi V$ we can write

$$P_{\mathcal{R}}(k) = \frac{2\pi}{m_P^6}\frac{V^3}{V'^2} = \frac{2}{3m_P^4}\frac{V}{\epsilon} \tag{6.89}$$

To relate this with observations, it is customary to consider $\delta_H^2(k) \equiv 4/25 P_{\mathcal{R}}(k)$. If we assume a power-law form for the curvature perturbation, and the normalisation provided by the CMB measurements of the COBE satellite, we find that

$$\delta_H(k) = 1.91 \times 10^{-5} \left(\frac{k}{k_*}\right)^{(n-1)/2} \tag{6.90}$$

where $k_* = 7.5 a_0 H_0$ is the scale at which the normalisation is independent of the spectral index, n. Assuming that the full COBE measurement is due to density fluctuations (neglecting gravity waves, see the next chapter), we can give a numerical value for the potential at k_*,

$$V^{1/4} = 6.6 \times 10^{16} \mathrm{GeV} \epsilon^{1/4}. \tag{6.91}$$

To calculate the spectral index n we define it as

$$n(k) - 1 \equiv \frac{d \ln P_{\mathcal{R}}}{d \ln k}. \tag{6.92}$$

Since we evaluate all quantities at $k = aH$ and since H is essentially constant, we have $d \ln k = H dt$. Using the slow-roll equations we find

$$\frac{d}{d \ln k} = -\frac{m_P^2}{8\pi} \frac{V'}{V} \frac{d}{d\phi}. \tag{6.93}$$

It is then easy to compute the derivatives of the slow-roll parameters,

$$\frac{d\epsilon}{d \ln k} = -2\epsilon\eta + 4\epsilon^2 \tag{6.94}$$

$$\frac{d\eta}{d \ln k} = 2\epsilon\eta - \xi^2, \tag{6.95}$$

where

$$\xi^2 \equiv \frac{m_P^2}{8\pi} \frac{V' V'''}{V^2}. \tag{6.96}$$

ξ^2 can be positive or negative, the superscript ^{2} only means that it is a second-order quantity. We find therefore that

$$n - 1 = -6\epsilon + 2\eta. \tag{6.97}$$

In slow-roll inflation, we see that the deviation from $n = 1$ is expected to be small. Indeed, the current limits are around $|n - 1| < 0.2$. In the same way we could compute higher and higher corrections to $n(k)$ using higher order slow-roll parameters like ξ^2.

If $n \neq 1$ it is called a *tilted spectrum*. The case $n > 1$ is often referred to as a *blue* spectrum, and the other case as a *red* spectrum.

6.3.5 Gravitational waves

General relativity predicts the existence of gravitational waves, which represent a perturbation of the metric:

$$ds^2 = dt^2 + a(t)^2 \left(\delta_{ij} + \frac{2}{a^2} h_{ij}\right) dx^i dx^j. \tag{6.98}$$

The h_{ij} are also called *tensor perturbations* while the density perturbations are called *scalar perturbations*. They are traceless, $h_i^i = 0$ and transverse, $\partial_i h_{ij} = 0$ so that there are only two independent components, commonly called h_+ and $h_\times$. In Fourier space and a coordinate system where the wave vector $\mathbf{k}$ is pointing into the 3-direction we find $h_{11} = -h_{22} = h_+$ and $h_{12} = h_{21} = h_\times$, while all other components vanish.

In the absence of anisotropic stresses (which is the case during inflation) the Einstein equations give their equation of motion as,

$$\ddot{h}_{+,\times} + 3H\dot{h}_{+,\times} + \left(\frac{k}{a}\right)^2 h_{+,\times} = 0, \tag{6.99}$$

which is the equation of motion of a massless scalar field. It is also the same equation as for the inflaton field. During inflation, the vacuum fluctuations of the h field will hence lead to a background of gravitational waves in precisely the same way as the vacuum fluctuations of the inflaton field. Their spectrum is just given by eq. (6.77), up to a numerical factor, and is

$$P_{\text{grav}}(k) = \frac{2}{m_P^2}\left(\frac{H}{2\pi}\right)^2\Bigg|_{k=aH}. \tag{6.100}$$

The factor of 2 is due to the two polarisations, $+$ and $\times$, and the Planck mass (the only scale in GR) ensures that the dimensions are correct.

In the same way as before we can define the spectral index

$$n_{\text{grav}} = \frac{d\ln P_{\text{grav}}(k)}{d\ln k} \tag{6.101}$$

or

$$n_{\text{grav}} = -2\epsilon. \tag{6.102}$$

So vacuum fluctuations of both the inflaton field and the gravitational waves lead to perturbations of the background universe. In that way, they also generate anisotropies in the cosmic microwave background. The relative size of their contributions can be computed and is approximately

$$r = \frac{T}{S} \approx 12.4\epsilon, \tag{6.103}$$

where the exact numerical prefactor depends on many things like the matter content in the universe and the scale at which it is measured. Using this expression, we can directly relate r and n_{grav}:

$$r = -6.2 n_{\text{grav}}. \tag{6.104}$$

This is called the *consistency relation*. It applies only to single-field inflationary models, and is different in more general cases. The existence of this relation is not surprising, since the creation of both the inflaton density perturbations and the gravitational waves have been are connected to the potential $V(\phi)$.

The experimental confirmation of the consistency relation would be a major success of the single field inflation model, all but proving it. Unfortunately it is very unlikely that we will ever be able to test it. Even the newest CMB experiments like WMAP set limits on r of the order of $r < 0.7$ (depending on assumptions about the underlying model), while most models of inflation predict only a very small amount of tensor waves. Even once they are measured it will still be very hard to determine their spectral index with any accuracy.

6.3.6 Example: quadratic potential

Returning to our old example of the last section, we easily compute the expected perturbations. First, we remember that the slow-roll parameters were

$$\epsilon(\phi) = \eta(\phi) = \frac{m_p^2}{4\pi\phi^2} \tag{6.105}$$

and the number of e-foldings until the end of inflation

$$N = \frac{2\pi\phi^2}{m_p^2} - \frac{1}{2} \tag{6.106}$$

(where we will neglect the $1/2$ from now on).

The scalar spectral index is then

$$\begin{aligned} n_S - 1 &= -6\epsilon + 2\eta = -4\epsilon & (6.107)\\ &= -\frac{m_p^2}{\pi\phi^2} = -\frac{2}{N} \approx -0.04 & (6.108)\end{aligned}$$

where we assumed in the last step that the scales of interest left the horizon 50 e-foldings before the end of inflation. We expect therefore $n_S \approx 0.96$, well compatible with the current WMAP results of $n_S = 0.99 \pm 0.04$ at 1σ.

The spectral index of the gravitational waves, or tensor spectral index n_T is found to be

$$n_T = -2\epsilon = \frac{n_S - 1}{2}, \tag{6.109}$$

and the ratio of scalar to tensor contributions is

$$r \approx 12.4\epsilon = -3.1(n_S - 1) \approx 0.12 \tag{6.110}$$

We see that this model leads to a potentially observable contribution from gravity waves.

We can also compute the mass of the inflaton field using the COBE normalisation of δ_H and assuming again that the COBE scales crossed the horizon 50 e-foldings before the end of inflation. We find that $m = 1.8 \cdot 10^{13}$ GeV.

Ex: Repeat these calculations for the case of power law inflation.

Chapter 7

Evolution of the perturbations

7.1 Definition of the perturbations

This discussion is motivated by the book of Padmanabhan. The fully general-relativistic treatment of perturbations is rather involved and leads to diverse technical issues. We are not going to work our way through the full calculation. Instead, we will have a general look at what is involved, and then use several well-known equations as a starting point.

In general, we need to define perturbations in the metric and the energy momentum tensor,

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + a^2 h_{\mu\nu} \qquad |h_{\mu\nu}| \ll 1 \tag{7.1}$$

$$T^{\mu}_{\nu} = \bar{T}^{\mu}_{\nu} + \Theta^{\mu}_{\nu} \qquad |\Theta^{\mu}_{\nu}| \ll \bar{\rho} \tag{7.2}$$

where we defined the averaged quantities

$$\bar{g}_{00} = a^2 \qquad \bar{g}_{ij} = -a^2\delta_{ij} \tag{7.3}$$

$$\bar{T}^{0}_{0} = \bar{\rho} \qquad \bar{T}^{i}_{j} = -\bar{p}\delta_{ij}. \tag{7.4}$$

These correspond to the metric and energy momentum tensor of the homogeneous and isotropic (flat) FLRW model.

We can then formulate the Einstein equations with these quantities and compute the result to first order in the perturbations. This is quite a long calculation, and it suffers from a technical issue, known as the gauge problem: We can only observe the real, perturbed universe but not the underlying "background" FLRW universe. Indeed, there is no unique background universe, it is not even certain that one exists. Even if we assume the existence of a background universe, we will always find a whole family of them. The different FLRW universes are connected by isomorphisms, known as *gauge transformations*.

One then has to either fix the gauge (the reference system) uniquely, or work with gauge-invariant variables. We are going to pretend following the second route in this section, although we will really avoid talking about the issue in any detail. At any rate, a measurable physical quantity cannot depend on the choice of a gauge. On scales much smaller than the horizon, where we reach an essentially Newtonian limit, we expect therefore that we can neglect the gauge issue.

A useful result is the Stewart-Walker lemma which says that any perturbation variable that belongs to a quantity vanishing in the background FLRW universe is automatically gauge invariant. One example are gravitational waves, which are generically created by inflation. In the exercises we will discuss this case in detail and compute the perturbed Einstein equations.

In the lecture we will instead concentrate on the quantities that can be written as scalar functions of space-time. Assuming as usual that only perfect fluids are present in the universe, we need to discuss three such quantities: The density, the velocity and the gravitational potential.

7.1.1 The density perturbation

We have already defined the density perturbations in the chapter on inflation,

$$\rho(\mathbf{x}, t) = \bar{\rho}(t) + \delta\rho(\mathbf{x}, t), \tag{7.5}$$

and $\delta \equiv \delta\rho/\bar{\rho}$. However, we will require the density perturbation for two different choices of background universe. The first one we define of hyper-surfaces of vanishing curvature,

$$D_g \equiv \delta|_{\mathcal{R}=0}. \tag{7.6}$$

The second one we define on hyper-surfaces that are moving with the fluid, ie.

$$D \equiv \delta|_{v_i=0}. \tag{7.7}$$

where $v_\mu = (v_0, v_i)$ with $v^2 = -1$ is the four velocity of the fluid. The connection between the two, in terms of the other perturbation variables, is given by

$$D = D_g + 3(1+w)\left(\Psi + H\frac{v}{k}\right). \tag{7.8}$$

7.1.2 The velocity perturbation

If the density is not constant, then we expect that the fluid will move around in space. A spatial velocity 3-vector v_j can always be decomposed into a divergence $\nabla \cdot \mathbf{v}$ and a rotation. In Fourier space this is very simple,

$$v_j(\mathbf{k}) = i\mathbf{k}_j v^{(S)}(\mathbf{k}) + v_j^{(V)}(\mathbf{k}), \tag{7.9}$$

with $k^j v_j^{(V)} = 0$. The divergence-part is a scalar quantity, and only this part is sourced (to first order) by a inhomogeneous density distribution (as we will see when we consider the equations).

We define the following scalar quantity:

$$H(\mathbf{x}, t) \equiv \frac{1}{3}\partial_j v^j(\mathbf{x}, t). \tag{7.10}$$

In the absence of perturbations, $H(t, \mathbf{x})$ reduces to $\bar{H}(t) = \dot{a}/a$ the Hubble constant of the background universe, since $\bar{\mathbf{v}} = \dot{a}/a\mathbf{x}$. We can therefore think of $\nabla \cdot \mathbf{v}$ as the "perturbed Hubble constant". Instead of $H = \bar{H} + \delta H$ we could also use $v = 3\delta H/k$.

7.1.3 The metric perturbations

We define the (scalar) metric perturbations again through

$$h_{00} = 2\Psi, \quad h_{ij} = -2\Phi\delta_{ij}. \tag{7.11}$$

For perfect fluids $\Phi = -\Psi$ and we will only use one of them. On distances much shorter than the horizon, Ψ corresponds to the Newtonian gravitational potential.

7.2 The perturbation equations

7.2.1 The conservation equation

The usual conservation equation holds,

$$\dot{\rho}(\mathbf{x}, t) = -3H(\mathbf{x}, t)\left(\rho(\mathbf{x}, t) + p(\mathbf{x}, t)\right). \tag{7.12}$$

We decompose all quantities into a (homogeneous) background contribution and the perturbations,

$$\begin{aligned} \partial_t(\bar{\rho} + \delta\rho) &= -3(\bar{H} + \delta H)(\bar{\rho} + \delta\rho + \bar{p}\delta p) & (7.13)\\ &= -3\left[\bar{H}(\bar{\rho} + \bar{p} + \bar{H}(\delta\rho + \delta p) + \delta H(\bar{\rho} + \bar{p}\right] + \text{2nd order} & (7.14) \end{aligned}$$

To lowest order we recover the usual conservation equation for the background quantities, $\dot{\bar{\rho}} = -3\bar{H}(\bar{\rho}+\bar{p})$. As before we introduce $w \equiv \bar{p}/\bar{\rho}$ and additionally also the speed of sound $c_s^2 = \delta p/\delta\rho$. For (pressureless) matter $w = c_s^2 \approx 0$ and for radiation $w = c_s^2 = 1/3$. The equation for the first order quantities, using that in a frame of reference where the curvature perturbation vanishes we have that $\delta\rho/\bar{\rho} = D_g$, is

$$\dot{D}_g + 3\bar{H}(c_s^2 - w)D_g + 3(1 + w)\delta H = 0. \tag{7.15}$$

Using the two other perturbation equations discussed below, we find after a rather long calculation that the conservation equation expressed in the variable D is

$$\dot{D} - 3\bar{H}wD + 3(1 + w)\delta H = 0. \tag{7.16}$$

7.2.2 The Poisson equation

In GR a flow of matter or radiation contributes also to the energy of a system. It is therefore not surprising that we recover the Newtonian equation of gravitation only in a frame of reference where we move with the fluid, the system in which we have defined the density perturbation variable D. Indeed, if we were to solve the Einstein equations, we would find just the usual Poisson equation,

$$\nabla^2\Psi = 4\pi G\delta\rho|_{\text{comoving}} = 4\pi G\bar{\rho}D. \tag{7.17}$$

If we deal with the dominant fluid, we can additionally use the Friedmann equation $8\pi G\bar{\rho} = 3\bar{H}^2$. In Fourier space the Laplace operator ∇^2 becomes $-(k/a)^2$ where k is the comoving wave number.

7.2.3 The Euler equation

The (special) relativistic Euler equation is

$$\frac{dv^i}{dt} = -\frac{1}{\rho + p}\partial^i p, \tag{7.18}$$

where the total time derivative is defined as

$$\frac{dv^i}{dt} = \frac{\partial v^i}{\partial t} + (v^j \partial_j)v^i. \tag{7.19}$$

The full equation in GR is rather complicated, as the total time derivative has to be replaced by a covariant derivative which includes all the Christoffel symbols, etc. Instead we can add the gravitational force $-\partial^i \Psi$ to the right hand side. However, this is only the contribution of the perturbations to the gravitational force, and so we cannot expect to recover the background equation, only the perturbation equation.

To recover a scalar equation, we consider the divergence of the Euler equation. The left-hand-side then becomes

$$\partial_i \dot{v}^i + \partial_i \left(v^j \partial_j v^i\right) = \partial_t(\partial_i v^i) + (\partial_i v^j)(\partial_j v^i) + (v^j \partial_j)(\partial_i v^i) \tag{7.20}$$

The first term is just $3\dot{H}$ and the last term vanishes as $v^j = 0$ in the comoving reference frame. To extract the scalar degree of freedom of the second term, we define $2\omega_{ij} = (\partial_j v_i - \partial_i v_j)$ and $2\sigma_{ij} = \partial_j v_i + \partial_i v_j - 2H\delta_{ij}$. The the second term is $3H^2 + 2(\sigma^2 - \omega^2)$ where $2\sigma^2 = \sigma_{ij}\sigma^{ij}$ and the same for ω. But σ^2 and ω^2 are both second order. The left-hand-side is therefore $3\dot{H} + 3H^2$ and the first order part is $3(\partial_t \delta H + 2H\delta H)$.

The divergence of the gravitational force is just the Laplacian applied to Ψ, $-\nabla^2\Psi = (k/a)^2\Psi = -4\pi G\bar{\rho}D$, where we used in addition the Poisson equation. In the third term we notice that $\partial^i \bar{p} = 0$ and $\partial^i \delta p = c_s^2 \bar{\rho}\partial^i D$. As we now have a purely first order quantity in the numerator, only the zero-order part of the denominator, $\bar{\rho} + \bar{p} = (1+w)\bar{\rho}$ is relevant. Overall, we find

$$\partial_t \delta H + 2\bar{H}\delta H = -\frac{4\pi G}{3}\bar{\rho}D + \left(\frac{k}{a}\right)^2 \frac{c_s^2}{3(1+w)}D. \tag{7.21}$$

Combining this equation with the continuity equation and eliminating δH, we find after some algebra, using the Friedmann equations, a single second-order differential equation for D:

$$\ddot{D} + \bar{H}\dot{\delta}\left(2 - 3(2w - c_s^2)\right) - \frac{3}{2}\bar{H}D\left(1 - 6c_s^2 - 3w^2 + 8w\right) = -\left(\frac{k}{a}\right)^2 c_s^2 D, \tag{7.22}$$

However, if we deal with several fluids that interact only gravitationally then we have to go back to our three initial equations. In this case we can write an equation for D and δH for each fluid independently. These equations are coupled together by the common metric perturbation Ψ which represents the gravitational interaction.

In the remainder of this section we will leave out the bar over the background quantities.

7.3 Gravitational instability in an expanding universe

We can now use these equations to derive the behaviour of the perturbations in the universe. Before we study specific cases, it is worthwhile to examine the general structure of the perturbation equation. Consider a perturbation labelled by a wave number k, in the linear approximation. We rewrite the perturbation equation in the following way:

$$\ddot{D}_A = -\alpha_A H \dot{D}_A + \left(\mu_A H^2 - \frac{k^2 c_{s,A}^2}{a^2}\right) D_A \tag{7.23}$$

where $H = (\dot{a}/a)$, $c_{s,A}$ is the sound speed relating the perturbed pressure and the perturbed density, and the variables μ_A, α_A are quantities which depend on the equation of state of the background matter.

The first term on the right hand side always dampens the growth, this is purely an effect of the expansion. The expansion rate is of course determined by all matter in a multicomponent fluid. If the dominant smooth component is not the species A, then the damping due to expansion may suppress the growth.

The second term represents the conflict between pressure support and gravity. This term will act as a "restoring force" and prevent growth if (ignoring the numerical factor μ_A)

$$a^2 H^2 < k^2 c_{s,A}^2. \tag{7.24}$$

With $H^2 = 8\pi G \rho_{dom}/3$ and $k^2 = (2\pi/\lambda)^2$ we can write this as

$$\frac{\lambda a}{c_s} = \frac{\lambda_{\rm phys}}{c_s} < \frac{1}{\sqrt{G\rho}}, \tag{7.25}$$

or

$$\lambda < \lambda_J \equiv \frac{c_s}{\sqrt{G\rho}} \tag{7.26}$$

where we introduced the *Jeans* length λ_J. This is the same as saying that the pressure time constant $t_{\rm pressure} = \lambda/c_s$, i.e. the time scale on which pressure can resist gravitational collapse, is smaller than the free-fall time scale $t_{\rm grav} = 1/\sqrt{G\rho}$. The Jeans length separates the two regimes.

Let us now discuss various special cases that we will need in order to understand the perturbations observed in the CMB and the distribution of the galaxies:

7.3.1 Perturbations in the dominant component, $\lambda > \lambda_J$

With this we mean radiation if $t < t_{eq}$ and matter if $t > t_{eq}$. In this case we can neglect the other fluid and use the single-fluid equation for D only,

$$\ddot{D} + H\dot{D}\left(2 - 3(2w - c_s^2)\right) - \frac{3}{2}H^2 D\left(1 - 6c_s^2 - 3w^2 + 8w\right) = -\left(\frac{k}{a}\right)^2 c_s^2 D, \tag{7.27}$$

and we remember that $H = (\dot{a}/a), w = p/\rho$ and $c_s^2 = \delta\rho/\delta p$. We start by changing the independent variable from t to a. Since

$$\frac{d}{dt} = \dot{a}\frac{d}{da} = H(a)a\frac{d}{da} \tag{7.28}$$

and

$$\dot{H} = -\frac{3}{2}(1+w)H^2 \tag{7.29}$$

the perturbation equation becomes

$$a^2 \frac{d^2 D}{da^2} + Aa\frac{dD}{da} + \left(B + \frac{k^2 c_s^2}{H^2 a^2}\right) D = 0 \tag{7.30}$$

where

$$A = \frac{3}{2}(1 - 5w + 2c_s^2)\,, \quad B = -\frac{3}{2}(1 - 6c_s^2 + 8w - 3w^2) \tag{7.31}$$

During the radiation dominated phase $w = c_s^2 = 1/3$, giving $A = 0, B = -2$ and the equation reduces to

$$a^2 \frac{d^2 D}{da^2} + \left(\frac{k^2}{3H^2 a^2} - 2\right) D = 0 \tag{7.32}$$

During matter domination, $w = c_s^2 = 0$, leading to $A = 3/2$ and $B = -3/2$. However we will keep c_s^2 in the the last term in order to discuss the effect of pressure support. The equation then is

$$a^2 \frac{d^2 D}{da^2} + \frac{3}{2}a\frac{dD}{da} + \left(\frac{k^2 c_s^2}{H^2 a^2} - \frac{3}{2}\right) D = 0 \tag{7.33}$$

As we are looking at modes with a wavelength larger than the Jeans length, we have that $k^2 c_s^2 < H^2 a^2$ and can neglect this term. The equations then become approximately

$$a^2 D'' - 2D \approx 0 \quad \text{(radiation dom.)} \tag{7.34}$$

$$a^2 D'' + \frac{3}{2}aD' - \frac{3}{2}D \approx 0 \quad \text{(matter dom.)} \tag{7.35}$$

Using an ansatz $D \propto a^n$ we find that in the first case $n(n-1) - 2 = 0$, leading to $n = 2$ (the growing mode) and $n = -1$ (the decaying mode). We are only interested in the growing mode, $D \propto a^2$. For the matter dominated case we find that $n = 1$ and $n = -3/2$. The growing modes are therefore

$$D \propto \begin{cases} a^2 & \text{(radiation dominated phase)} \\ a & \text{(matter dominated phase)} \end{cases} \tag{7.36}$$

Firstly, notice that the condition $(k^2 c_s^2 / H^2 a^2) \ll 1$ has a different meaning in the radiation dominated phase and in the matter dominated phase. In the radiation dominated phase, this condition selects the modes which are bigger than the *Hubble* radius, while in the matter dominated case it selects the modes are bigger than the *Jeans length*. Thus the Hubble radius plays the role of the Jeans length in the radiation dominated phase, which should not come as a surprise, since photons move with the speed of light.

Secondly, the perturbations grow only as a power law of the scale factor (and therefore of time), even during matter domination. This is an important difference to the case of a static universe where they grow exponentially fast. If we had not already discussed the creation of initial fluctuation by inflation, we would now realise that we need such a mechanism to explain the COBE measurements of fluctuations of the size 10^{-5} at last scattering.

7.3.2 Perturbations in the relativistic component, $\lambda < d_H$, $t < t_{eq}$

Radiation is dominant, so we can use the equations above again, this time neglecting the factor of 2 in the second term,

$$\frac{d^2 D}{da^2} + \frac{k^2}{3H^2a^4}D \approx 0. \tag{7.37}$$

Notice that $H^2 \propto \rho \propto a^{-4}$ during the radiation dominated phase, making the coefficient of D a constant. The solution is therefore just a plane wave,

$$D = \exp\left(i\frac{k}{\sqrt{3}Ha^2}a\right) = \exp\left(i\frac{d_H(a)}{\lambda_{\rm phys}(a)}\right) \tag{7.38}$$

These oscillations (which persist also after t_{eq}) are visible on sub-horizon scales in the cosmic microwave background as the so-called *acoustic oscillations*.

The solution of (7.32) behaves very differently in the two limiting cases. In fact, we can solve it exactly and show the transition in the behaviour of the solution. Writing $D = aQ$ the equation becomes

$$\frac{d^2 Q}{da^2} + \frac{2}{a}\frac{dQ}{da} + \left[\mu^2 - \frac{2}{a^2}\right]Q = 0 \tag{7.39}$$

where $\mu = k^2c_s^2/(H_{eq}^2 a_e q^4)$ and where we have used the fact that $H^2a^2 = H_{eq}^2(a_{eq}^4/a^2)$. The general solution of this equation is

$$Q = Cj_1(\mu a) + Dy_1(\mu a) \tag{7.40}$$

where j_1 and n_1 are spherical Bessel functions. The transition from oscillatory behaviour (for $\lambda \ll d_H$) to power law growth (for $\lambda \gg d_H$) can then be seen from the behaviour of the Bessel functions. For large μa the Bessel functions oscillate with the amplitude decreasing as $1/a$, hence $D = Qa$ behaves as an oscillatory function with constant amplitude. For small μa the asymptotic form of the Bessel functions leads to $Q \approx c_1 a + c_2/a^2$, and therefore

$$D \approx c_1 a^2 + c_2/a, \tag{7.41}$$

with the growing mode scaling as a^2. The parameter $\mu^2a^2 = k^2c_s^2/(H^2a^2) \sim (d_H^2/\lambda^2)$ is essentially the square of the ratio between the Hubble radius and the proper wavelength of the perturbation. The nature of the solution - whether it grows or oscillates - depends on this ratio.

We should expect these oscillations to be visible in the cosmic microwave background radiation, which we will discuss in the last chapter.

7.3.3 Dark matter perturbations, $t < t_{eq}$

This is a more complicated case as the dark matter is sub-dominant during this epoch. The evolution of its perturbations can be affected, and indeed is affected, by both the different evolution of the background universe and by the presence of the fluctuation in the dominant radiation fluid. In general we need to consider the perturbations in both fluids separately,

$$\dot{D}_m + 3\delta H_m = 0 \tag{7.42}$$
$$\delta\dot{H}_m + 2H\delta H_m = -\frac{1}{3}\nabla^2\Psi \tag{7.43}$$
$$\dot{D}_r - HD_r + 4\delta H_r = 0 \tag{7.44}$$
$$\delta\dot{H}_r + 2H\delta H_r - \left(\frac{k}{a}\right)^2 \frac{D_r}{12} = -\frac{1}{3}\nabla^2\Psi \tag{7.45}$$
$$\nabla^2\Psi = 4\pi G\left(\rho_m D_m + \rho_r D_r\right). \tag{7.46}$$

We are going to assume that the radiation fluid behaves as if no dark matter was present. This will allow us to find the overall behaviour of the dark matter perturbations. The background Hubble parameter H is now determined by the radiation dominated evolution of the universe. The metric perturbation Ψ provides the coupling between the radiation and matter perturbations. As the perturbations in the radiation behave radically different for $\lambda > d_H$ and for $\lambda < d_H$ we need to discuss these two regimes separately *even though* the Jeans length of the dark matter is not related to d_H.

super-horizon perturbations, $\lambda > d_H$

In general the discussion is pretty involved and depends on the relative initial conditions between the dark matter and dark energy. However, for the case of interest to us we can derive the leading order behaviour of the dark matter perturbations as follows: We assume that the initial relative fluctuations in both the dark matter and the radiation are comparable. This is sensible as they are both due to the inflaton perturbations. Then, as $\rho_r \gg \rho_m$ during radiation domination, we can treat the radiation perturbation separately and just use the result of our previous discussion, $D_r \propto a^2$. The equation for the matter perturbations then is approximately

$$\ddot{D}_m + 2H\dot{D}_m = 4\pi G\rho_r D_r. \tag{7.47}$$

During radiation domination $a \propto t^{1/2}$, $H = 1/(2t)$ and $4\pi G\rho_r \approx 3/2H^2 \propto 1/t^2$. The leading order behaviour in powers of t is therefore

$$\ddot{D}_m + \frac{1}{t}\dot{D}_m \propto t^{-1}, \tag{7.48}$$

and clearly $D_m \propto t \propto a^2 \propto D_r$. The perturbations in the radiation fluid "drag" the perturbations in the matter fluid along.

sub-horizon perturbations, $\lambda < d_H$

What would happen if there were no perturbations in the radiation? An interesting question in itself, it is also relevant as on small scales $\langle D_r\rangle \approx 0$ due to the rapid oscillations which we have encountered earlier. We can therefore just ignore D_r on sub-horizon sizes – but of course the evolution of H is still governed by the radiation dominated universe. Therefore the perturbation equation becomes

$$\ddot{D}_m + \frac{2\dot{a}}{a}\dot{D}_m = 4\pi G\rho_m D_m, \tag{7.49}$$

where the background is governed by the equation

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}(\rho_r + \rho_m). \tag{7.50}$$

Introducing the variable $x \equiv a/a_{eq}$ and combining the two equations we can recast them into the form

$$2x(1+x)\frac{d^2 D_m}{dx^2} + (2+3x)\frac{dD_m}{dx} = 3D_m \tag{7.51}$$

One solution to this equation is obvious,

$$D_m = 1 + \frac{3}{2}x. \tag{7.52}$$

In other words, $D_m \approx$ constant for $a \ll a_{eq}$ (hence no growth in the radiation dominated phase) and $D_m \propto a$ (growth proportional to the scale factor) during the matter dominated epoch. Thus D_m does not grow significantly during the radiation dominated phase even though $\lambda > \lambda_J$. Indeed, we see that the growth factor is

$$F = \frac{1 + (3/2)(a/a_{eq})}{1 + (3/2)(a_{\rm enter}/a_{eq})}. \tag{7.53}$$

Assuming $a_{\rm enter} \ll a_{eq}$ and $a \approx a_{eq}$, we find that the maximal growth factor until equality is about $5/2$.

This statement however needs to be examined in more detail. All differential equations used so far are second order. Hence they have two linearly independent solutions, and the general solution is only determined by choosing two initial conditions. So far we have evaded this problem by considering only the growing solution as *the* solution to the equation. This solution is certainly justified when we are not interested in transitory phenomena. However in our current case, we need to consider both solutions, as this will change the result slightly.

Given the growing solution, we can easily find the decaying solution D by the Wronskian condition $Q'/Q = -(2+3x)/[2x(1+x)]$ where $Q = D_m D' - D_m' D$. Rewriting the decaying solution as $D(x) = f(x)D_m(x)$ we find a differential equation for f,

$$\frac{f''}{f} = -2\frac{D_m'}{D_m} - \frac{2+3x}{2x(1+x)}. \tag{7.54}$$

This can be solved to give

$$f(x) = -\int \frac{dx}{x(1+3x/2)^2(1+x)^{1/2}}. \tag{7.55}$$

Integrating the function, we find the decaying solution to be,

$$D = fD_m = \left(1 + \frac{3x}{2}\right)\ln\left[\frac{(1+x)^{1/2}+1}{(1+x)^{1/2}-1}\right] - 3(1+x)^{1/2} \tag{7.56}$$

Thus the general solution is a linear superposition of the two solutions,

$$D_{\rm gen}(x) = AD_m(x) + BD(x) = \begin{cases} A + B\ln(4/x) & (x \ll 1) \\ (3/2)Ax + (4/5)Bx^{(-3/2)} & (x \gg 1) \end{cases} \tag{7.57}$$

The constants A and B have to fixed by matching this solution to the growing solution which was valid when the mode was bigger than the Hubble radius. Since the latter solution is given by $D(x) = x^2$ in the radiation dominated phase, we get the matching conditions

$$\begin{aligned} x_{\rm enter}^2 &= [AD_m(x) + BD(x)]_{x=x_{\rm enter}} && (7.58)\\ 2x_{\rm enter} &= \left[AD_m'(x) + BD'(x)\right]_{x=x_{\rm enter}} && (7.59)\end{aligned}$$

As an example, consider the a mode for which $x_{\rm enter} \ll 1$. The "decaying solution" has the asymptotic form $D(x) = \ln(4/x)$ for $x \ll 1$. The matching conditions then become,

$$\begin{aligned} x_{\rm enter}^2 &= A\left(1 + \frac{3}{2}x_{\rm enter}\right) + B\ln\frac{4}{x_{\rm enter}}, && (7.60)\\ 2x_{\rm enter} &= \frac{3}{2}A - B/x_{\rm enter}. && (7.61)\end{aligned}$$

Solving these to determine A and B (to the lowest order in $x_{\rm enter}$) we find

$$A = x_{\rm enter}^2\left[1 + 2\ln\left(\frac{4}{x_{\rm enter}}\right)\right] \ ; \quad B = -2x_{\rm enter}^2 \tag{7.62}$$

and the full solution, properly matched, inside the Hubble radius is

$$D(x) = x_{\rm enter}^2\left[1 + 2\ln\left(\frac{4}{x_{\rm enter}}\right)\right]\left(1 + \frac{3x}{2}\right) - 2x_{\rm enter}^2 \ln\frac{4}{x}. \tag{7.63}$$

During the radiation dominated phase, that is until $a \approx a_{eq}$, or $x \approx 1$ this mode can grow by a factor

$$\frac{D(x\approx 1)}{D(x_{\rm enter})} = \frac{1}{x_{\rm enter}^2}D(x=1) = 5\ln\left(\frac{1}{x_{\rm enter}}\right) = 5\ln\left(\frac{a_{eq}}{a_{\rm enter}}\right) = \frac{5}{2}\ln\left(\frac{t_{eq}}{t_{\rm enter}}\right). \tag{7.64}$$

Since the time $t_{\rm enter}$ is fixed for a mode with wavelength λ by the condition $\lambda a_{\rm enter} \sim \lambda t_{\rm enter}^{1/2} = d_H(t_{\rm enter}) \sim t_{\rm enter}$ it follows that $\lambda \propto t_{\rm enter}^{1/2}$. Hence

$$\frac{D_{\rm final}}{D_{\rm enter}} = 5\ln\left(\frac{\lambda_{eq}}{\lambda}\right) \tag{7.65}$$

for a mode with wavelength λ. The growth in the matter dominated phase is therefore logarithmic. Notice that the matching procedure has brought in an amplification factor which depends on the wavelength.

As a second example of coupled density fluctuations consider the fluctuations in the baryon-dark matter system just after decoupling. This evolution is now governed by the equations,

$$\begin{aligned} \ddot{D}_m + \frac{2\dot{a}}{a}\dot{D}_m &= 4\pi G(\rho_m D_m + \rho_B D_B) \approx 4\pi G\rho_m D_m && (7.66)\\ \ddot{D}_B + \frac{2\dot{a}}{a}\dot{D}_B &= 4\pi G(\rho_m D_m + \rho_B D_B) \approx 4\pi G\rho_m D_m. && (7.67)\end{aligned}$$

Here we have used the fact that just after decoupling $\rho_m D_m \gg \rho_B D_B$. From our previous analysis we know that the dark matter fluctuations behave as

$$D_m = \alpha a \tag{7.68}$$

where α is a constant. Substituting this result into the second equation, we can rewrite it as

$$a^{3/2}\frac{d}{da}\left(\frac{1}{a^{1/2}}\frac{dD_B}{da}\right)+2\frac{dD_B}{da}=\frac{3}{2}\alpha, \tag{7.69}$$

where we have also used the relation $a \propto t^{2/3}$. This equation has the growing solution

$$D_B=\alpha(a-a_0)=D_m(a)\left(1-\frac{b}{a}\right) \tag{7.70}$$

where b is some constant. This solution shows that $D_B \to D_m$ for $a \gg b$, even if $D_B \approx 0$ at some $a = b = a_{\rm dec}$ (for example). In other words, baryonic perturbations catch up with dark matter perturbations after the decoupling, just like the radiation perturbations on super-horizon scales during radiation domination enhanced the dark matter perturbations.

7.3.4 Free streaming and Silk damping

In the discussion of the perturbations so far, we have taken the matter content of the universe to be an ideal fluid. This approximation breaks down for wavelengths smaller than a particular critical value. At smaller wavelengths, energy is drained away by certain dissipative processes which we will now discuss.

The physical origin of dissipation is different in baryons and dark matter. In collisionless dark matter, dissipation occurs through a process called 'free streaming', while in baryons it arises due to coupling between radiation and matter.

Let us consider the dark matter component first. If the dark matter is made up of weakly interacting particles, then these particles do not feel each others presence via collisions (unlike an ordinary gas where collisions are significant). Each dark matter particle therefore moves along a geodesic in space-time. Perturbations modify the space-time metric and consequently the geodesic orbits. We have been studying the response of dark matter particles to such perturbations by invoking an 'effective pressure' and treating dark matter as an ideal fluid. Such an approximation is only valid for sufficiently long wavelengths. At small scales, the free geodesic motion of the particles will wipe out any structure, because the particles can freely propagate from an overdense region to an underdense region equalising the densities. The process is called 'free streaming'.

Let $l_{FS}(t)$ be the proper distance which a dark matter particle can travel in time t in the background spacetime, and let $\lambda(t)$ be the proper wavelength of a perturbation at time t. Then all modes for which $\lambda(t) < l_{FS}(t)$ will suffer dissipation due to free streaming. We therefore need to compute this distance. This can be done as follows: The proper distance travelled by a particle in time t can be written as

$$l_{FS}(t)=a(t)\int_0^t \frac{v(t')}{a(t')}dt', \tag{7.71}$$

where the proper velocity is defined through $adL = vdt$. The dark matter particles are taken to be relativistic for $t < t_{nr}$ and have $v \approx 1$. Using $a(t) \propto t^{1/2}$ we find

$$l_{FS}(t<t_{nr})=a\int_0^t \frac{dt}{a_{nr}}\left(\frac{t_{nr}}{t'}\right)^{1/2}=2t\propto a^2. \tag{7.72}$$

For $t_{nr} < t < t_{eq}$, $v \propto 1/a$ and we get

$$l_{FS}(t_{nr} < t < t_{eq}) = \left(\frac{l_{FS}(t_{nr})}{a_{nr}} + \int_{t_{nr}}^{t} \frac{dt'}{a(t')} \frac{a_{nr}}{a(t')}\right) a(t) = \frac{2t_{nr}a}{a_{nr}} \left(1 + \ln \frac{a}{a_{nr}}\right). \tag{7.73}$$

Finally for $t > t_{eq}$ we use $a(t) \propto t^{2/3}$,

$$l_{FS}(t > t_{eq}) = \left[\frac{2t_{nr}}{a_{nr}} \left(1 + \ln \frac{a_{eq}}{a_{nr}}\right) + \frac{3t_{nr}}{a_{nr}} \left(1 - \frac{a_{eq}^{1/2}}{a^{1/2}}\right)\right] a(t). \tag{7.74}$$

Today therefore the free streaming length is

$$\lambda_{FS} \equiv l_{FS}(t_0) = \left(\frac{a_0}{a_{nr}}\right) (2t_{nr}) \left(\frac{5}{2} + \ln \frac{a_{eq}}{a_{nr}}\right). \tag{7.75}$$

Since l_{FS} always grows, all proper wavelengths $\lambda > \lambda_{FS}$ will survive the process of free streaming.

Next we need to identify t_{nr}. We may take this to be the time at which the temperature of the dark matter falls to $T_{DM} \approx m/3$, where m is the mass of the dark matter particles. In order to compute the temperature of the dark matter to the one of the radiation, T, we use that the quantity

$$R^3 \equiv \left(\frac{T_{DM}}{T}\right)^3 = \frac{n_{DM}}{n_\gamma} \tag{7.76}$$

is conserved during the expansion. But since

$$\Omega_{DM} = \left(\frac{m n_{DM}}{\rho_c}\right)_{\text{now}} = \frac{m n_\gamma}{\rho_c} \left(\frac{n_{DM}}{n_\gamma}\right) \approx 30 \left(\frac{m}{1\text{keV}}\right) \left(\frac{n_{DM}}{n_\gamma}\right) h^{-2}, \tag{7.77}$$

we can write

$$R^3 = \left(\frac{\Omega_{DM} h^2}{30}\right) \left(\frac{m}{1\text{keV}}\right)^{-1}. \tag{7.78}$$

Knowing this ratio, all other can quantities can be determined. For example, $(a_{nr}/a0) = (T_0/T)_{nr} = (T_0/T_{DM})_{nr}(T_{DM}/T) = (3T_0/m)R$, etc. We therefore obtain the free streaming length

$$\lambda_{FS} = 40\text{Mpc} \left(\Omega_{DM} h^2\right)^{-1} \left(\frac{T_{DM}}{T_R}\right)^4 = 0.5\text{Mpc} \left(\Omega_{DM} h^2\right)^{1/3} \left(\frac{m}{1\text{keV}}\right)^{-4/3}. \tag{7.79}$$

This result is of crucial importance. It shows that the length scale below which perturbations are wiped out, λ_{FS}, is essentially decided by the mass m (or, equivalently, Ω_{DM}) and the temperature T_{DM} of the dark matter. If, for example, the dark matter is made of neutrinos with a mass $m \approx 30$eV, $(T_{DM}/T_R) = 0.71$ and $\Omega_\nu h^2 \approx (m_\nu/91\text{eV})$ then

$$\lambda_{FS} = 28\text{Mpc} \left(\frac{m_\nu}{30eV}\right)^{-1}. \tag{7.80}$$

Thus in such a universe all small-scale perturbations will be wiped out. This specific case (called *hot dark matter*) is strongly disfavoured by data. In general, a heavier dark matter candidate will let the power survive at smaller scales. A scenario with very heavy dark matter is called a *cold dark matter* model.

epoch	D_r	$D_{\rm DM}$	D_B
$t < t_{\rm enter} < t_{\rm eq}$ $\lambda > d_H$	grows as a^2	grows as a^2	grows as a^2
$t_{\rm enter} < t < t_{\rm eq}$ $\lambda < d_H$	oscillates	grows as $\ln a$	oscillates
$t_{\rm eq} < t < t_{\rm dec}$ $\lambda < d_H$	oscillates	grows as a	oscillates
$t_{\rm dec} < t$ $\lambda < d_H$	oscillates	grows as a	grows as a

Table 7.1: Growth of perturbations

Fluctuations in the baryons are damped as well: At $t \ll t_{dec}$, photons and baryons are very tightly coupled due to Thomson scattering. Clearly, any perturbations on scales smaller than the photon mean free path will be damped by photon free streaming. But on somewhat larger time scales, photons will diffuse out of overdense regions into underdense regions and thereby damp fluctuations in the baryons on these larger scales (by dragging them along). This process is called *Silk damping*. Its length scale (given by the random walk of photons before t_{dec} is

$$l_S \approx 3.5\text{Mpc} \left(\frac{\Omega}{\Omega_B}\right)^{1/2} (\Omega h^2)^{-3/4}. \tag{7.81}$$

Table 7.1 shows the collected behaviour of the perturbations in radiation (D_r), dark matter ($D_{\rm DM}$) and the baryons (D_B).

In the final two chapters, we will apply this theory first to the fluctuations in the dark matter and then to the anisotropies in the cosmic microwave background.

7.4 The power spectrum of the fluctuations in the dark matter

Let $D_\lambda(t)$ denote the amplitude of the dark matter perturbation corresponding to some wavelength λ at a time t. To each λ we can associate a wavenumber $k \propto \lambda^{-1}$ and a mass $M \propto \lambda^3$, accordingly we may label the perturbation as $D_M(t)$ or $D_k(t)$. We are interested in the value of $D_\lambda(t)$ at some $t > t_{dec}$ (and thus also $t > t_{eq}$).

To begin with, note that the process of free streaming will wipe out the perturbations on all scales smaller than λ_{FS}. So we have the first result,

$$D_\lambda(t) \approx 0 \qquad \text{for } \lambda < \lambda_{FS}. \tag{7.82}$$

Consider next the range of wavelengths $\lambda_{FS} < \lambda < \lambda_{eq}$. These modes enter the Hubble radius in the radiation dominated phase. However, as we have seen this means that their growth is suppressed due to the rapid expansion of the universe. Therefore, they do not grow significantly until $t = t_{eq}$, resulting in $D_\lambda(t_{eq}) \approx D_\lambda(t_{\rm enter})$. After matter begins to dominate, the amplitude of these modes grows in proportion to the scale factor a. Thus

$$D_\lambda(t) = D_\lambda(t_{\text{enter}}) \left(\frac{a}{a_{eq}}\right) \quad \text{for } \lambda_{FS} < \lambda < \lambda_{eq} \ , \ a > a_{eq}. \tag{7.83}$$

Then we study the modes with $\lambda_{eq} < \lambda < \lambda_H$ where $\lambda_H \equiv H^{-1}(t)$ is the Hubble radius at time t when we are studying the spectrum. These modes enter the Hubble radius in the matter dominated phase and grow proportional to a afterwards. So,

$$D_\lambda(t) = D_\lambda(t_{\text{enter}}) \left(\frac{a}{a_{\text{enter}}}\right) \quad \text{for } \lambda_{eq} < \lambda < \lambda_H, \tag{7.84}$$

which may be rewritten as

$$D_\lambda(t) = D_\lambda(t_{\text{enter}}) \left(\frac{a_{eq}}{a_{\text{enter}}}\right) \left(\frac{a}{a_{eq}}\right). \tag{7.85}$$

But notice that, since t_{enter} is fixed by the condition $\lambda a_{\text{enter}} \propto t_{\text{enter}} \propto \lambda t_{\text{enter}}^{2/3}$ we have $t_{\text{enter}} \propto \lambda^3$. Furthermore, $(a_{eq}/a_{\text{enter}}) = (t_{eq}/t_{\text{enter}})^{2/3}$, giving

$$\left(\frac{a_{eq}}{a_{\text{enter}}}\right) = \left(\frac{\lambda_{eq}}{\lambda}\right)^2, \tag{7.86}$$

and find overall

$$D_\lambda(t) = D_\lambda(t_{\text{enter}}) \left(\frac{\lambda_{eq}}{\lambda}\right)^2 \left(\frac{a}{a_{eq}}\right). \tag{7.87}$$

Comparing (7.83) with (7.87) we see that a mode which enters the Hubble radius after t_{eq} has its amplitude decreased by a factor of λ^{-2} relative to its original value.

Finally consider the modes with $\lambda > \lambda_H$ which are still outside the Hubble radius at time t and will enter the Hubble radius at some future time $t_{\text{enter}} > t$. During the time interval (t, t_{enter}) they will grow by a factor (a_{enter}/a), thus

$$D_\lambda(t_{\text{enter}}) = D_\lambda(t) \left(\frac{a_{\text{enter}}}{a}\right) \tag{7.88}$$

or

$$D_\lambda(t) = D_\lambda(t_{\text{enter}}) \left(\frac{a}{a_{\text{enter}}}\right) \quad \text{for } \lambda > \lambda_H. \tag{7.89}$$

Thus the behaviour of the modes is the same for the cases $\lambda_{eq} < \lambda < \lambda_H$ and $\lambda_H < \lambda$, i.e. for all $\lambda > \lambda_{eq}$. Combining all these pieces of information, we can state the final result as follows:

$$D_\lambda(t) = \begin{cases} 0 & (\lambda < \lambda_{FS}) \\ D_\lambda(t_{\text{enter}})(a/a_{eq}) & (\lambda_{FS} < \lambda < \lambda_{eq}) \\ D_\lambda(t_{\text{enter}})(a/a_{eq})(\lambda_{eq}/\lambda)^2 & (\lambda_{eq} < \lambda) \end{cases} \tag{7.90}$$

It remains to specify $D_\lambda(t_{\text{enter}})$, or equivalently, $D_\lambda(t_i)$. Let us assume a power-law

$$D_\lambda(t_{\text{enter}}) = A\lambda^\alpha \propto k^{-\alpha}. \tag{7.91}$$

In the inflationary standard model, we have seen that we expect an initial curvature spectrum close to scale invariant, i.e.

$$P_{\mathcal{R}} \propto k^{n-1} \tag{7.92}$$

with $n \approx 1$. The perturbation in the gravitational potential is proportional to the curvature perturbation, $Psi_k \propto \mathcal{R}_k$, and the Poisson equation $4\pi G\rho D = \nabla^2\Psi \propto k^2\Psi$ shows that $\mathcal{P}_D \propto k^4\mathcal{P}_\mathcal{R} \propto k^{n+3}$. Since we defined the spectrum as $P_D \propto k^3|D_k|^2$, we find that

$$|D(k,t)|^2 \propto k^n \tag{7.93}$$

where D is now specified at a *fixed* time. We need to related this to the spectral index of D at the time when a given scale enters the horizon.

There are two regimes which we need to check, radiation domination and matter domination. We have seen that during matter domination $t_{\text{enter}} \propto \lambda^3 \propto k^{-3}$. Furthermore, with $a \propto t^{2/3}$,

$$D_m(k,t) \propto k^{n/2}a \propto k^{n/2}t^{2/3}, \tag{7.94}$$

and therefore

$$D_m(k,t_{\text{enter}}) \propto k^{n/2}k^{-2} \propto k^{-\alpha}. \tag{7.95}$$

During radiation domination a given scale enters the horizon at $t_{\text{enter}} \propto k^{-2}$, the matter perturbations grow on super-horizon scales as $D_m \propto a^2$ and $a \propto t^{1/2}$, which leads to the same relation between the two spectral indices, $n = (4 - 2\alpha)$. The scale invariant spectrum $n = 1$ corresponds therefore to the case $\alpha = 3/2$.

Using the power law ansatz and $\lambda \propto k^{-1}$, we find for the perturbation spectrum

$$k^3|D_k(t)|^2 = \begin{cases} 0 & (k > k_{FS}) \\ k^{(3-2\alpha)}(a/a_{eq})^2 & (k_{FS} > k > k_{eq}) \\ k^{(7-2\alpha)}(a/a_{eq})^2 & (k_{eq} > k) \end{cases} \tag{7.96}$$

At the time of entering the Hubble radius, this quantity has the dependence

$$(k^3|D_k(t)|^2)_{t=t_{\text{enter}}} \propto k^{3-2\alpha}, \tag{7.97}$$

which shows that in the scale invariant case $\alpha = 3/2$ neither small nor large scales dominate.

The exact form of the spectrum, for both hot and cold dark matter scenarios, can be found by numerical integration of the relevant equations. Two fitting functions are: for hot dark matter

$$|D_k|^2 = Ak^{4-2\alpha}\exp\left[-4.61(k/k_{FS})^{3/2}\right], \tag{7.98}$$

where $k_{FS} = 0.16\text{Mpc}^{-1}(m_\nu/30\text{eV})$ is the free streaming length and A is an unknown amplitude. For cold dark matter

$$|D_k|^2 = \frac{Ak^{(4-2\alpha)}}{(1 + Bk + Ck^{3/2} + Dk^2)^2}, \tag{7.99}$$

with $B = 1.7(\Omega h^2)^{-1}\text{Mpc}$, $C = 9\text{Mpc}^{3/2}(\Omega h^2)^{-3/2}$ and $D = 1\text{Mpc}^2(\Omega h^2)^{-2}$. A is again the unknown amplitude. It is easy to verify that these functions have the correct form for small and large k.

Figure 7.1 shows a plot of $P(k) \equiv |D_k|^2$ for a cold dark matter model, compared to recent data. We can see very well how for small k, $P(k) \propto k$ and for large k, $P(k) \propto k^{-3}$ as predicted by eq. (7.96). There is no free streaming cutoff visible in the data, and we can conclude that hot dark matter is sub-dominant.

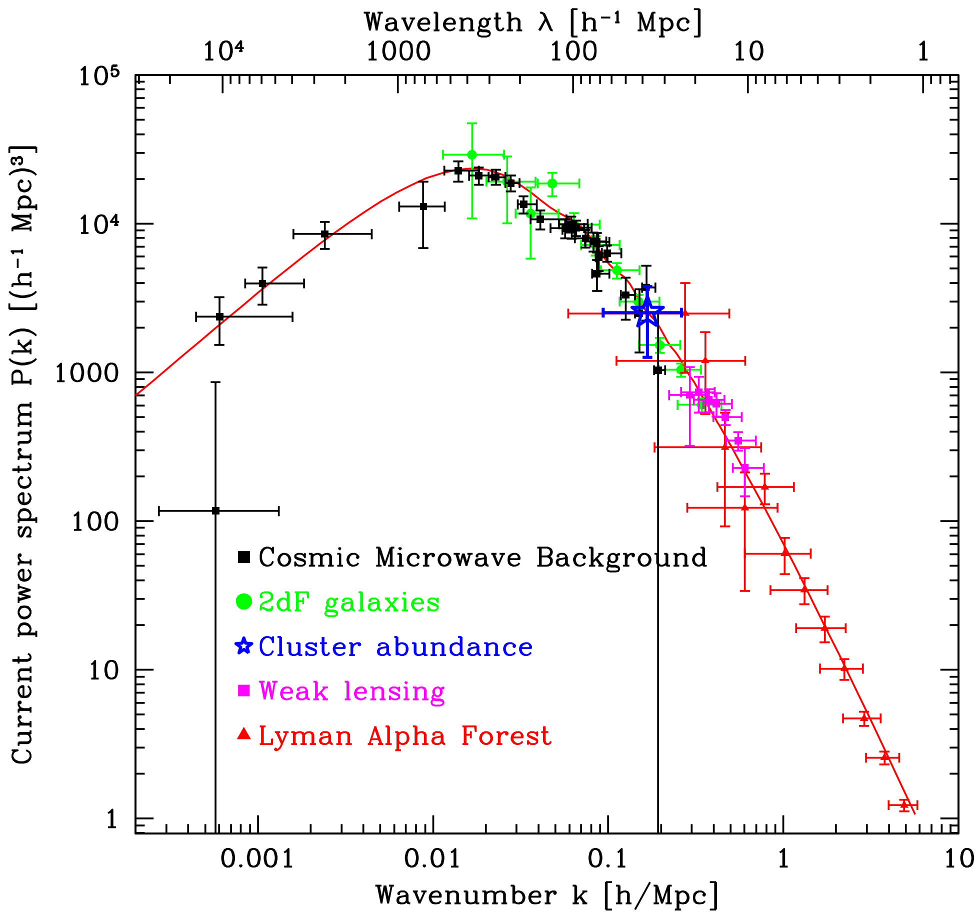

Figure 7.1: The predicted power spectrum for a cold dark matter model compared to current data. From Tegmark & Zaldarriaga, `astro-ph/0207047`.

Chapter 8

The cosmic microwave background

In the first part of these lectures we saw that the cosmic microwave background radiation is basically a photograph of the radiation fluid at decoupling. The inflationary perturbations were of course already present at that time, and we should see their imprint. For two reasons the CMB is a particularly powerful tool to study the perturbations. Firstly, as we see the fluctuations at $t_{\rm dec} \ll t_0$ they are still very small and can be described very well by linear perturbation theory. Secondly, we observed directly the photons, while we cannot observe the dark matter directly. Not surprisingly the CMB is the most direct test of our cosmological model, and leads to the strongest limits on the cosmological parameters.

8.1 The anisotropies in the CMB

What effects do we expect to see in the CMB? For a detailed calculation, we would need to follow the perturbations with the Boltzmann equation. But the final result is easy to understand intuitively. As $\rho_\gamma \propto T^4$, we have that $\delta T/T \propto 1/4 D_\gamma$. This is called the *intrinsic* temperature perturbation. Additionally, there are two possible sources of redshifts, the Doppler shift and the gravitational redshift.

If a photon scatters off an electron which is moving with a velocity v_j with respect to the observer, then its temperature will be shifted by $v_j n^j$ where n^j is a unit vector along the line of sight. This will lead to a measurable contribution if there is a bulk motion of the baryons, $v^{(b)}$.

The photons will pick up a gravitational redshift if there is a difference in the gravitational potential at decoupling and today. (We normalise the gravitational potentials so that there is no perturbation today.) Additionally, if there is a time-varying gravitational potential perturbation along the photon path, this will also lead to a gravitational red- (or blue-) shift. If the perturbation is constant in time, then the photons will pick up identical red- and blueshifts upon entering and leaving the perturbations, so that the net-shift vanishes.

Combining these contributions, we conclude that the temperature perturbation in the CMB is of the form

$$\frac{\delta T}{T} = \left(\frac{1}{4} D_\gamma + v_j^{(b)} n^j + (\Psi - \Phi)\right)(t_{\rm dec}, x_{\rm dec}) + \int_{t_{\rm dec}}^{t_0} \left(\dot{\Psi} - \dot{\Phi}\right)(t, x(t)) dt \tag{8.1}$$

where the integral is understood to be taken along the photon path $x(t)$. General relativity allows us to place a different perturbations in the spatial and temporal part of the metric, hence the presence of two gravitational potentials, Φ and Ψ. In a Newtonian approximation, we find that $\Psi \approx -\Phi$.

To this we have to add the "Silk damping" discussed earlier, which is due to photon free-streaming out of the overdense regions. This will damp exponentially all temperature perturbations on small scales.

Let us find some rough estimates of the different contributions. Going back to the discussion about the evolution of perturbations in a radiation fluid, we find that their amplitude oscillates on scales smaller than the horizon, see eq. (7.38):

$$D_\gamma \propto e^{i\frac{k}{\sqrt{3}Ha^2}a} \propto \cos\left(\frac{k}{k_0}a\right). \tag{8.2}$$

Before decoupling, the baryons and the photons are tightly coupled. As an estimate of $v^{(b)}$ we assume that the baryons are dragged along by the photons, so that they follow the bulk movement of the photon perturbation, $\dot{D}_\gamma$. We see that the Doppler shift is out of phase ($\sim \sin(k/k_0 a)$) compared to the intrinsic perturbation. A detailed calculation also shows that it is subdominant.

We shouldn't forget at this point that we observe the power spectrum of the fluctuations, which corresponds to $(\delta T/T)^2$, so that the oscillations behave like $\cos^2$ and $\sin^2$. The Doppler contribution mainly ensures that the minima do not reach down to zero.

An additional interesting point is that the presence of baryons makes the gravitation wells deeper. This leads to a difference in the amplitude of the compression and rarefaction peaks of the density oscillations. By comparing the amplitudes it is possible to estimate the amount of baryons in the universe. The results are consistent with those from big-bang nucleosynthesis.

The integrated part of the gravitational redshift (also called *integrated Sachs-Wolfe effect*, ISW) is usually subdominant for the inflationary models considered here, except at very late time when the accelerated expansion due to the dark energy leads to the decay of the gravitational perturbations. As this happens at late times, it only affects large scales. The contributions from the potential difference at decoupling also affects predominantly large scales. This can be seen by looking at the Poisson equation, $\nabla^2\Psi = 4\pi G\delta\rho$. As $\nabla^2 \sim k^2$ in Fourier space and as $\delta\rho = \rho D$, we find that it scales like $1/k^2$ and is largest for small k, i.e. on large scales.

8.2 Power spectra

We therefore expect to see a sequence of oscillations that are damped towards small scales and which have an extra contribution on large scales. Although we don't directly observe the three- dimensional power spectrum, but its projection onto the sky sphere, the structure is largely conserved. The main difference is that the wavelength is translated into an angle on the sky.

To relate the observed angular power spectrum and the power spectrum in Fourier space, we expand the sky temperature into spherical harmonics,

$$\frac{\delta T}{T}(x_0, n) = \sum_{\ell,m} a_{\ell m}(x_0) Y_{\ell m}(n). \tag{8.3}$$

Here x_0 is the position of the observer in the universe, and n the direction in which we observe the temperature. In the expansion, ℓ is related to the angular size ϑ by approximately $\vartheta \sim \pi/\ell$. $\ell = 0$ is the average temperature, also called the *monopole*, $\ell = 1$ is called the *dipole*, $\ell = 2$ the *quadrupole*, and so on. m is related to the angular orientation of a given configuration and runs from $-\ell$ to ℓ.

The convention is to plot the power spectrum of the angular fluctuations. If the fluctuations are created by a (statistically) homogeneous and isotropic random process, then the (averaged) result depends neither on x_0 nor on m and there is no coupling between different scales and orientations. We write

$$\langle a_{\ell m} a_{\ell' m'} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_\ell. \tag{8.4}$$

It is easy to show that the C_ℓ are related to the temperature correlations through

$$\left\langle \frac{\delta T}{T}(n) \frac{\delta T}{T}(n') \right\rangle = \frac{1}{4\pi} \sum_\ell (2\ell+1) C_\ell P_\ell(\mu), \tag{8.5}$$

where $\mu = \cos(n \cdot n')$. As expected, this does not depend on the absolute orientation, only on the angle between the two directions that we observe the CMB in.

It remains to relate the real space correlation above to the fluctuations in Fourier space. We are going to compute this for a simple but relevant example. For adiabatic initial conditions, the dominant contribution on super horizon scales (neglecting the integrated Sachs Wolfe effect $\int(\dot{\Phi} - \dot{\Psi})$) is

$$\frac{\delta T}{T}(\mathbf{x}_0, n, \eta_0) = \frac{1}{3} \Psi(x_{\rm dec}, \eta_{\rm dec}). \tag{8.6}$$

Here η is the conformal time, defined by $dt^2 = a^2 d\eta^2$.

The Fourier transform of (8.6) gives

$$\frac{\delta T}{T}(\mathbf{k}, n, \eta_0) = \frac{1}{3} \Psi(k, \eta_{\rm dec}) \cdot e^{i\mathbf{k}n(\eta_0 - \eta_{\rm dec})} \;. \tag{8.7}$$

Using the mathematical identity

$$e^{i\mathbf{k}n(\eta_0 - \eta_{\rm dec})} = \sum_{\ell=0}^{\infty} (2\ell+1) i^\ell j_\ell(k(\eta_0 - \eta_{\rm dec})) P_\ell(\widehat{\mathbf{k}}n) \;,$$

and replacing the ensemble average $\langle\rangle$ by a spatial average over all observer positions x_0 (homogeneity!) we obtain

$$\begin{aligned}
\left\langle \frac{\delta T}{T}(\mathbf{x}_0, n, \eta_0) \frac{\delta T}{T}(\mathbf{x}_0, n', \eta_0) \right\rangle &= \frac{1}{V} \int d^3x_0 \frac{\delta T}{T}(\mathbf{x}_0, n, \eta_0) \frac{\delta T}{T}(\mathbf{x}_0, n', \eta_0) \\
&= \frac{1}{(2\pi)^3} \int d^3k \frac{\delta T}{T}(\mathbf{k}, n, \eta_0) \frac{\delta T}{T}(\mathbf{k}, n', \eta_0)
\end{aligned}$$

$$= \frac{1}{(2\pi)^3 9} \int d^3k \, |\Psi|^2 \sum_{\ell,\ell'=0}^{\infty} (2\ell+1)(2\ell'+1) j_\ell(k(\eta_0 - \eta_{\rm dec})) j_{\ell'}(k(\eta_0 - \eta_{\rm dec})) i^{\ell-\ell'} P_\ell(\hat{\mathbf{k}}n) P_{\ell'}(\hat{\mathbf{k}}n').$$

The Legendre polynomials can be decomposed into a sum of spherical harmonics, $P_\ell(\hat{\mathbf{k}}n) = \frac{4\pi}{2\ell+1} \sum_m Y_{\ell m}(\hat{\mathbf{k}}) Y^*_{\ell m}(n)$. These obey orthonormality conditions, so that we integrate over all directions of the k vector, $d\Omega_{\hat{k}}$. As $Y(\hat{k})$

appears twice, it drops out and we pick up a factor $\delta_{\ell\ell'}\delta_{mm'}$ instead, allowing us to perform those sums. Using as well $\sum_m Y^*_{\ell m}(n)Y_{\ell m}(n') = \frac{2\ell+1}{4\pi}P_\ell(\mu)$, where $\mu = n \cdot n'$, we find finally

$$\left\langle \frac{\delta T}{T}(\mathbf{x}_0, n, \eta_0)\frac{\delta T}{T}(\mathbf{x}_0, n', \eta_0)\right\rangle_{nn'=\mu} = \sum_\ell \frac{2\ell+1}{4\pi}P_\ell(\mu)\frac{2}{\pi}\int\frac{dk}{k}\left\langle\frac{1}{9}|\Psi|^2\right\rangle k^3 j_\ell^2(k(\eta_0-\eta_{\rm dec})). \quad (8.8)$$

Comparing this equation with (8.5), we obtain for *adiabatic perturbations* on scales $\ell \ll \chi(\eta_0 - \eta_{\rm dec})/\eta_{\rm dec} \sim 100$

$$C_\ell^{(SW)} \simeq C_\ell^{(OSW)} \simeq \frac{2}{\pi}\int_0^\infty \frac{dk}{k}\left\langle\left|\frac{1}{3}\Psi\right|^2\right\rangle k^3 j_\ell^2\left(k\left(\eta_0-\eta_{\rm dec}\right)\right). \quad (8.9)$$

If Ψ is a pure power law, the integral (8.9) can be performed analytically. For the ansatz

$$\left\langle|\Psi|^2\right\rangle k^3 = Ak^{n-1}\eta_0^{n-1}. \quad (8.10)$$

one obtains

$$C_\ell^{(SW)} = \frac{A}{9}\frac{\Gamma(3-n)\Gamma(\ell-\frac{1}{2}+\frac{n}{2})}{2^{3-n}\Gamma^2(2-\frac{n}{2})\Gamma(\ell+\frac{5}{2}-\frac{n}{2})}. \quad (8.11)$$

Of special interest is the *scale invariant* spectrum, $n = 1$. This spectrum with a time and scale independent gravitational potential has first been investigated by Harrison and by Zel'dovich. It is therefore called the Harrison-Zel'dovich (HZ) spectrum and leads to

$$\ell(\ell+1)C_\ell^{(SW)} = \text{const.} \simeq \left\langle\left(\frac{\delta T}{T}(\vartheta_\ell)\right)^2\right\rangle, \quad (8.12)$$

which is just the behaviour observed by the DMR experiment aboard COBE. This also shows why we plot $\ell(\ell+1)C_\ell$. Indeed, this is the power per logarithmic interval in scale, and it shows that the HZ spectrum has the same amount of power on large and small scales.

On smaller scales, $\ell \gtrsim 100$, the contribution to $\delta T/T$ is dominated by acoustic oscillations, the first two terms in eq. (8.1). Instead of (8.9) we then obtain

$$C_\ell^{(AC)} \simeq \frac{2}{\pi}\int_0^\infty\frac{dk}{k}\left[k^3\left\langle\left|\frac{1}{4}D_\gamma(k,\eta_{\rm dec})j_\ell(k(\eta_0-\eta_{\rm dec})) + v^{(b)}(k,\eta_{\rm dec})j'_\ell(k,\eta_0-\eta_{\rm dec})\right|^2\right\rangle\right]. \quad (8.13)$$

D_γ and $v^{(b)}$ are oscillating like sine or cosine waves on sub-horizon scales, but they are out of phase with respect to each other: Whenever the density contrast is extremal, the fluid is maximally compressed or decompressed and the oscillation is about turn around and so the fluid doesn't move. A more detailed calculation shows that the Doppler contribution is smaller than the intrinsic one. It therefore sets the minima of the oscillations visible in the power spectrum (which would reach down to 0 otherwise).

The location of the first peak is also of great importance. Remembering that D_γ grows as a power law outside the horizon (and hence has *no* maximum on very large scales), the first peak is at $k/(Ha) = 2\pi\sqrt{3}$, corresponding to the size of the sound horizon at last scattering. This is the largest perturbation that had enough time to collapse to maximal density, as perturbations evolve at the speed of sound, $1/\sqrt{3}$. Its true size is therefore known. That means that we can use it as a standard ruler. The angular size under which we observe it on the sky will depend on the geometry of the universe. In an open universe, the photon trajectories will diverge and we will see it under a smaller angle (at higher ℓ), while it will appear at smaller ℓ in a closed universe. The CMB allows us therefore to probe the overall curvature of space, although there is a caveat that a combination of other parameters can lead to a similar shift in the peak positions. But combined with other data sets, we find that the universe is very close to flat.

The Bessel functions j_ℓ oscillate themselves, the resulting superposition smears out the result somewhat. But the main contribution for each ℓ and $x = kt$ of $j_\ell(x)$ comes from $x \approx \ell$. Correspondingly the $C_\ell^{(AC)}$ will show peaks and minima.

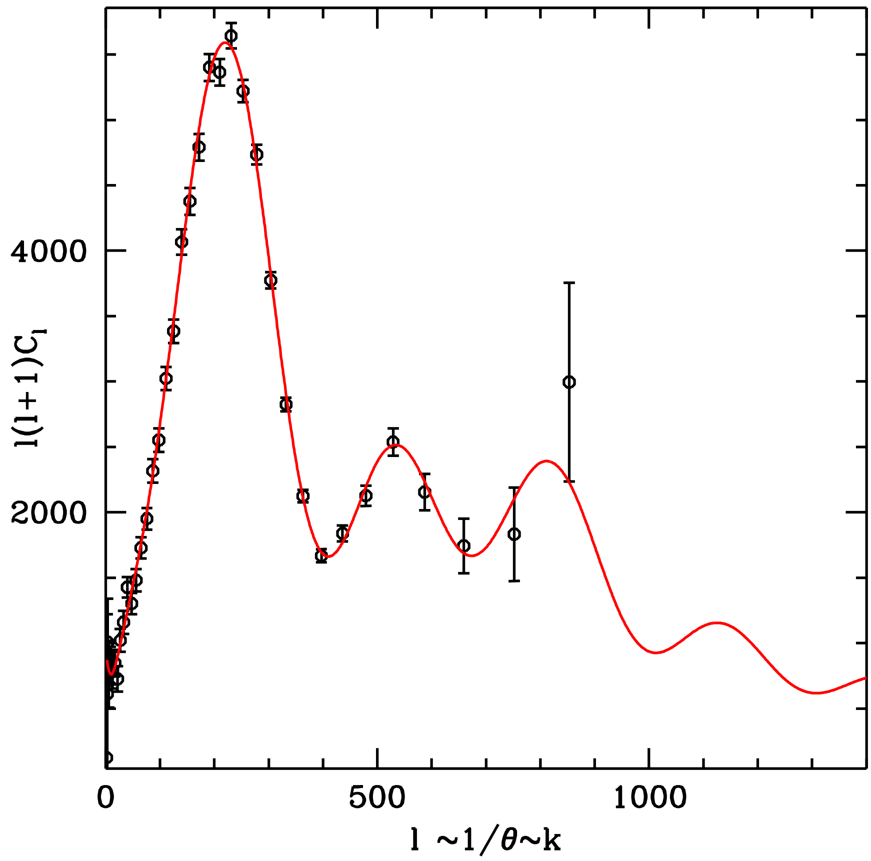

Figure 8.1: The angular power spectrum of the temperature anisotropies in the CMB for the current best-fit model (red line) and the WMAP data (in black). The curve clearly shows the features discussed in the text.

8.3 Dependence on other cosmological quantities

As we have seen, the wavelength and therefore the angular power spectrum of the temperature anisotropies depends in many ways on the cosmological model. We are going to quickly recapitulate the most important examples.

8.3.1 Curvature

The presence of a non-zero curvature determines if initially parallel light-rays stay parallel (flat space), diverge (open universe) or converge (closed universe). In this way, it changes the angle under which we see a given scale at last scattering (angular diameter distance), and therefore the position of the first peak. The peaks appears to be at lower ℓ (larger scales) for a closed universe, and at higher ℓ (smaller scales) for an open universe.

Curvature also changes the integrated Sachs-Wolfe contribution, which is visible only on very large scales (very low ℓ).

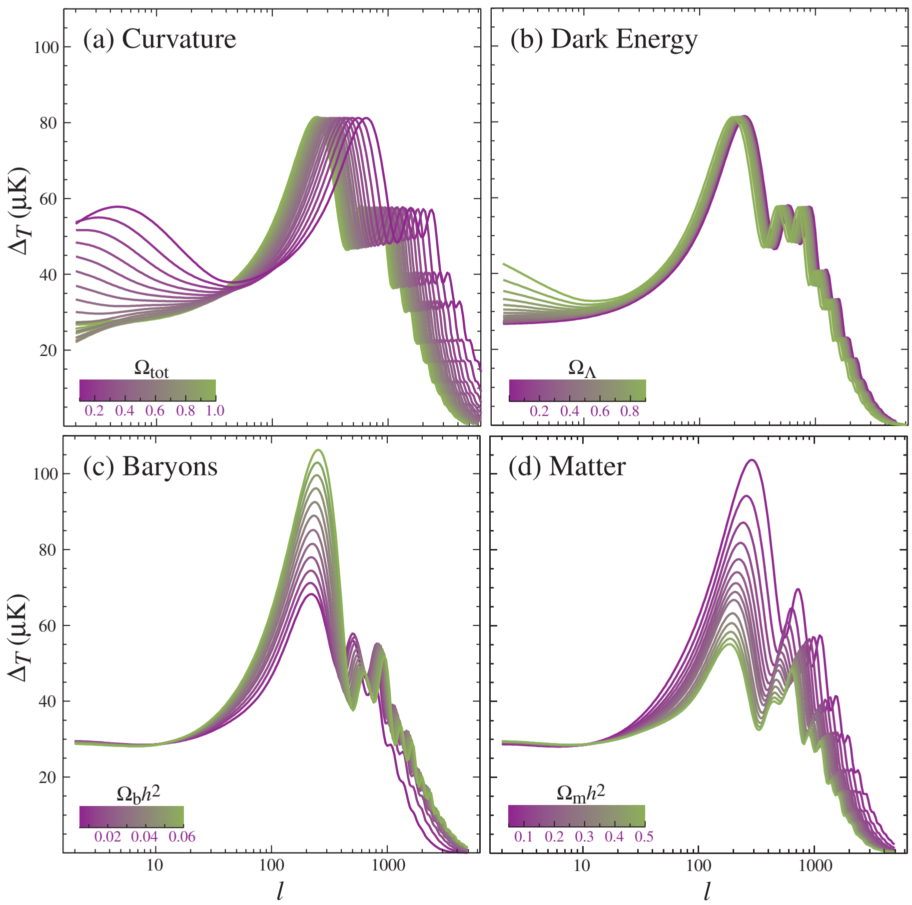

Figure 8.2: Sensitivity of the acoustic temperature spectrum to four fundamental cosmological parameters, (a) curvature, (b) dark energy, (c) baryon density and (d) matter density. The models were varied around a fiducial model with $\Omega_{\rm tot} = 1$, $\Omega_\Lambda = 0.65$, $\Omega_b h^2 = 0.02$, $\Omega_m h^2 = 0.147$, $n_s = 1$ and $\tau = 0$. Figure from Hu & Dodelson.

8.3.2 Baryon density

The presence of baryons decreases the sound velocity, leading to a longer oscillation period. This in turn increases the spacing between the acoustic peaks. But the most distinctive feature is the enhancement of the compression peaks and the decrease of the expansion peaks as mentioned earlier.

8.3.3 Cosmological Constant

For a given total density, the presence of a cosmological constant with $\Omega_\Lambda = \Omega_{\rm tot} - \Omega_m$ means that the matter density is lower and so delays the epoch of equal matter and radiation. As the gravitational potential decays during radiation domination, this leads to an ISW effect, boosting mainly the first acoustic peak. When the cosmological constant starts to dominate, it also leads to a further decay of the gravitational potential and so adds a late (large scale) ISW contribution.

8.3.4 Scalar spectral index

By construction, the differential equations for the perturbation variables are linear. The initial (or *primordial*) power spectrum of the perturbations, $\mathcal{P}_\Psi(k) = k^3|\Psi(k)|^2 \sim k^{n_s-1}$, therefore directly determines their behaviour as a function of k. The spectral index n_s then governs the overall scaling of the perturbations with size. Changing the scalar spectral index induces an overall tilt in the power spectrum. The same is true for the tensor spectral index with respect to the gravity wave power spectrum.

8.3.5 Gravitational waves

Inflation not only perturbs the energy densities and the curvature, but also different parts of the metric that lead to gravity waves. These are also called *tensor perturbations* to differentiate them from *scalar* quantities like the curvature and the density perturbations. The evolution of these different kinds of perturbations decouples in linear perturbation theory, and they can be treated independently.

In the metric, they appear as the transverse, traceless perturbation of the spatial part of the metric: h_{ij} with $h^i_i = 0$ and (in flat space) $k^i h_{ij} = 0$. There are two independent modes, h_+ and $h_\times$. For a wave propagating in the z direction, the perturbation of the spatial metric can be written as

$$\begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{8.14}$$

which satisfies obviously the conditions given above. The physical interpretation is that they represent the two polarisations of a gravitational wave (which is a massless spin 2 wave). They have the same perturbation equation, which turns out to be

$$\ddot{h} + 3H\dot{h} + \left(\frac{k}{a}\right)^2 h = 0. \tag{8.15}$$

During the radiation dominated era $a \propto t^{1/2}$ and the Hubble function is $H = 1/(2t)$. Defining $x = kt/a$ and $' = d/dx$ we find the equation

$$h'' + \frac{2}{x}h' + h = 0 \tag{8.16}$$

the solution of which is proportional to $\sin(x)/x$. On super-horizon scales ($x \ll 1$) the amplitude is conserved. On sub-horizon scales ($x \gg 1$) the energy in the field is proportional to $|h|^2/a^2$, which scales like $1/a^4$. That is no surprise, as the gravitons are massless particles like photons. This also means that we expect a gravitational wave background similar to cosmic microwave background.

Any given inflationary model makes specific predictions about the gravity wave spectrum. Their detection might allow us to probe the physics of inflation. Their effect on CMB can be calculated similarly to the scalar case with the geodesic equation for the photons, the source for the perturbation is proportional to $\dot{h}$ integrated over the photon path, somewhat akin to the integrated Sachs-Wolfe effect. They only affect the low-ℓ region of the CMB power spectrum, as they decay when they enter the horizon.

8.4 Observational results

8.4.1 CMB experiments

The current experiments can be classified into three different categories: Satellite experiments, balloon-borne experiments and ground based experiments.

The ground based experiments have the advantage that they can observe a given region on the sky for a very long time, and that there are few size and weight constraints. They are also relatively cheap, especially compared to satellite experiments. Their main drawback is the presence of atmospheric fluctuations, which strongly limits their fundamental sensitivity. Examples of ground-based CMB experiments are Python, White Dish, DASI, CAT, VSA.

For this reason, several groups have opted to use balloons to lift their instruments above most of the atmosphere. The main problems here are weight and integration time restrictions, also any issues that occur during a flight cannot be corrected. Examples of balloon-borne CMB experiments are MAXIMA, BOOMERanG and TopHat. Arguably, MAXIMA and BOOMERanG provided the first clear measurement of the first acoustic peak.

A very expensive, but ultimately by far the most powerful solution is to put the detectors into space. COBE was the first CMB satellite mission, and it both confirmed the black body spectrum nature of the CMB and measured for the first time the anisotropies (on very large scales). The second mission, WMAP, has increased dramatically the precision of the CMB data after its first year already. This satellite is positioned at the second Lagrange point, L2, where it can keep both the sun and the earth behind its "back". Also the third planned mission, Planck, will go to L2. Its launch is scheduled for 2007, and it should measure the C_ℓ out to $\ell = 1500$ with a good precision for both temperature and E type polarisation, as well as try to detect B type polarisation from gravitational waves. It also measures the sky at more frequencies, which should allow a better subtraction of foreground emissions. We plot in figure 8.3 some of the data available before early 2003 (i.e. before WMAP). This should be contrasted with figure 8.1.

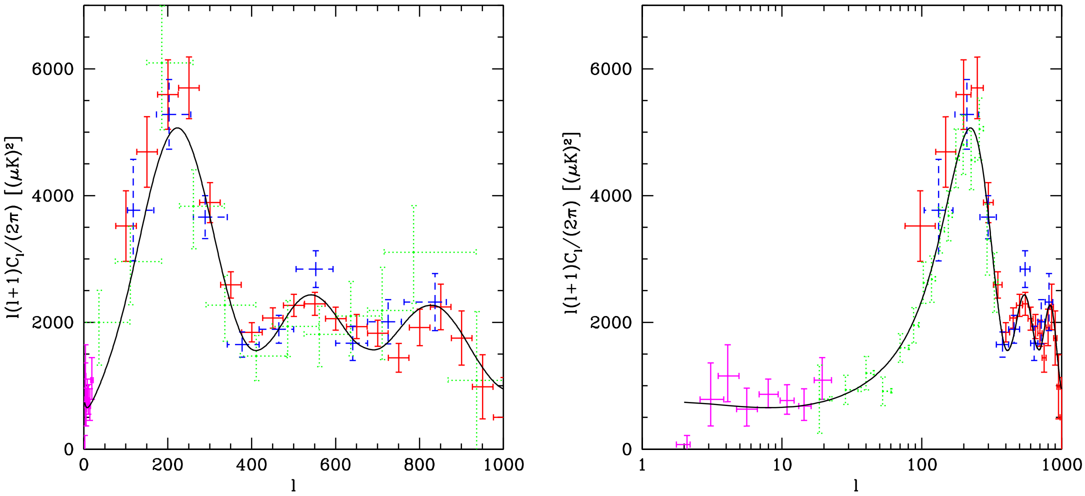

Figure 8.3: The measured CMB temperature anisotropies by the end of 2002. The left figure shows a linear ℓ axis, while the right one shows a logarithmic one to emphasise the low-ℓ region and the COBE data. The theoretical curve is a standard cold dark matter model. The data points shown are COBE/DMR (solid/magenta), Archeops (dotted/green on right hand panel), BOOMERanG (solid/red) and MAXIMA-1 (dotted/green on left hand panel). (figure from review paper by Ruth Durrer)

8.4.2 Cosmic variance

There is a fundamental limitation to how accurately we can measure the C_ℓ: The anisotropies arise from a random process, and so are statistical in nature. Only the average over an infinity of realisations coincides precisely with the theoretical curve. Unfortunately we can only measure the $a_{\ell m}$ of a single universe. As we have seen, the C_ℓ do not depend on m, so we can instead average over m. This corresponds to averaging over directions. For each ℓ, m runs from $-\ell$ to ℓ, so that we have $2\ell+1$ independent estimates of the C_ℓ (assuming full sky coverage, in reality only satellite missions achieve this, and even there we have to cut out the galactic region as it is dominated by foregrounds). The statistical error on each C_ℓ is then

$$\Delta C_\ell = \sqrt{\frac{2}{2\ell+1}} C_\ell. \tag{8.17}$$

This is called *cosmic variance*. No experiment can ever measure the power spectrum to a higher precision. Real experiments tend to do significantly worse, especially at high ℓ, since they cannot resolve the small angular scales, and since the noise tends to be stronger. On the other hand, if we average over adjacent ℓ in bins of a given width, then we can decrease the error bars. But of course we can't increase the statistical power. WMAP has basically reached this maximal sensitivity on large scales, $\ell \lesssim 300$. Planck should extend this to considerably larger ℓ.

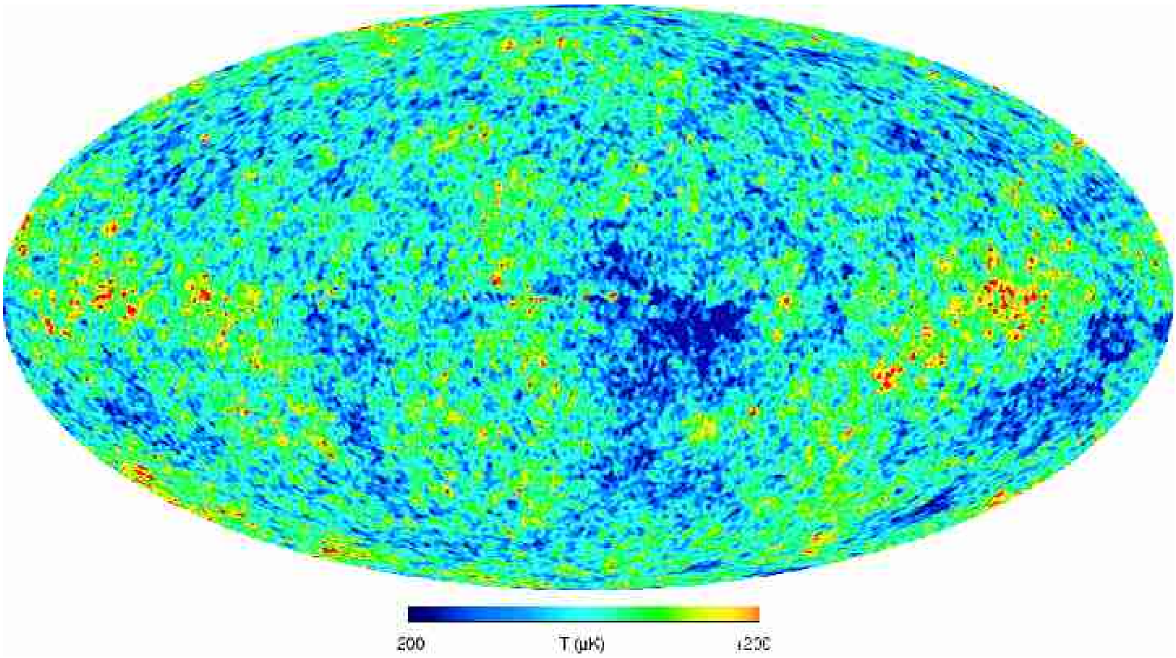

Figure 8.4: The sky according to WMAP. (figure from astro-ph/0302207)

8.4.3 From δT to the cosmological parameters

A measurement generically scans over the sky and returns a so-called *time-ordered data stream*, i.e. the sky temperature (and maybe polarisation) as a function of time. This is connected to the actual temperature map through a *pointing matrix*, which records to which location on the sky the experiment was pointing at a given time. Additionally, the experiment is generally noisy. The first step in data analysis is to recover a map from the time ordered data. This is a hard problem, as it would require inverting a huge matrix as well as estimating the noise properties from the data, and much work has been done on this in recent years. Typically, one uses maximum likelihood methods (mentioned below) and approximate or iterative matrix inversion to solve for the most likely map given the data.

The map typically contains a lot less data than the initial data stream, typically by a factor of several thousand. E.g. for the BOOMERanG ballon experiment, the data stream contained 2×10^8 numbers, while the map only had 57000 pixels. If the experiment measures at different frequencies, then we end up with several maps. These, as well as information from other studies, can then be used to remove foregrounds like point sources and dust emission, as they behave differently from the CMB as a function of frequency. The galaxy is usually just cut out, as it completely covers the signal.

For most applications we are interested in the power spectrum of the temperature fluctuations, the C_ℓ, which results in another dramatic reduction of data. BOOMERanG as an example measured the C_ℓ in some 19 bins from the 57000 pixels. This presents similar problems as the construction of the map from the time ordered data, and again much progress has been made in recent years in order to speed this process up. Apart from maximum likelihood methods one can for example just transform the map into harmonic space and then use many simulations (where we know the map and the C_ℓ) to remove the systematic shift introduced by incomplete sky coverage, and to estimate the error.

Without this strong compression of the measurements (by a factor of about 10^7 in our example) it would be much more difficult to use the CMB as a tool for cosmological statistics. In general, the structure of the experiments and of the CMB itself (see non-Gaussianity in the next section) ensures that we do not lose any critically important parts of the data.

Given both a model of the universe (e.g. encoded in a set of cosmological parameters) and the measured $C_\ell^{(d)}$, we can then first calculate the theoretical power spectrum $C_\ell^{(t)}$ and then compute a goodness of fit, called a *likelihood* function $\mathcal{L}$. The simplest one is the χ^2 estimator, with $\mathcal{L} \approx \exp(-\chi^2/2)$,

$$-2\log\mathcal{L} \approx \chi^2 = \sum_\ell \left(\frac{C_\ell^{(d)} - C_\ell^{(t)}}{\Delta C_\ell}\right)^2 . \tag{8.18}$$

There are better approximations to the true likelihood. Typically, the errors for different ℓ are not independent, so that we need to use a *covariance matrix* instead of the simple, diagonal form given above. WMAP for example distributes a short program that, given $C_\ell^{(t)}$, returns a good approximation to $\mathcal{L}$.

The probability of a given model is then the likelihood for that model times the "prior probability". In the absence of further information, we often take the latter to be constant, so that the probability is directly proportional to the likelihood. In general (as opposed to the map-making and the C_ℓ estimation), the model C_ℓ depend on the cosmological model in a complicated and non-linear way. In many cases we use a program to compute them – some of these programs are publicly available, for example `cmbfast`, `camb` (both written in fortran) or `cmbeasy` (in C++). In this case, we cannot simply invert the likelihood function to compute the best model.

Instead we can a priori compute the likelihood for a large number of models. One possibility is to run a huge grid where we vary the value of each parameter, although todays preferred method is a Monte-Carlo approach using a Markov-Chain to sample the parameter space. The latter method predominantly stays in the most likely region, so that the sampling probability is proportional to the likelihood. This greatly speeds up the exploration of the model space for more than about six parameters and simplifies the subsequent error computation.

We reproduce the results from the first-year WMAP data (astro-ph/0302209) in tables 8.1 and 8.2. These assume a *flat* ΛCDM standard model, containing photons, baryons, three massless neutrinos, dark matter and a cosmological constant. Figure 8.5 shows the limits on the matter density and the dark energy (and therefore the curvature) from different combinations of data.

Whenever we use results derived in this way, we need to be aware of all the assumptions that enter the calculations. At a very basic level we need to accept that General Relativity is the correct theory to describe the universe (except of course if we modify the programs to test this assumption). A very strong assumption is that the power spectrum of the initial perturbations has a very simple form, typically $P(k) \propto k^n$. Of course if we left $P(k)$ to be a free function, we would not be able to recover any results, so that we have to find an acceptable compromise. Sometimes it is possible to use Bayesian statistics to get some handle on which parameters are reasonable, but often we have to rely on our intuition.

Parameter		Mean (68% confidence range)	Maximum Likelihood
Baryon Density	$\Omega_b h^2$	0.024 ± 0.001	0.023
Matter Density	$\Omega_m h^2$	0.14 ± 0.02	0.13
Hubble Constant	h	0.72 ± 0.05	0.68
Amplitude	A	0.9 ± 0.1	0.78
Optical Depth	τ	$0.166^{+0.076}_{-0.071}$	0.10
Spectral Index	n_s	0.99 ± 0.04	0.97
	χ^2_{eff}/ν		1431/1342

Table 8.1: Parameter values from the WMAP CMB data for a ΛCDM model. (taken from astro-ph/0302209)

Parameter	Mean (68% confidence range)
Amplitude of Galaxy Fluctuations	$\sigma_8 = 0.9 \pm 0.1$
Characteristic Amplitude of Velocity Fluctuations	$\sigma_8 \Omega_m^{0.6} = 0.44 \pm 0.10$
Baryon Density/Critical Density	$\Omega_b = 0.047 \pm 0.006$
Matter Density/Critical Density	$\Omega_m = 0.29 \pm 0.07$
Age of the Universe	$t_0 = 13.4 \pm 0.3$ Gyr
Redshift of Reionization	$z_r = 17 \pm 5$
Redshift at Decoupling	$z_{dec} = 1088^{+1}_{-2}$
Age of the Universe at Decoupling	$t_{dec} = 372 \pm 14$ kyr
Thickness of Surface of Last Scatter	$\Delta z_{dec} = 194 \pm 2$
Thickness of Surface of Last Scatter	$\Delta t_{dec} = 115 \pm 5$ kyr
Redshift at Matter/Radiation Equality	$z_{eq} = 3454^{+385}_{-392}$
Sound Horizon at Decoupling	$r_s = 144 \pm 4$ Mpc
Angular Diameter Distance to the Decoupling Surface	$d_A = 13.7 \pm 0.5$ Gpc
Acoustic Angular Scale	$\ell_A = 299 \pm 2$
Current Density of Baryons	$n_b = (2.7 \pm 0.1) \times 10^{-7}$ cm^{-3}
Baryon/Photon Ratio	$\eta = (6.5^{+0.4}_{-0.3}) \times 10^{-10}$

Table 8.2: Derived parameters from the WMAP CMB data for a ΛCDM model. (taken from astro-ph/0302209)

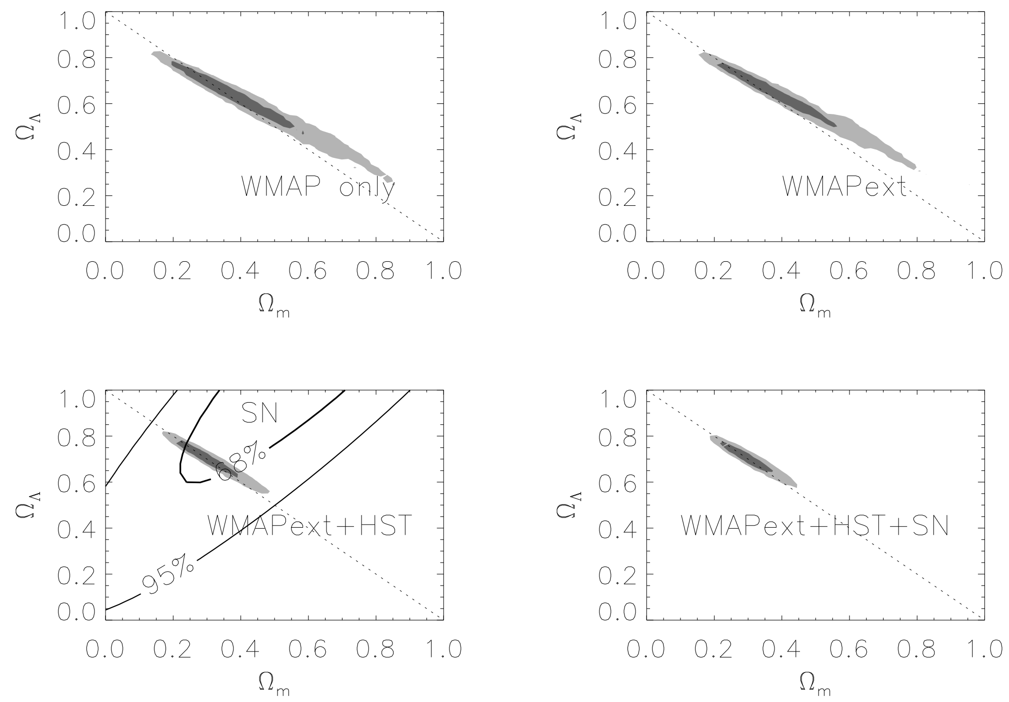

Figure 8.5: Constraints on the matter energy density Ω_m and the dark energy density Ω_Λ for WMAP combined with other data sets. The peak-position degeneracy of the CMB data (top two panels) is clearly visible. The preferred region corresponds to a flat universe with $\Omega_m \approx 0.3$ and $\Omega_\Lambda \approx 0.7$.

8.5 The future of CMB observations

Apart from the primary temperature anisotropies discussed so far, there is a range of further effects that affect the cosmic microwave background radiation. As the data improves, their relevance will increase. We just discuss some of the most important additional observables.

8.5.1 Polarisation

Photons are not scalar particles, they are vector bosons. In addition to an intensity (temperature), they also oscillate in a given transverse direction. This is called polarisation. If all incoming photons have a random and uncorrelated polarisation, then the CMB would be called unpolarised. But this is not the case, as Thomson scattering is not isotropic. For a photon with incoming polarisation vector ϵ_1 and outgoing polarisation vector ϵ_2, the cross section is proportional to $d\sigma_T/d\Omega \propto |\epsilon_1 \cdot \epsilon_2|^2$. Therefore, the polarisation vector of the scattered photon cannot have changed by 90 degrees.

If the incoming photon distribution was perfectly isotropic, this would not be important. But there are the tiny anisotropies that we have been discussing. Imagine that (cf fig. **??**) we have colder (less intense) regions just above and below a given point on the sky, and colder regions to the right and left. This corresponds to a local quadrupole on the last scattering surface. The photons that reach us from this point pick up the "left-right" polarisation from cold region, as the the hot region has no polarisation state in that direction (remember that photons, being massless, have only transverse polarisation). Conversely, the "up-down" polarisation is due to the photons from the hot region. As there are more of the latter, the radiation from this point will pick up an overall "up-down" polarisation. Even if the radiation was not polarised initially, it is now.

The polarisation vectors themselves depend on the the orientation of our coordinate system, while temperature doesn't. This makes actual calculations very complicated. It is easier to decompose the polarisation vector-field into a rotation-free part which corresponds to a divergence of a scalar field (called *electric* or E polarisation) and the rotational part (called *magnetic* or B polarisation). Scalar perturbations do not produce any B polarisation, while gravitational waves do. This leads to the hope that a measurement of the B-type polarisation would allow us to measure the gravitational wave background. Unfortunately the signal is expected to be very small, and may well be swamped by foregrounds. It is currently believed that $r \approx 10^{-5}$ is achievable, which means that we could measure the B modes from gravity waves if the scale of inflation is larger than about 2×10^{15} GeV.

The E-mode polarisation on the other hand helps to break certain degeneracies, and it also carries a distinct signal from late scattering processes like reionisation. Many non-standard processes like topological defects, or other mechanisms that rely on purely causal late-time effects to create anisotropies can be fine-tuned to match the observed temperature power spectrum, but will in general lead to a different polarisation power spectrum. Although the data is at the moment still rather limited, we expect a first true E polarisation power spectrum very soon, from the second-year data of the WMAP satellite.

8.5.2 Secondary Anisotropies

There are several additional mechanisms that can significantly change the photon temperature on their way from last scattering to us. These are called *secondary anisotropies*. We can roughly classify them into gravitational and scattering secondaries.

Gravitational secondaries

The primary anisotropies consist of those that contribute in first order perturbation theory. But as the matter perturbations collapse and start to form galaxy halos, they leave the range of linear perturbations as $\delta \gg 1$. When such a perturbation has virialised, we expect that $\rho\delta \approx$ constant, and the corresponding gravitational potential Ψ grows like a^2. Even though the density perturbation is large, the gravitational potential is usually still very small, and we can use linear perturbation theory to compute the redshifting of the photons that pass through this evolving gravitational potential. The integrated Sachs-Wolfe contribution is

$$\frac{\delta T}{T} = -2\frac{\Phi_i}{a_i^2}\left(a_f^2 - a_i^2\right) \approx 4\Psi\frac{\delta z}{1+z}, \tag{8.19}$$

where δz is the redshift difference across the structure (usually very small) and z the averaged redshift. This contribution to the temperature anisotropy is called the *Rees-Sciama* (RS) effect. An additional contribution, sometimes called *moving halo* effect, is due to a possible bulk motion of the galactic halos across the line of sight on large scales.

A second gravitational secondary perturbation is *gravitational lensing* from the gravitational potential of the large-scale structure. Lensing does not change the surface brightness and so only affects anisotropies. It is therefore second order in perturbation theory. The lensing deflects the photons in a pseudo-random fashion. In this way, it smoothes the features in the temperature and polarisation maps. It also does not respect the E/B-type decomposition of the polarisation, and creates B polarisation from E polarisation. This can mask the "real" B polarisation from gravity waves, and sets a limit on the smallest gravitational wave background that we can detect.

Scattering secondaries

Photons are predominantly scattered by free electrons, requiring the ionisation of hydrogen. There two major ways how this takes place. Firstly, we know that hydrogen is ionised for redshifts smaller than at least 6 – otherwise, neutral hydrogen would absorb radiation in quasar spectra and lead to the so-called Gunn-Peterson trough, which is not observed. This scattering at late times suppresses the anisotropies roughly exponentially, parametrised by an optical depth τ which represents the fraction of electrons that were re-scattered after "last scattering". As with the initial generation of the CMB polarisation signal, the local quadrupole at rescattering will generate polarisation. Of course the local quadrupole corresponds now to a larger angular scale, which allows us to distinguish the two mechanisms and to pinpoint the redshift of reionisation. Although the first-year WMAP polarisation data is not very good, there is a tentative detection of the reionisation signal in the temperature-polarisation cross power spectrum.

Secondly, the gas in galaxy clusters is very hot ($T \sim$ keV) and ionised. Some of CMB photons that pass through clusters are scattered by the very hot cluster electrons. In the process, they gain energy and the photon distribution function is distorted and has no longer a perfect black body form. Instead, it has a non-zero Compton y parameter, as discussed in eq. (5.6). The temperature change is roughly

$$\frac{\delta T}{T} \approx -2y, \qquad y = \int_i^f \left(\frac{T_e - T}{m_e}\right) n_e \sigma_T dt, \tag{8.20}$$

integrated along the photon path. We immediately notice that this expression does not depend on the redshift of the cluster, allowing us to detect clusters at arbitrary redshifts (as long as we can resolve the angular size, which of course is smaller for high-redshift clusters)! This is called the *Sunyaev-Zel'dovich* (SZ) effect.

The SZ effect is expected to be main contribution to the CMB power spectrum beyond the damping tail. But its distinctive spectral signature should allow us to subtract it quite well from anisotropy maps and to probe the other, lower secondary contributions.

8.5.3 non-Gaussianity

So far we have only discussed the power spectrum of the temperature distribution. This is the two-point function, i.e. we compare the temperature at two points on the sky. Of course nothing stops us from comparing the temperature at three, four or more points.

If the random process that produced the initial fluctuations is Gaussian, and all subsequent effects are linear (preserving the Gaussian property), then there is no need to go beyond the two-point description. All higher order correlation functions are uniquely described by the power spectrum. This means that **if** the statistics of the CMB is Gaussian, then we can use the power spectrum alone to compute the likelihood of cosmological models.

We can also turn the argument around: by testing if the statistical nature of the fluctuations is Gaussian, we can gain some knowledge of the initial fluctuations and detect, constrain and maybe subtract the non-linear secondary anisotropies. An important example is the subtraction of the lensing anisotropies of the B mode polarisation in order to recover the B mode polarisation from the gravitational wave background.

The current consensus is that the initial fluctuations from single-field slow-roll inflation models are very close to Gaussian. This provides a testable prediction for these cases. Multi-field models and more exotic sources may exhibit a strong non-Gaussian signal.

Of course, as we describe non-Gaussian through what it isn't (namely, not Gaussian), there is an infinity of ways in which a given signal can be non-Gaussian. Correspondingly, it is very difficult to develop general methods and to compute general predictions. Furthermore, many methods tend to be much more computationally intensive than the power spectrum.

There are currently three main groups of methods which are used to test for the presence of non-Gaussianity. Firstly the n-point correlation functions. They are often used in Fourier (or harmonic) space where they are called n-spectra, e.g. the bispectrum (three-point function) and the trispectrum (four-point function). Secondly, wavelet methods, which represent scale-filters somewhere in between pixel space and Fourier space in that they are localised in both. And thirdly the so-called Minkowski functionals which characterise properties such as the number of pixels above a certain temperature threshold.

Although there have been detections of systematic effects through non-Gaussian methods, there is so far no sign of any primordial non-Gaussianity in the CMB.