Cosmology 2012: Lecture Notes

Andrew H. Jaffe, Imperial College

Contents

Preface

Very little of the following is new. Most of it comes, in one form or another, from one of the following books.

- Rocky Kolb & Mike Turner, The Early Universe
- Andrew Liddle, An Introduction to Modern Cosmology
- Michael Rowan-Robinson, Cosmology
- Peter Schneider, Extragalactic Astronomy and Cosmology: An Introduction

In some cases I explicitly give a reference at the beginning of a chapter or section.

Chapter 1

Introduction: Homogeneity and Isotropy

MRR 4.1-4.2

See introductory slides.

Chapter 2

The Expanding Universe

2.1 Expansion

Before we start discussing a physical and mathematical model to describe an expanding Universe, it is useful to be able to picture just what we mean by "expansion" in this context. On "small" scales, corresponding to relatively nearby galaxies, it is completely appropriate to think of the expansion as a relative velocity between any two galaxies, increasing with time and distance. On larger scales, however, such that the time between the emission of light at the distance galaxy and its reception here is large, then we are really observing the galaxy at a very different time and we cannot say that we are observing a quantity having anything to do with its velocity relative to us. Hence, we will see that there are very many galaxies in the distance Universe for which we would *naively* define a velocity $v > c$, but this is just not the correct way to interpret our observations.

A similar confusion occurs when talking about the *distance* to that galaxy. We simply cannot lay out relativistic meter sticks between us, now, and the distant galaxy, then, which is the appropriate thought-experiment to do in order to define the distance. Hence we will also find that there are a myriad of different ways to define the distance, each numerically different, although none of them is any more correct than the others. Nonetheless even at this early stage it is very useful to lay down a coordinate system in our expanding Universe. We could of course imagine a proper coordinate system, with the coordinates given by the number of actual, physical meters, from some fiducial point which we could choose to be the origin. But in this coordinate system, due to expansion, the galaxies are moving. Instead, try to picture a coordinate system with the grid lines expanding along with all of the objects in the Universe. If we could freeze time, this coordinate system would just look like a normal cartesian (or polar, or however we decide to draw our axes), but the grid lines are further apart from one another in the future, and closer together in the past. So for this case we don't really want to measure things in physical units (meters or megaparsecs), but really just numbers. Conventionally, however, we do choose the units so that they are equivalent to physical units today — since many of our measurements are done now.

Real galaxies or other objects may have small movements (called "peculiar velocities") with respect to this coordinate system, but *on average* — due to the homogeneity and

isotropy of the Universe — they will be at rest with respect to these coordinates as the Universe expands. We call this a *comoving coordinate* system, and hypothetical observers expanding along with these coordinates are called *comoving observers*. We show this (in two dimensions) in Figure 2.1.

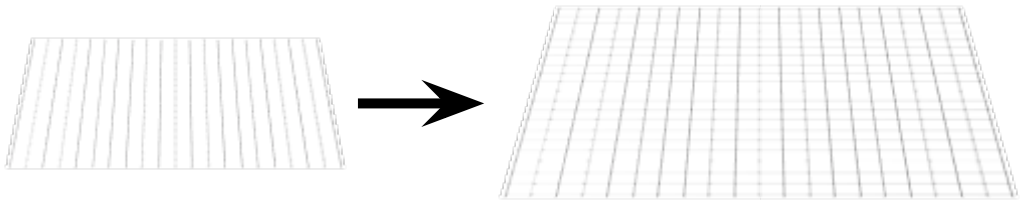

Figure 2.1: Comoving coordinates and observers in an expanding Universe.

2.2 The Scale Factor and Hubble's law

MRR 4.3

We want to work out the mathematical description of an expanding, homogeneous and isotropic Universe. First consider a bunch of points in the Universe at some time t_1, and then later at some time t_2, as in Figure 2.2. If the Universe is to remain homogeneous, the points can't "bunch up" relative to one another. If $\mathbf{r}_{ij}(t) = \mathbf{r}_i(t) - \mathbf{r}_j(t)$ is the distance between points i and j at a given time, we must *scale* all of these relative distances as a function only of time, but not position. That is

$$\begin{aligned} r_{ij}(t_2) &= a(t_2) r_{ij}(t_0) \\ r_{ij}(t_1) &= a(t_1) r_{ij}(t_0) \end{aligned} \tag{2.1}$$

where t_0 is some arbitrary time, which we can take to be the present day. We can combine these to get

$$r_{ij}(t_0) = a^{-1}(t_1) r_{ij}(t_1) = a^{-1}(t_2) r_{ij}(t_2) = \text{const} \, . \tag{2.2}$$

(It's a constant since the value at any specific time is fixed.) So, at a general time, t,

$$a^{-1}(t) r_{ij}(t) = \text{const} \tag{2.3}$$

and we can take the derivative:

$$a^{-1}(t) \dot{r}_{ij}(t) - a^{-2}(t) \dot{a}(t) r_{ij}(t) = 0 \tag{2.4}$$

or

$$\frac{\dot{a}}{a} = \frac{\dot{r}_{ij}}{r_{ij}} \, . \tag{2.5}$$

This equation looks trivial, but note that the right-hand side depends on i and j, and hence which two points you've chosen, whereas the left-hand side doesn't. Therefore, the

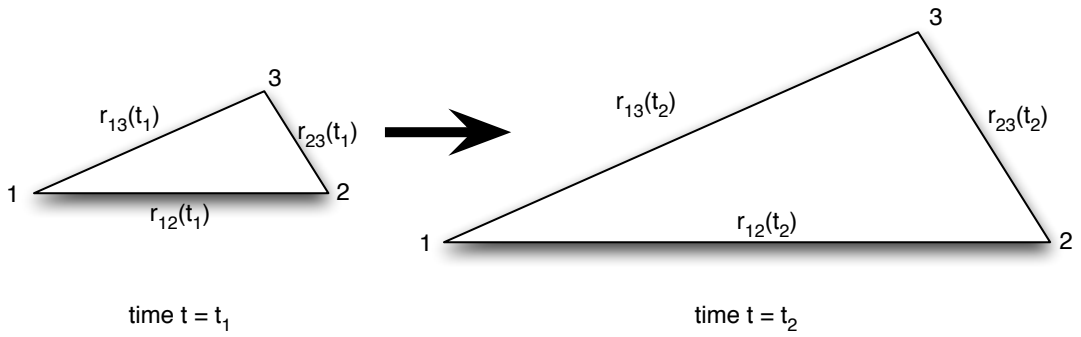

Figure 2.2: Homogeneous and isotropic expansion.

scaling between *any* two points is just a function of time, and we can write most generally

$$\frac{\dot{a}}{a} = \frac{\dot{r}}{r} \tag{2.6}$$

for an arbitrary distance r. (Note that we started our derivation assuming that a has no units, but in fact in this last equation we see that we could also give a units of length which is sometimes done.)

The quantity $a(t)$ — the **scale factor** — is going to be with us throughout the course.[1] It describes the evolution of the Universe, and for a homogeneous and isotropic Universe, it tells us almost everything we need to know. We're also going to continually encounter the combination $\dot{a}/a$, which is the **expansion rate** (also referred to as the *Hubble parameter*) and we will sometimes write

$$H(t) = \left.\frac{\dot{a}}{a}\right|_t \tag{2.7}$$

and in particular

$$H_0 = H(t_0) = \left.\frac{\dot{a}}{a}\right|_{t=t_0} \tag{2.8}$$

is the expansion rate *today*, one of the most important quantities in cosmology. Today, then, we can rewrite this equation as

$$\dot{r}(t_0) = H_0 r(t_0) \ . \tag{2.9}$$

These are all *relative* distances (recall the ij indices we've got rid of), but by convention we can take one of the points to be the location of the earth from which we observe, and the other to be the location of some galaxy (since we can only measure the distances to objects, not arbitrary points!), so we can then write,

$$\boxed{v = H_0 d} \tag{2.10}$$

[1] In some texts, including *MRR*, the scale factor is denoted $R(t)$ rather than $a(t)$.

where v is the velocity of the galaxy relative to us, and d is its distance. This is *Hubble's law*, which we've derived assuming homogeneity, but in fact Hubble found it through observations as we discussed last time.

Comments about there not being a true centre...

2.3 Newtonian Dynamics

What do we need to do in order to have a theory which predicts the scale factor $a(t)$ and hence the expansion rate? In truth, we need the full artillery of General Relativity (GR), which we'll touch on later. But we can do a surprising amount using Newtonian mechanics and gravity.

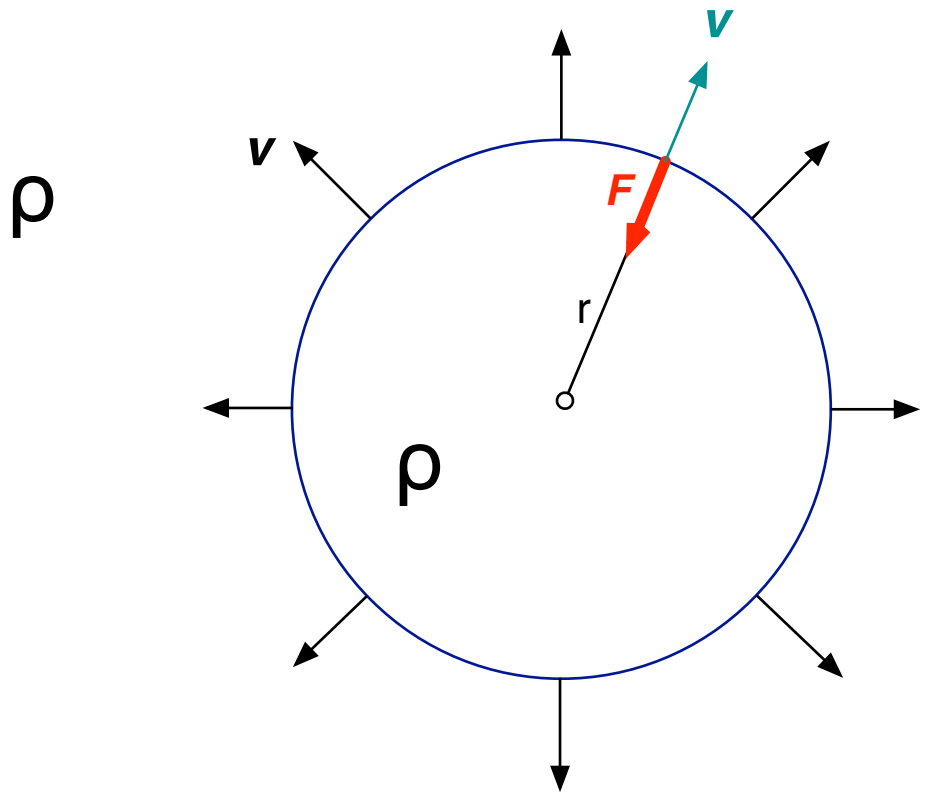

Figure 2.3: An expanding uniform-density sphere of matter, itself embedded in a medium (an infinite sphere) of the same density.

In addition to the usual $F = ma$, and Newton's law of gravitation, we'll need to remember two very important corollaries from Newtonian gravity. First, there is no gravitational field *inside* a spherical shell of matter. Second, we can treat the gravitational attraction of a sphere as if it were concentrated at the centre. Hence, if we're inside a

homogeneous sphere of matter of density ρ (assume it's gas so we can move around) at some distance r from the centre, we can ignore everything at greater radii. We will also assume that the matter is pressureless so it takes no $P\,dV$ work for it to move — this is conventionally called "dust" or simply "non-relativistic matter" in cosmology and relativity.

This is the setup for the Universe, as shown in Figure 2.3, except that we'll take the *outer* sphere to be arbitrarily (infinitely) large. This is a bit of a cheat in Newtonian gravity, but in fact this derivation is correct, and indeed goes through to GR. So consider a point P at a distance r from the "centre". The gravitational force on a test particle of mass m at P is

$$F = m\ddot{r} = -\frac{GMm}{r^2} = -\frac{Gm}{r^2}\frac{4}{3}\pi r^3\rho = -\frac{4\pi Gm\rho r}{3}\ , \tag{2.11}$$

using the mass interior to r, $M = 4\pi r^3\rho/3$. We can rewrite this as

$$\frac{\ddot{r}}{r} = -\frac{4}{3}\pi G\rho \tag{2.12}$$

Now, if we assume that the sphere is expanding homogeneously, our test mass at point P is actually moving with time, $r(t) \propto a(t)$, but so is all the matter both interior and exterior to it. Hence, although the density, ρ, of the interior sphere is changing with time, the total mass in that sphere (M, above) remains constant. So in fact we can just go back to the second equality above and write

$$\ddot{r} = -\frac{MG}{r^2} \tag{2.13}$$

where now we note that M is a constant with time. This is a differential equation that we can integrate, and find

$$\dot{r}^2 = 2GM/r + A \tag{2.14}$$

where A is an integration constant. We can rewrite this in the form

$$\left(\frac{\dot{r}}{r}\right)^2 = \frac{2GM}{r^3} - \frac{A}{r^2} = \frac{8\pi G}{3}\rho - \frac{A}{r^2} \tag{2.15}$$

where we have replaced $M = 4\pi r^3\rho/3$. With this substution, Eq. 2.12 and Eq. 2.14 only contain factors of r in combinations $\dot{r}/r$, $\ddot{r}/r$ or with an integration constant. Recall that we can relate the distance between any two points as a function of time, $r(t)$, to the scale factor via $r(t) = a(t)r(t_0)$. Hence in these terms the factors of $r(t_0)$ cancel or can be absorbed into a constant, and so we can rewrite these differential equations as

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\rho \tag{2.16}$$

and

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) - \frac{kc^2}{a^2} \tag{2.17}$$

where we have written the integration constant as kc^2 where c is the speed of light so that the combination a^2/k has units [length]^2 — some authors use that to make k a number and give a units of length, while others make a a number and k have units [length]$^{-2}$. For now only the overall combination k/a^2 matters.

Exercise: solve Eq. 2.13 to get Eq. 2.14. This is a slightly simplified version of the **Friedmann Equation**, and is one of the most important equations in cosmology. We will encounter it again and again over the entire course. Despite the Newtonian derivation (which required some hand-waving) it is in fact generally true for a homogeneous and isotropic space-time, and was first derived by Friedmann using General Relativity.

We have used the fact that the mass in a sphere that expands along with the Universe remains constant. This is called a *comoving* sphere, another concept that we will encounter again and again. We can use this observation to write down an equation for the evolution of the density:

$$\frac{d}{dt}\rho(t) = \frac{d}{dt}\left[\frac{M}{4\pi r(t)^3/3}\right] = \frac{-3M}{4\pi r^4/3}\frac{dr}{dt} = -3\rho(t)\frac{\dot{a}}{a} = -3H\rho \tag{2.18}$$

or

$$\dot{\rho} + 3H\rho = 0 \tag{2.19}$$

which is a differential equation expressing the conservation of ρ for pressureless matter (although we derived it by putting in the solution!) and eventually we will see how it can be generalized for other forms of matter that behave in a more complicated way due to relativity (where *pressure* actually contributes to the energy density). The solution to this equation can be written as

$$\rho(t) \propto 1/a(t)^3 \qquad \text{or} \qquad \rho(t) = \rho(t_1)\frac{a(t_1)^3}{a(t)^3} \tag{2.20}$$

where t_1 is some arbitrary but fixed time. Plugging this into the Friedmann Equation,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t_1)\frac{a(t_1)^3}{a(t)^3} - \frac{kc^2}{a(t)^2} \tag{2.21}$$

Up until now, we haven't said much about a, but in fact we know that our Universe is expanding (a is growing, or $\dot{a} > 0$), so if we go back far enough, a gets smaller and smaller. Eventually, then, the first term above will dominate. (Alternately, we can look at the case $k = 0$ which we will see is a very special case, and in fact may describe the

Universe.) For this case,

$$\begin{aligned}
\left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho(t_1)\frac{a(t_1)^3}{a(t)^3} \\
\dot{a} &= \sqrt{\frac{8\pi G}{3}\rho(t_1)a(t_1)^3}\; a(t)^{-1/2} \\
a^{1/2}\, da &= \sqrt{\frac{8\pi G}{3}\rho(t_1)a(t_1)^3}\, dt \\
\frac{2}{3}(a^{3/2} - a(0)^{3/2}) &= \sqrt{\frac{8\pi G}{3}\rho(t_1)a(t_1)^3}\; t \\
a &\propto t^{2/3}
\end{aligned} \tag{2.22}$$

where in the last line we have assumed $a(0) = 0$ (which we can take to *define* $t = 0$ — when the distance between *any* two points was zero).

As an aside, there's another interesting case, if $k > 0$ in Eq. 2.17 there's a time when $\dot{a} = 0$; the Universe stops expanding! We will see soon that this case corresponds to a "closed" Universe which expands to a finite size and then recollapses.

2.4 Thermodynamics

MRR 4.3

If we go back to our conservation equation, Eq. 2.19, we can write it in an illuminating form, multiplying the left and right sides by $a^3\, dt$:

$$\begin{aligned}
a^3\, d\rho + 3\rho a^2\, da &= 0 \\
d(\rho a^3) &= 0 \propto dE/c^2
\end{aligned} \tag{2.23}$$

where in the last line we have first used the fact that, for any comoving volume, ρa^3 is proportional to M, the mass in that volume, and then used relativity ($E = Mc^2$) to write that mass in terms of the total energy in the volume. This is really just the first law of thermodynamics — energy conservation — since for this "dust" matter we have both $dQ = T\, dS = 0$ and $dW = P\, dV = 0$. But let's generalize to $P \neq 0$:

$$\begin{aligned}
dE &= T\, dS - dW = 0 - p\, dV \\
dE/c^2 &= -p\, dV/c^2 \propto -p\, d(a^3)/c^2
\end{aligned} \tag{2.24}$$

where we have assumed reversible, adiabatic expansion ($dS = 0$), and $dW = p\, dV$ is the work done by the system on the surroundings as it expands. So, combining with the above,

$$a^3\, d\rho + 3\left(\rho + \frac{p}{c^2}\right) a^2\, da = 0 \tag{2.25}$$

or

$$\frac{d\rho}{dt} + 3\left(\rho + \frac{p}{c^2}\right)\frac{\dot{a}}{a} = 0 \tag{2.26}$$

Note that the units here make sense:

$$\rho c^2 = \frac{\text{Energy}}{\text{Volume}} = \frac{\text{Force} \times \text{length}}{(\text{length})^3} = \frac{\text{Force}}{\text{Area}} = \text{Pressure} \tag{2.27}$$

The presence of the speed of light, c, indicates that this is essentially an effect of relativity, and in fact we have to modify our force equation in addition to this conservation equation. The source in relativity (even special relativity) isn't just mass, but "stress energy", encoded in the **Stress-Energy tensor**. For the fluids we will be concerned with here, this is given by

$$T_{\mu\nu} = \begin{pmatrix} \rho & & & \\ & p/c^2 & & \\ & & p/c^2 & \\ & & & p/c^2 \end{pmatrix} \tag{2.28}$$

(empty entries are zero). What we really need is the Trace of this tensor,

$$\text{Tr}\, T = \rho + 3p/c^2 \tag{2.29}$$

which changes the force equation, Eq. 2.16 to

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p/c^2\right). \tag{2.30}$$

Note that, in fact, Eq. 2.13 for $\dot{a}/a$ *still holds*, but is now independent of this acceleration equation for $\ddot{a}$

We still need a relation to link the density and the pressure, known as the **equation of state**. For many kinds of stuff, we will assume

$$\frac{p}{\rho c^2} = w = \text{const} \tag{2.31}$$

so, for example, $w = 0$ for the "dust" we've been dealing with, and a radiation-dominated medium has radiation pressure $p = \rho c^2/3$ so $w = 1/3$. For these cases, we can also calculate the sound speed:

$$c_s^2 = \frac{\partial p}{\partial \rho} = \frac{p}{\rho} = wc^2 \tag{2.32}$$

2.4.1 Radiation-Dominated Universe

MRR 5.1

Going back to the radiation-dominated case with $p = \rho c^2/3$ (so $w = 1/3$), the fluid equation, Eq. 2.26, becomes

$$\begin{aligned} 0 &= \dot{\rho} + 3\rho(1 + 1/3)\frac{\dot{a}}{a} \\ \dot{\rho} &= -3 \times \frac{4}{3}\rho\frac{\dot{a}}{a} \\ \frac{\dot{\rho}}{\rho} &= -4\frac{\dot{a}}{a} \end{aligned} \tag{2.33}$$

which has the solution

$$\rho_{\rm rad} \propto a^{-4} \ . \tag{2.34}$$

Compare $\rho_{\rm mat} \propto a^{-3}$ for dust — we will understand the difference when we discuss the redshift in a few lectures.

Now, let's insert this into Eq. 2.17, again considering $k = 0$ appropriate for sufficiently early times:

$$\left(\frac{\dot{a}}{a}\right)^2 \simeq \frac{8\pi G}{3}\rho \tag{2.35}$$

which, for our radiation-dominated case, is

$$\left(\frac{\dot{a}}{a}\right)^2 \simeq \frac{8\pi G}{3}\frac{a_1^4}{a^4}\rho_1 \tag{2.36}$$

where the subscript 1 refers to some fixed time at which the quantities are evaluated. We can solve this:

$$\begin{aligned} \dot{a} &= \sqrt{\frac{8\pi G}{3}a_1^4}\, a^{-1} \\ a\, da &= \sqrt{\frac{8\pi G}{3}a_1^4}\, dt \\ a^2/2 &= \sqrt{\frac{8\pi G}{3}a_1^4}\, t \end{aligned} \tag{2.37}$$

or

$$a \propto t^{1/2} \qquad \text{radiation} \tag{2.38}$$

which we can compare to

$$a \propto t^{2/3} \qquad \text{non-relativistic matter} \tag{2.39}$$

where from now on we will usually refer to the "dust matter" as "non-relativistic matter".

2.5 General and Special Relativity

MRR 4.4-4.5

It turns out that we've been able to use Newtonian Gravity and a few "handwavey" arguments from Special Relativity to derive equations that do, in fact, apply to the real Universe.

These are the **Friedmann Equations**:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) - \frac{kc^2}{a^2} \ , \tag{2.40}$$

and

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p/c^2\right) \ , \tag{2.41}$$

along with the conservation equation

$$\frac{d\rho}{dt} + 3\left(\rho + \frac{p}{c^2}\right)\frac{\dot{a}}{a} = 0 \tag{2.42}$$

But the real Universe is governed by General Relativity (GR). So to really understand what these equations mean (and in particular the significance of the constant of integration, k) we need to use a little bit of GR.

2.5.1 Metrics and Curvature

As you probably know, GR is a theory which links gravity to the underlying curvature of the spacetime manifold. This curvature, in turn, is described by a *metric*, which encodes the distances between any two points on the manifold. We can build this up using the invariant interval,

$$ds^2 = \sum_{\mu,\nu} g_{\mu\nu} dx^\mu dx^\nu \tag{2.43}$$

where the indices μ, ν run over 0,1,2,3, corresponding to time and three spatial coordinates, $g_{\mu\nu}$ is the metric, and dx^μ is a differential spacetime coordinate interval (in order to have units of length, we will take the time coordinate, $\mu = 0$, to be $dx^0 = c\,dt$ where c is the speed of light). Because we are free to use different coordinates (for example, polar vs. Cartesian), the actual form of the metric depends on the coordinates used — and the requirement that the physics not depend on this is one of the central facts of GR.

Note that we will occasionally use the "Einstein Summation Convention" and write this as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \qquad \text{summed over repeated indices .} \tag{2.44}$$

You have probably already encountered a metric in *special* relativity, which we used to describe the fact that the time coordinate can mix with the spatial coordinates. In this case, the appropriate metric is the Minkowski metric, which, for Cartesian coordinates $(x_0, x_1, x_2, x_3) = (ct, x, y, z)$ is just

$$g_{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} +1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \tag{2.45}$$

[note that we will use the "signature" $(+---)$ but you will often encounter $(-+++)$ in the literature]. Here, then,

$$ds^2 = c^2 dt^2 - ds_3^2 \qquad \text{where } ds_3^2 = dx^2 + dy^2 + dz^2 \ . \tag{2.46}$$

If instead we had decided to use polar coordinates, we could instead write the spatial three-dimensional part as

$$ds_3^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \ . \tag{2.47}$$

Note that the angular part of this can itself be written as

$$r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right) = r^2 d\Omega_2^2 \tag{2.48}$$

where, finally $d\Omega_2^2$ represents the metric on the surface of the $r = 1$ sphere (more specifically, of the two-sphere, where "two" represents the number of dimensions of the surface of the sphere) — this is our first example of the metric of a curved surface, even though it is a surface embedded in a larger, flat, spacetime.

To get a feeling for this, let's take a look at a metric that describes a circle — that is, distances along the perimeter of a circle of radius R. This is just a one-dimensional, curved manifold. It is pretty clear that we should be able to write

$$ds^2 = R^2 \ d\theta^2 \tag{2.49}$$

which simply says that distances along the perimeter of the circle satisfy $ds = R\, d\theta$. But it is also case that we can write the metric in a different coordinate system as

$$ds^2 = \frac{dy^2}{1 - ky^2} \ . \tag{2.50}$$

This is clear if we make the substitution $ky^2 = \sin^2\theta$, in which case we can write

$$ds = \frac{k^{-1/2}\cos\theta \ d\theta}{\sqrt{1 - \sin^2\theta}} = k^{-1/2} \ d\theta \tag{2.51}$$

so we can identify $k = 1/R^2$ and we can also interpret our second set of coordinates to mean that $y = R\sin\theta$ represents y in a Cartesian coordinate system (with θ measured counter-clockwise from $y = 0$, $x = R$).

Exercise: The above discussion assumes $k > 0$. Show that $k = 0$ corresponds to distances along a line, and $k < 0$ to distance along a hyperbola.

The FRW Metric

Now, we wish to expand this discussion to General Relativity in a cosmological context. From the first lecture, we know that, to a good approximation, the Universe is *homogeneous and isotropic* — the cosmological principle. Thus, on large scales, we want to find a spacetime manifold (basically, a shape in three spatial and one time dimensions) that, at any given time, looks the same in all directions and from all places. To guide our intuition, let's think about this in two dimensions first. What two-dimensional shapes satisfy the requirements of homogeneity and isotropy? Certainly, the infinite plain does so: it looks the same everywhere on it. So does the sphere: there are no special points

anywhere. It turns out there is one other, not so obvious, possibility, which is the "hyperbolic paraboloid" or saddle shape. All of these possibilities are shown in Figure 2.4. For the sphere, we can obviously label the manifold by its radius in some units, and in fact the flat plane corresponds to the infinite-radius limit. We can also define a "radius of curvature" for the hyperbolic manifold. (Indeed, many calculations on the hyperboloid just correspond to using the hyperbolic trigonometric functions in place of the spherical functions.) For both of the curved cases, the manifold is self-similar: it is just a uniform scaling between different values of the radius of curvature.

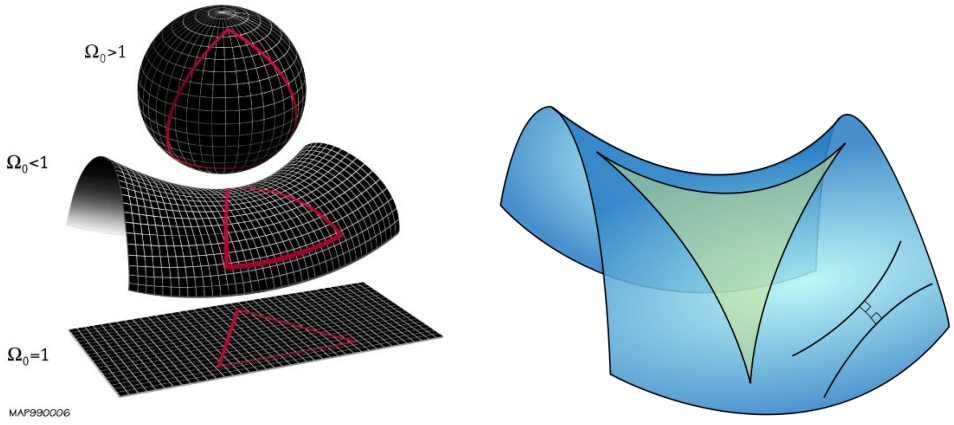

Figure 2.4: Homogeneous and isotropic manifolds in two dimensions. Left: All three constant-curvature manifolds, courtesy WMAP/NASA GSFC. Right: Hyperbolic paraboloid courtesy Wikipedia: `http://en.wikipedia.org/wiki/File:Hyperbolic_triangle.svg`

It is worth spending a little time understanding the meaning of "curvature" in relativity. We can usually only picture curvature when the manifold is embedded in a higher-dimensional spacetime (for example, Figure 2.4 shows the two-manifolds embedded in flat three-dimensional space — and then projected down to two-dimensional flat space), but we can define curvature more generally, based on concepts defined only on the metric itself. All we need is the concept of a locally straight line — a null geodesic curve. We are used to these from living on the surface of something that is approximately a two-sphere, the Earth. We know what it means to walk in a straight line locally, and we find that if we continue walking that would take us along a great circle. So we would find that two walkers starting off on parallel tracks would eventually meet. Furthermore, we would find that a triangle made of segments of three such great circles does not have a total interior angle of 180° but rather some number greater than that. To see an example of this, consider one of the segments to be along the equator, and note that *any* two lines perpendicular to this meet at the pole, with the angle at the pole just depending on how far apart the two perpendiculars started. So we have a triangle with *at minimum* 180° total angle and at maximum 360° for the case when the two other legs start at antipodal points. Similarly, on the saddle-shaped surface, parallel lines can converge and diverge, and a triangle always contains less than 180°. Since one of the results of GR is that light

follows geodesic paths, we can use lightrays, and therefore astronomical observations, to probe the geometry of the Universe. Later on, we will use this very fact to measure the curvature of the Universe as a whole with the Cosmic Microwave Background radiation.

It turns out that in $3+1$ dimensions, we have the same set of homogeneous and isotropic manifolds, but promoted to higher dimensionality (and so effectively impossible to picture with our 3D brains). The most general spatially homogeneous and isotropic manifolds are indeed flat space, the three-sphere, and the three-hyperboloid. Note in particular that the three-sphere is *not* a "ball" in three dimensions — it is a three-dimensional "surface" with no boundary that can be observed in three dimensions, just as there is no boundary as you walk along the surface of the approximately-spherical earth. The metric for this case is given by

$$ds^2 = c^2dt^2 - a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2\right) \tag{2.52}$$

where we use polar coordinates in three dimensions, and we now have a scale factor $a(t)$, since all that the cosmological principle tells us is that at any given time the Universe is homogeneous and isotropic, but it lets us change the overall scale of the universe as a function of time. In this form, the scale factor has no units. We also have a number k in our equation, and this tells us which of the three possible manifolds we have. As we have written it here, this is a real number with units [length]$^{-2}$, just as in the case of the circle manifold. So in fact it is just the overall sign of k that determines which of the manifolds applies. [You can see this by making the substitution $r \to r/\sqrt{|k|}$ in which case we can rescale $a \to a/\sqrt{|k|}$ as well, and only the overall sign of k matters. Because of this freedom, you will sometimes see this metric written with the terms and factors scaled in this way.] This is the famous Friedmann-LeMaïtre-Robertson-Walker metric (FLRW, although sometimes LeMaïtre is left out to give only FRW). Amazingly, the scale factor a and the constant k that we have been using in our Newtonian calculation correspond to exactly the same quantities in this full analysis.

We can examine the spatial part of this metric separately:

$$d\chi^2 = \frac{dr^2}{1-kr^2} + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2\ . \tag{2.53}$$

We have removed the common $a^2(t)$ factor which converts to physical coordinates; this is the metric of a "comoving" constant-time spatial slice of the metric. This looks a bit like some of the metrics we considered in the previous Subsection 2.5.1: it shares the $dr^2/(1-kr^2)$ term with the circle (exchanging r for y), and the remaining two $r^2d\Omega^2$ terms with the spherical manifolds. Combining these, for $k>0$, gives the metric of a **three-sphere**, that is a *three-dimensional* manifold all of whose points are a distance $R = 1/\sqrt{k}$ from a fixed point in higher dimensions (I emphasize that this is not a normal two-sphere). The discussion above tells us in particular that for $k \neq 0$ the r coordinate does not represent distances along that sphere, but can be interpreted as a Cartesian coordinate in the higher-dimensional space. The r coordinate does, however, represent an important distance: it gives the radius (in comoving length) of a sphere centred at

$r = 0$. This is evident when we go back and recall the interpretation of the dr^2 term as the metric along the circle. In two dimensions we can think of a sphere as a series of "nested" circles of radius r (i.e., as if looking down from the top at lines of constant latitude), and in three dimensions of nested spheres of radius r. Note that the dr term doesn't give the metric along each of these spheres or circles, but how they are related to one another.

The possibilities for the sign of k thus each correspond to one of our constant-curvature manifolds. The choice $k = 0$ corresponds to the flat manifold, which should be clear from the form of the metric, which in this case just looks like Minkowski with the extra factor of a^2 on the spatial part. So at any one time, the Universe acts like Minkowski space, although of course anything that actually happens takes a finite amount of time, so the evolution of the Universe must be taken into account.

If $k > 0$, we say that the Universe has positive curvature, and it takes the form of a three-sphere at any given time. Analogously to the two-dimensional sphere discussed above, triangles have a total of more than 180 degrees, lines that start out parallel will converge and eventually cross, any seemingly-straight line returns to its starting point eventually, and the Universe has a finite volume at any one time. This corresponds to the case discussed a few lectures ago where we saw that $a'(t) = 0$ at some time, corresponding to a maximum expansion followed by recollapse: the sphere grows with time from $R = 0$ and then shrinks.

If $k < 0$, we say that the Universe has negative curvature, and it is the three-dimensional saddle shape or hyperboloid. Here, a triangle has less than 180 degrees, and lines that start out parallel may cross and will diverge, but, like the flat case, the Universe is infinitely large.

2.5.2 Dynamics of the FRW metric

This FRW metric describes the Universe at a given time, but we haven't discussed how it evolves with time, i.e., the evolution of the scale factor $a(t)$. For this we need Einstein's Field Equations for GR. For completeness, we can write these down, but we will not really be able to explain them here:

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{2.54}$$

where $G_{\mu\nu}$ is the "Einstein Tensor" and encodes information about the curvature of the Universe (it is a function of the metric and its derivatives) and the right-hand side is the stress-energy tensor that we have encountered already. We will not to understand this equation in detail for this course, but an excellent introduction from Baez and Bunn, which concentrates more on the dynamics of the Einstein equation rather than the metric, is available at `http://math.ucr.edu/home/baez/einstein/`.

We won't reproduce the original arguments of FLRW here, but the result of their analysis of the field equations for their metric is in fact exactly the same as Equations 2.40-2.42. Actually, there is one important change that we can make. Einstein's Field Equations allow (but do not require) an additional term to be added, giving an extra $\Lambda g_{\mu\nu}$ on

the left-hand side, where $g_{\mu\nu}$ is the metric, and Λ is the infamous cosmological constant ("Einstein's biggest blunder") — and it really is a constant, a single number given once and for all. In fact, we'll see that even this term is surplus to requirements and in fact can really be subsumed into the stress-energy tensor itself, and despite being a "blunder" is probably needed to describe the Universe. So, if we include this term, the Friedmann Equations become

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) - \frac{k}{a^2} + \frac{1}{3}\Lambda \,, \tag{2.55}$$

and

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p/c^2\right) + \frac{1}{3}\Lambda \,. \tag{2.56}$$

In these equations, ρ, p and a are all functions of time, and they must be supplemented with enough information about the behaviour of the mass-energy to solve these equations (for example, fluid conservation, Eq. 2.42 along with the equation of state parameter, $wc^2 = p/\rho$).

2.5.3 Redshift

MRR 3.3, 7.4

Almost all of the information we have about the Universe comes to us in the form of light — photons — that have propagated to us from distant objects. Because of the finite speed of light, of course, this means that we see objects as they were at some time in the past (and it also means that defining the *distance* to an object that you see can be complicated). Light travels along the geodesic curves of a manifold, essentially locally straight lines, satisfying $ds = 0$.

So let's return to our FRW metric, Eq. 2.52, and look at the case where light travels along a "radial" curve with constant θ and ϕ (so $d\theta = d\phi = 0$), from the point of emission at $t = t_e$, $r = r_e$ to our observation today at $t = t_0$, $r = r_0 = 0$. We then have

$$0 = ds^2 = c^2dt^2 - a^2(t)\frac{dr^2}{1-kr^2} \tag{2.57}$$

or

$$\frac{c\,dt}{a(t)} = -\frac{dr}{\sqrt{1-kr^2}} \tag{2.58}$$

where we have chosen the negative sign since both a and t increase to the future, whereas r increases away from the observer, i.e., toward the past. We can integrate this along the path from emission to observation:

$$\int_{t_e}^{t_0} \frac{c\,dt}{a} = \int_{r_e}^{0} \frac{-dr}{\sqrt{1-kr^2}} = \int_0^{r_e} \frac{dr}{\sqrt{1-kr^2}} = f_k(r_e) \tag{2.59}$$

where we note that the dr integral is a fixed function of r_e (and k). Because of this, the equality must still hold for photons emitted at a later time, $t_e + \delta t_e$ and observed at $t_0 + \delta t_0$ (note that the two δs are *different*). The δs correspond to, for example, successive

peaks in the lightwave, since we'll see that we want to measure the photon wavelength or frequency. Hence

$$\int_{t_e}^{t_0} \frac{dt}{a} = \int_{t_e+\delta t_e}^{t_0+\delta t_0} \frac{dt}{a} \tag{2.60}$$

We can now split up the range of integration on both sides to eliminate the "overlap" between $t_0 + \delta t_0$ to t_e:

$$\int_{t_e}^{t_e+\delta t_e} \frac{dt}{a} + \int_{t_e+\delta t_e}^{t_0} \frac{dt}{a} = \int_{t_e+\delta t_e}^{t_0} \frac{dt}{a} + \int_{t_0}^{t_0+\delta t_0} \frac{dt}{a} \,. \tag{2.61}$$

The two "inner" terms are equal, so we are just left with

$$\int_{t_e}^{t_e+\delta t_e} \frac{dt}{a} = \int_{t_0}^{t_0+\delta t_0} \frac{dt}{a} \,. \tag{2.62}$$

Now, let's assume that the δs are small and Taylor expand the integrands:

$$\frac{1}{a(t+\delta t)} = \frac{1}{a(t)} \left[1 - \delta t \left(\frac{\dot{a}}{a}\right) + \mathcal{O}(\delta t^2)\right] \simeq \frac{1}{a(t)} \tag{2.63}$$

where the approximation holds if

$$\frac{\delta t}{(\dot{a}/a)^{-1}} \simeq \frac{c^{-1} \times \text{photon wavelength}}{\text{age of Universe}} \ll 1 \tag{2.64}$$

and we've used $H = \dot{a}/a \sim 1/t_0$ and the fact that we're considering successive peaks in the sinusoidal lightwave. This inequality obviously holds for any reasonable times. Now, applying this to $t = t_0$ and $t = t_e$,

$$\frac{\delta t_0}{a(t_0)} = \frac{\delta t_e}{a(t_e)} \tag{2.65}$$

which we can rewrite as

$$\frac{a(t_0)}{a(t_e)} = \frac{\delta t_0}{\delta t_e} = \frac{\lambda_0/c}{\lambda_e/c} = \frac{\nu_e}{\nu_0} \tag{2.66}$$

or

$$\frac{a_0}{a} = 1 + z \tag{2.67}$$

where we define the **redshift**

$$\begin{aligned} z &= \frac{\nu_e}{\nu_0} - 1 = \frac{\nu_e - \nu_0}{\nu_0} = \frac{\delta\nu}{\nu} \\ &= \frac{\lambda_0 - \lambda_e}{\lambda_e} = \frac{\delta\lambda}{\lambda} \end{aligned} \tag{2.68}$$

What does redshift mean? Most importantly, it means that the wavelength of freely-propagating photons increases with the expansion of the universe, proportional to the scale factor. This is a purely General-relativistic effect. Recalling our heuristic Newtonian

derivation of the expansion of the Universe, we can see that the expansion is very much like climbing out of a potential well, and so it is plausible that moving objects — photons, in this case — would lose energy as they propagate. Note that although our derivation required approximations when we used the Taylor expansion, in fact the cosmological redshift of Eq. 2.67 is *exact* in a full GR calculation.

We also see that when $z > 1$ (which we certainly do observe for sufficiently distant objects), if we use the usual correspondence between redshift and speed, $v = cz$, we seem to have a relative speed greater than that of light. On the one hand, there is no problem here: the equations that give this result are unambiguous. But what does it mean? We are taught that one of the underlying principles of relativity is $v \leq c$, and yet we seem to contradict this. The answer is that this is a matter of interpretation, not physics: Is $v = cz$ correct in this circumstance? What does the prohibition against superluminal velocities mean for very distant objects? We will not go into the details here, but I will point to a recent paper by two of my colleagues, Bunn and Hogg, available at `http://arxiv.org/abs/0808.1081` and discussed further in Bunn's blog at `http://blog.richmond.edu/physicsbunn/2009/12/02/interpreting-the-redshift/`.

Interpretation aside, the redshifting of radiation as the Universe expands has some important and interesting effects. Let us compare photons to our "dust" or "non-relativistic matter". As the Universe expands, the number density of matter particles decreases: $n_m = N/V \propto N/L^3 \propto a^{-3}$. Hence the energy density, which is just $\rho_m c^2 = m n_m c^2$ for particle of mass m, also scales as

$$\rho_m c^2 \propto a^{-3} \, . \tag{2.69}$$

For photons, the number density is still $n \propto a^{-3}$. However, the energy per particle is $E_\gamma = pc = h\nu = hc/\lambda \propto a^{-1}$ as the wavelength redshifts. Hence, for photons, $\rho_\gamma c^2 = n_\gamma E_\gamma \propto a^{-3} \times a^{-1}$ or

$$\rho_\gamma c^2 \propto a^{-4} \, . \tag{2.70}$$

So far, we have discussed the case of photons, but really this distinction is more general. Consider the fully relativistic energy per particle, $E^2 = p^2c^2 + m^2c^4$. If the first term dominates, ($mc^2 \ll pc$), the particles behave like photons, which we will call, generically, *radiation* — this applies to anything with zero mass, but also to very low-mass particles, where "low-mass" really means "high-speed", since in the early Universe we will see that all particles move with greater and greater speed as the Universe gets hotter and hotter as we look earlier and earlier. Conversely, when the second term dominates ($mc^2 \gg pc$), their energy is dominated by their mass, and they behave like non-relativistic matter.

We will very often use the redshift as a proxy for the time parameter — if the Universe has always been expanding, then the redshift of a more distant object (hence with an earlier t_e is always greater). Obviously, this is the redshift as observed by us, today; it would have a different value if observed at a very different time. (As we can see from our derivation, however, as long as the time difference is small compared to the age of the Universe — which it will be for any observations made in the recent past or foreseeable future! — the redshift will be the same. We may explore this further in a problem set.)

Chapter 3

Cosmological Models and Parameters

3.1 Cosmological Parameters and the Expansion of the Universe

MRR 4.8

Let us return to our first-order Friedmann Equation (Eq. 2.55)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) - \frac{kc^2}{a^2} + \frac{1}{3}\Lambda \; . \tag{3.1}$$

Consider what this equation looks like if we evaluate everything today, $t = t_0$. We define the **Hubble Constant** as

$$H_0 = \left.\frac{\dot{a}}{a}\right|_{t=t_0} = 100h \text{ km s}^{-1} \text{ Mpc}^{-1} \tag{3.2}$$

where the second equality just parameterizes our ignorance: we are pretty sure from observations that H_0 is between 50 and 100 km s^{-1} Mpc^{-1}, so this way the quantity that we don't know is $0.5 \lesssim h \lesssim 1$, which is a bit easier to work with. With this, the Friedmann equation is

$$H_0^2 = \frac{8\pi G\rho_0}{3} - \frac{kc^2}{a_0^2} + \frac{1}{3}\Lambda \; . \tag{3.3}$$

Now, consider a *flat* Universe with no cosmological constant, $k = \Lambda = 0$, in which case we will say that the density today is the "critical density", $\rho = \rho_c$, given by

$$H_0^2 = \frac{8\pi G\rho_c}{3} \tag{3.4}$$

or

$$\rho_c = \frac{3H_0^2}{8\pi G} = 1.9h^2 \times 10^{-29} \text{ g cm}^{-3} \; . \tag{3.5}$$

Even for a universe with $\rho \neq \rho_c$, we can define the ratio of the actual density to the critical density, the **density parameter**

$$\Omega_0 = \frac{\rho_0}{\rho_c} \tag{3.6}$$

The two parameters H_0 and Ω_0 have traditionally been the most important numbers in modern cosmology, and until the last decade or so, their values were not really known to better than 50% or so. However, we'll see right away that we can't really use only a single number, Ω_0, to represent the density; we really a need a separate number for each "kind" of matter.

To see this, let's rewrite the Friedmann equation in a very useful form. First, let's write the expansion rate as

$$H(t) = \frac{\dot{a}}{a} = H_0 E(z) \tag{3.7}$$

which defines a new function $E(z)$, first popularized by James Peebles in his books and papers, which we will call the "Hubble Function". The Friedmann equation becomes

$$\begin{aligned} E^2(z) &= \frac{8\pi G\rho}{3H_0^2} - \frac{kc^2}{H_0^2 a^2} + \frac{\Lambda}{3H_0^2} \\ &= \frac{\rho(z)}{\rho_c} - \frac{kc^2}{a_0^2 H_0^2}(1+z)^2 + \frac{\Lambda}{3H_0^2} \ . \end{aligned} \tag{3.8}$$

From the previous chapter, we know that, for non-relativistic matter,

$$\rho_m(z) = \rho_m(0)\left(\frac{a}{a_0}\right)^{-3} = \rho_m(0)(1+z)^3 = \rho_c\Omega_m(1+z)^3 \tag{3.9}$$

and for radiation

$$\rho_r(z) = \rho_r(0)\left(\frac{a}{a_0}\right)^{-4} = \rho_r(0)(1+z)^4 = \rho_c\Omega_r(1+z)^4 \ . \tag{3.10}$$

We can similarly *define*

$$\Omega_k = -\frac{kc^2}{a_0^2 H_0^2} \qquad \text{and} \qquad \Omega_\Lambda = \frac{\Lambda}{3H_0^2} \tag{3.11}$$

(we really shouldn't think of Ω_k, especially, as having anything to do with an energy density, although we will see that Ω_Λ actually is related to an appropriate energy density $\rho_\Lambda c^2$), which finally gives

$$E^2(z) = \Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_\Lambda \ . \tag{3.12}$$

In particular, at $z = 0$ (today), $E(0) = 1$, since $\dot{a}/a|_0 = H_0$, so

$$1 = \Omega_m + \Omega_r + \Omega_k + \Omega_\Lambda \tag{3.13}$$

where I will always use the quantities Ω_i to refer to the value *today*. Note that this assumes that Eqs. 3.9-3.10 are exact — that there is no conversion between components due to particle decay, for example. Because all of the Ω_i *except* Ω_k can refer to the actual density of some sort of "stuff", we shall also define the total density parameter

$$\Omega_{\rm tot} = \Omega_m + \Omega_r + \Omega_\Lambda = 1 - \Omega_k \ . \tag{3.14}$$

3.1.1 The time-dependence of the density

Let us now start to consider how the Universe evolves with time. Note that each of the terms in our Hubble Function has a different power of $(1 + z)$ and hence as we move further and further back in time, the higher powers will dominate. In particular, if we go back far enough, the Ω_r and Ω_m terms will dominate over the Ω_Λ and Ω_k terms — and hence the Universe will look like a geometrically flat ($k = 0$) universe with no cosmological constant ($\Lambda = 0$).

Indeed, if we consider observations of the Universe today, we find that the matter density is $\Omega_m \simeq 0.3$ and that the radiation density can be found by considering the *Cosmic Microwave Background radiation*, which has a temperature of about $T_{\rm CMB} = 2.73$ K, this corresponds to a density (for a blackbody)

$$\Omega_{\rm CMB} = \frac{\rho_{\rm CMB}}{\rho_c} = \frac{1}{\rho_c c^2}\left[\frac{\pi^2 k_B^4}{15(\hbar c)^3}\right] T_{\rm CMB}^4 \simeq 2.5 \times 10^{-5} h^{-2} \tag{3.15}$$

which is much smaller than the matter density today.

Consider, for example, $1 + z = 1000$, which we will see corresponds to a particularly interesting time in the early Universe. Then,

$$\begin{aligned} E^2(z) &= \Omega_m(10^3)^3 + \Omega_r(10^3)^4 + \Omega_k(10^3)^2 + \Omega_\Lambda \\ &\simeq 10^9\Omega_m + 10^{12}\Omega_r \\ &\simeq 10^9\Omega_m \end{aligned} \tag{3.16}$$

Because the evolution at this time is controlled by the value of Ω_m, we say that the Universe is *matter-dominated* (MD). However, if we go to a somewhat earlier time (any $z \gg 0.3/2.5 \times 10^5$), the Ω_r term is largest and we say that the Universe is *radiation-dominated* (RD). Similarly, if we wait long enough, and if Ω_k and Ω_Λ are nonzero, we expect them to dominate eventually. We show all of these in Figure 3.1, although of course the real universe wouldn't have such sharp transitions between the different phases. Note that the actual Universe need not have all of these phases, if one of the Ω_i is sufficiently small or zero. In fact, we're pretty sure that $\Omega_k = 0$ in our Universe, so the curvature-dominated phase probably doesn't happen. Moreover, it looks like $\Omega_\Lambda \simeq 0.7$, so we are experiencing the change-over from MD to ΛD *now*.

3.2 Cosmological Models

MRR 4.6

Although Figure 3.1 gives a good picture of the Universe in its various phases, let's see what happens in a bit more detail as we vary the parameters. It will be useful to go back to our first-order Friedmann Equation in its original form, (Eq. 2.55):

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) - \frac{kc^2}{a^2} + \frac{1}{3}\Lambda\,. \tag{3.17}$$

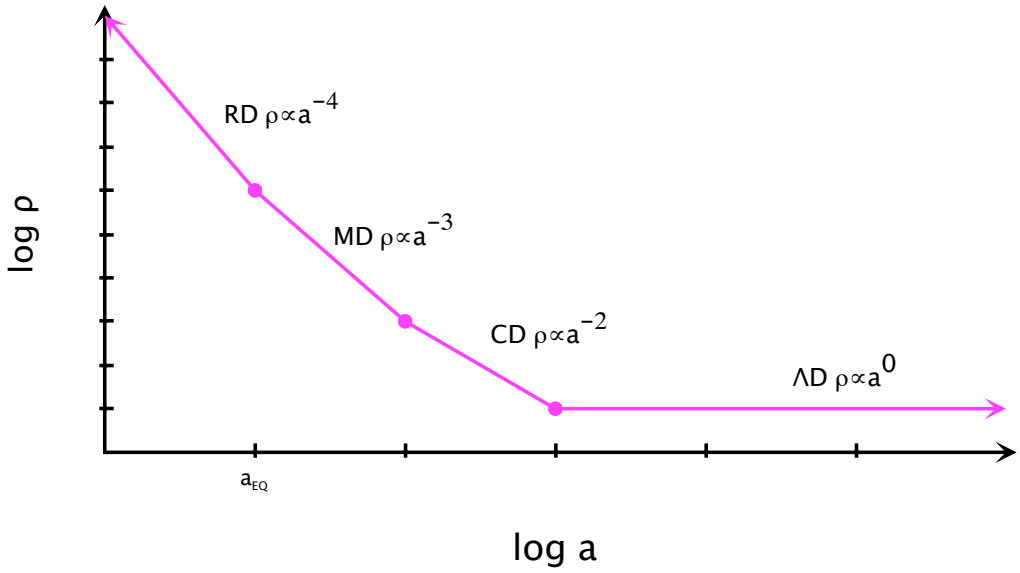

Figure 3.1: The evolution of the density of various components. RD, MD, CD, and ΛD refer to Radiation-, Matter-, Curvature- and Λ-domination, respectively.

The Empty Universe. First, we will consider a manifestly non-physical Universe, an open ($k < 0$) one with no matter, no radiation, and no cosmological constant. This just gives the very simple equation

$$\dot{a}^2 = kc^2 \,, \qquad \text{or} \quad a = \pm\sqrt{|k|}\, ct \tag{3.18}$$

so

$$\frac{\dot{a}}{a} = t^{-1} \,. \tag{3.19}$$

This is the "wide open", or *Milne* Universe. Note that it also corresponds to the far future evolution of any open Universe ($k = -1$ or $\Omega_k > 0$ or $\Omega_{\rm tot} < 1$) in which the matter and radiation densities have diluted so much as to become negligible, as long as there is no cosmological constant $\Omega_\Lambda = 0$.

The flat, matter-dominated Universe. Now, we take $k = 0$ and $\Omega_m = 1$. We already considered this in Section 2.3 above. In this case, we have

$$a \propto t^{2/3} \tag{3.20}$$

which gives

$$\frac{\dot{a}}{a} = \frac{2}{3} t^{-1} \,. \tag{3.21}$$

This corresponds to early times for a Universe with non-relativistic matter (but not so early that a radiation component dominates), and is often called the "Einstein-de Sitter" Universe (not to be confused with the "de Sitter" Universe, which is, unfortunately, a different case!).

The Closed Universe. Now, consider $k > 0$, $\Omega_{\rm tot} > 1$, or $\Omega_k < 0$. We saw earlier that in this case there is a value of a such that $\dot{a} = 0$. This occurs when

$$\frac{8\pi G\rho}{3} = \frac{kc^2}{a^2} \tag{3.22}$$

or (assuming non-relativistic matter so $\rho \propto a^{-3}$)

$$a_{\rm max} = \frac{8\pi G}{3kc^2}\rho_0 a_0^3 \; . \tag{3.23}$$

In fact, the full solution for the evolution of $a(t)$ is symmetric around this point: it climbs from $a = 0$ at $t = 0$ to $a_{\rm max}$ at $t = t_{max}$ and then falls back to $a = 0$ at $= 2t_{\rm max}$.

So far, we've ignored the possibility of a radiation-dominated Universe. In fact, in terms of the long-term evolution of the Universe, it appears that the radiation-dominated phase is very short and we are now long after it. *Exercise: show that the effect on the age of the Universe (i.e., the relationship between a and t) of an early RD phase is negligible.* However, we will see that much of the most important "microphysics" of cosmology happens during the RD phase, so it is worth recalling its properties.

The Flat, radiation-dominated Universe. If we go back far enough in a Universe with both matter and radiation, we find that is still flat but dominated by radiation ($k = 0$ and $\Omega_r = 1$), so we have

$$a \propto t^{1/2} \tag{3.24}$$

and therefore

$$\frac{\dot{a}}{a} = \frac{1}{2}t^{-1} \; . \tag{3.25}$$

We will see that this applies to sufficiently early times for our Universe.

Let us put together these different cosmological models onto a single graph, Figure 3.2 (although we'll actually have to use some of the calculations from the next chapter in order to it). Note a few crucial things. All the Universes start out looking like a flat Universe, but diverge depending upon the total density (i.e., the sign of k). Note also that the $k = 0$ case does *not* asymptote to a constant value, but it grows more slowly than $k < 0$ the open case. Finally, note that if we normalize all three curves to the same expansion rate at some particular time (e.g., today), at that time the open Universe is older than the flat Universe, which is older than the closed Universe.

3.2.1 Models with a cosmological constant

MRR 4.7

So far, we've assumed no cosmological constant, but there are many interesting possibilities if we allow $\Lambda \neq 0$. More importantly, this seems to describe the real Universe!

We have already seen by examining our Hubble function $E(z)$ as well as Figure 3.1 that the late-time behaviour in such a Universe is dominated by the cosmological constant. Going back to the Friedmann Equation for this case, we will have

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\Lambda}{3} \; ; \tag{3.26}$$

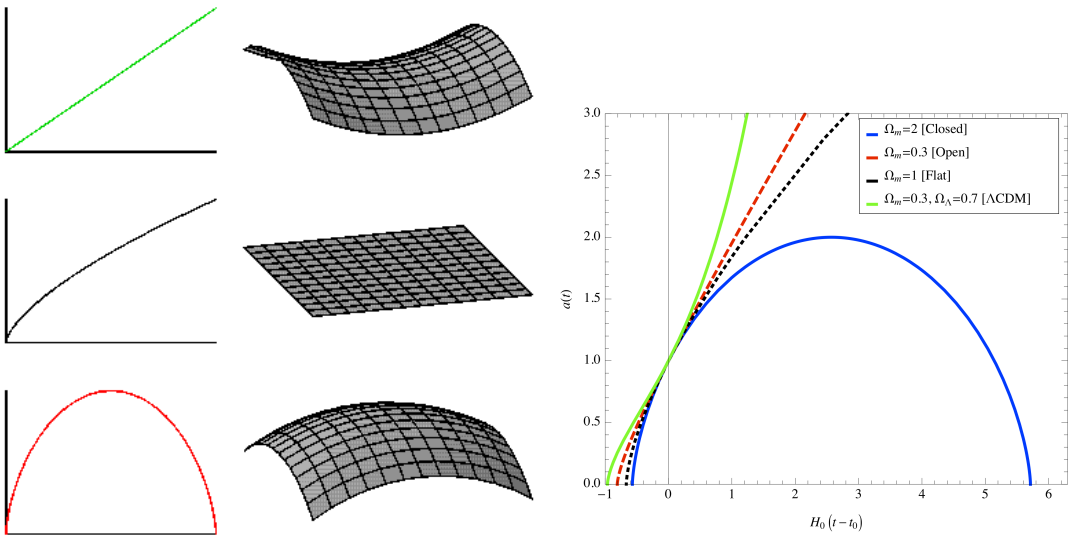

Figure 3.2: Left: The scale factor as a function of time, for MD universes with $k < 0$, $k = 0$, $k > 0$. [Courtesy Ned Wright, `http://www.astro.ucla.edu/~wright/cosmolog.htm`.] Right, the scale factor for various universes, marked by their values of Ω_m (with $\Omega_\Lambda = 0$ except as marked), normalized to have the same scale factor and expansion rate today.

this is a constant — the expansion rate does not evolve with time. This has solution

$$a \propto \exp\left[\sqrt{\Lambda/3}\, t\right] \ . \tag{3.27}$$

Unlike the cases we have looked at so far, which had power-law evolution for $a(t)$, this Universe grows *exponentially*. We will see that this has important consequences.

However, note that the early time behaviour of the Universe is almost completely independent of the cosmological constant: the very early Universe looks like a flat RD Universe, followed by an MD phase.

The Concordance Universe As noted, the real Universe seems to have matter, radiation and a cosmological constant. Roughly,

$$\begin{aligned} \Omega_k &= 0 \\ \Omega_m &\simeq 0.3 \\ \Omega_r &\simeq 10^{-5} \\ \Omega_\Lambda &= 1 - (\Omega_m + \Omega_r) \simeq 0.7 \ . \end{aligned} \tag{3.28}$$

This is occasionally called the "standard cosmological model" or the "concordance cosmology". More details on the current best measurements of these numbers (and others you will encounter in this course) are at `http://lambda.gsfc.nasa.gov`.

Note that in such a multicomponent Universe the behaviour can be quite complicated. In Figure 3.2 we see this "ΛCDM Universe"; note that $a(t)$ starts out decelerating but (around now, $t \simeq t_0$) the Cosmological Constant is beginning to dominate the expansion, getting closer and closer to the exponential expansion of Eq. 3.27. Note also that this enables the Universe to be considerably older than other possibilities, even an open Universe with the same matter density today.

Usually, when we write Ω_i we refer to the value today. But obviously this is just a convenience to refer to quantities that we can measure easily. The density itself is obviously a function of time, and we can define the critical density as a function of time, as well, and so make the Ω_i functions of time or redshift as well:

$$\rho_c(z) = \frac{3}{8\pi G}\left(\frac{\dot{a}}{a}\right)^2 = \frac{3H_0^2}{8\pi G}E^2(z) = \rho_{c,0}E^2(z)\ , \tag{3.29}$$

so we can define

$$\begin{aligned}\Omega_i(z) &= \frac{\rho_i(z)}{\rho_c(z)} = \frac{\rho_{c,0}\Omega_{i,0}(1+z)^{n_i}}{\rho_{c,0}E^2(z)} \\ &= \Omega_{i,0}\frac{(1+z)^{n_i}}{E^2(z)}\end{aligned} \tag{3.30}$$

where we've used a zero subscript to make sure it is evident we are referring to today, $t = t_0$ (but probably will not in the future). For a Universe with radiation, matter and a cosmological constant, this is shown in Figure 3.3. In a flat Universe, the components are $\Omega_i = 0$ or $\Omega_i = 1$ for much of the evolution of the Universe. This is not the case for an open or closed Universe.

In fact the concordance Universe is a fairly "boring" example of a Universe with a cosmological constant. As you will see in the problem sheet, if we balance the densities of the various components just right, we can get a wide variety of behaviours (including the one that led Einstein to his "greatest blunder".)

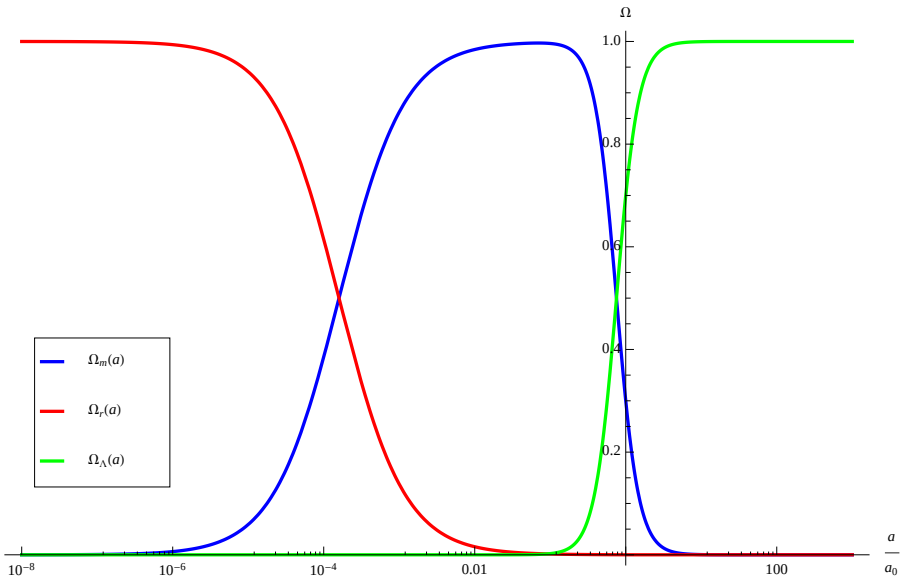

Figure 3.3: The density dependence of the various components of the concordance Universe. Note that the horizontal axis gives the scale factor of the Universe, with $a/a_0 = 1$ today.

Chapter 4

Cosmography

In this chapter we will discuss the description — and measurement — of times and distances in a Friedmann-Robertson-Walker univese. How old is the Universe? How far away is some object that we see? We will see that these questions do not necessarily have unambiguous answers, and that we must very carefully describe what we want to measure.

4.1 The Age of the Universe

MRR 4.9

We want to start pinning down the numbers corresponding to the possible FRW Universes we have been examining. One of the most important is the *age*. We can see how to calculate it from some simple manipulations:

$$t_0 = \int_0^{t_0} dt = \int_{a=0}^{a=a_0} \frac{dt}{da}\, da\, \frac{a}{a} = \int_0^{a_0} da\, a^{-1} \frac{a}{\dot{a}}\,. \tag{4.1}$$

There are many ways to actually do this integral, but one particularly simple one is to use the fact that

$$\frac{a}{a_0} = \frac{1}{1+z} \qquad \text{so } da = -a_0(1+z)^{-2}\, dz \tag{4.2}$$

and we have defined

$$\frac{\dot{a}}{a} = H_0 E(z) \tag{4.3}$$

so

$$\begin{aligned} t_0 &= \int_0^{\infty} \frac{dz}{(1+z)^2}(1+z)\left[H_0 E(z)\right]^{-1} \\ &= H_0^{-1} \int_0^{\infty} \frac{dz}{(1+z)E(z)}\,, \end{aligned} \tag{4.4}$$

where $E(z)$ is the Hubble Function of Eq. 3.12. (Since $E(z)$ depends on the quantity $1+z$ it may be useful to use $1+z$ or $1/(1+z)$ as an integration variable.) *Exercise: Show*

that the time $\Delta t(z_1, z_2)$ between two redshifts z_1 and z_2 is given by the same expression with the limits of integration changed to z_1 and z_2.

Remember that we saw (in a problem sheet) that if we had a constant expansion velocity $\dot{a}$ we would have found an age $1/H_0$, so we can think of the combination

$$H_0 t_0 = \int_0^\infty \frac{dz}{(1+z)E(z)} \tag{4.5}$$

as being entirely due to the change in expansion velocity, which is of course due to the presence of matter in the Universe. For most of the models we will consider, the quantity $H_0 t_0$ is of order one. For example, in a flat, matter-dominated Universe, $H_0 t_0 = 2/3$

Considering Figure 3.2, this is the difference between a tangent line stretching back from today to the horizontal axis compared to the actual curve. In this figure, $a(t)$ always has a negative second derivative (expansion decelerates). This is because we only showed Universes with no cosmological constant, but in fact once we allow $\Omega_\Lambda > 0$, we can have $\ddot{a} > 0$ which can actually make the Universe "older" than it would be with constant $\dot{a}$, as we saw in the previous chapter, Figure 3.2.

This fact has been crucial in our understanding of our actual Universe. We find from measurements that $H_0^{-1} \simeq 14$ Gyr and in fact find objects in the Universe that we believe to be roughly 11–12 Gyr old. Since this gives $H_0 t_0 \simeq 0.8 \gtrsim 2/3$, it seems that the simplest FRW Universe — flat and matter-dominated — cannot be the case! Either our measurements are wrong, or we need the Universe to be dominated by a form of matter which gives a larger value of $H_0 t_0$. But we have seen that this will have the effect of an accelerating scale factor, or a positive value of $\ddot{a}$. This seems a very strange thing when we recall the equation that governs the acceleration of the expansion:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p/c^2\right) + \frac{\Lambda}{3} \tag{4.6}$$

Obviously $\Lambda > 0$ will do this, but it is worth considering the strange-seeming possibility that $\Lambda = 0$ but that the first term is overall positive — which means that somehow $\rho + 3p/c^2 < 0$.

4.2 Horizons

MRR 4.10

Information cannot travel faster than the speed of light, or more precisely causality cannot act faster than the speed of light. This is one of the core principles of physics since the early twentieth century, affirmed in relativity and quantum mechanics alike.

Light rays travel on null geodesics, which in our metric notation corresponds to $ds^2 = 0$. If we assume $d\phi = d\theta = 0$ as in our redshift calculation of Section 2.5.3. We can first ask, at what distance would a light ray have had to start to arrive here at time t? We show a spacetime diagram of this situation in Figure 4.1.

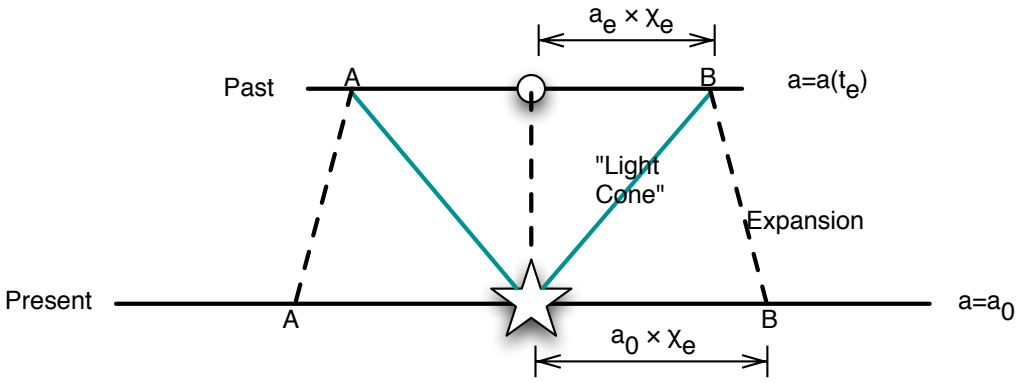

Figure 4.1: Two-dimensional spacetime diagram of horizons in the expanding Universe — space is horizontal, time vertical. The star represents our place of observations, here, today, and the diagonal solid lines represent our past light cone. χ_e represents the comoving coordinate distance (as if measured at a particular time with the scale factor $a = 1$) between us (at the star) and point B on the light cone, so $a_0 \times \chi_e$ gives the proper distance to the present-day location of B. If $t_e = 0$ then $\chi_e = \chi_H$ is the comoving distance to the horizon, and $d_H = a_0 \times \chi_H$ gives the proper distance to the present-day location of our horizon. (It is important also to realize that except for the flat $k = 0$ case, the lines of constant t represent a *curved* universe, and thus there is a distinction between the value of the radial coordinate, r_e, and the comoving distance, χ_e.)

As we have emphasized, whenever we talk about distances, we need to be very precise in our definition. So first, let us recall that the form of the FRW metric is

$$ds^2 = c^2\, dt^2 - a^2(t) d\chi^2 = c^2\, dt^2 - a^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2\, d\Omega^2 \right) \tag{4.7}$$

where $d\chi^2$ gives the metric of a three-dimensional spatial "slice" of spacetime, in comoving coordinates, and $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ gives the angular metric. We need to then use the same equation as for the redshift calculation, for a light ray with $d\theta = d\phi = 0$:

$$\frac{c\, dt}{a(t)} = -\frac{dr}{\sqrt{1 - kr^2}}\,. \tag{4.8}$$

Note that this equation is just $d\chi$, a differential element of comoving distance. As in the redhift case, we put $t = 0$ at $r = r_H$, and $t = t_0$ at $r = 0$ and hence a negative sign appears. We can integrate this along the line of sight from $t = 0, r = 0$ to $t, r = r_H$, after switching the limits of integration because of that negative sign (equivalently, we could just use $r = 0$ at $t = 0$ and $r = r_H$ at t):

$$\int_0^t \frac{c\, dt}{a(t)} = \int_0^{r_H} \frac{dr}{\sqrt{1 - kr^2}} \equiv \chi_H(t, r_H) \tag{4.9}$$

where we define χ_H, the comoving *distance* to the horizon at comoving *coordinate value* r_H. For a flat Universe ($k = 0$) these two are the same. However, for $k \neq 0$, they are

not, and in particular, χ_H corresponds to the comoving distance to the sphere — along a $t = \text{const}$ surface — defined by coordinate r_H. As we discussed in Section 2.5, r does not correspond to a distance along the manifold; for example in the spherical case it is basically the distance perpendicular to the axis defining $r = 0$, or alternately the radius of the two-sphere labelled by r. (To see this, think of the r coordinate as referring to the distance from the axis to a constant latitude circle of a 2-sphere.)

Finally, we convert this coordinate distance to a *proper* distance to the horizon by multiplying by the scale factor, $a(t)$,

$$d_H(t) = a(t)\chi_H(t) = a(t)\int_0^t \frac{c\,dt'}{a(t')}\ . \tag{4.10}$$

We can recast this in a useful form by returning to our Hubble function, $E(z)$, defined by $\dot{a}/a = H_0 E(z)$, so

$$\begin{aligned} d_H(t) &= ca\int_0^a \frac{dt}{da'}\frac{a'}{a'^2}da' = ca\int_0^a \frac{a'}{\dot{a}'}\frac{1}{a'^2}da' = cH_0^{-1}\frac{a}{a_0}\int_z^\infty \frac{dz'}{E(z')} \\ &= \frac{cH_0^{-1}}{1+z}\int_z^\infty \frac{dz'}{E(z')}\ . \end{aligned} \tag{4.11}$$

In particular, the horizon distance today is just this evaluated at $z = 0$. For a flat, $\Omega_m = 1$ Universe, we have $d_H = cH_0^{-1}\int_0^\infty dz/(1+z)^{3/2} = 2cH_0^{-1} = 3ct_0$ where we have done the integral and then used $H_0 t_0 = 2/3$ for this case from above. Note that this is greater than the naive expectation of $d_H \sim ct_0$, the simple distance traveled by a photon in the time from $t = 0$ to $t = t_0$, as d_H is the distance today from $r = 0$ to the present-day position defined by coordinate $r = r_H$. As we see in the Figure 4.1, expansion means that this position is further away than ct_0.

4.3 Distances

It is often useful to be able to describe the distance between different points (events) on a spacetime manifold, but in GR there is usually no unambiguous definition of *the* distance between two points. Rather, we must state very explicitly what we mean. If we could freeze the Universe at a particular time, it might make sense to use the *spatial* coordinate distance between those two points, and indeed that is an easy quantity to work with. However, in practice we can only observe points that are on our "light cone" — at a given redshift we observe objects at a particular coordinate distance from us at a particular time in the evolution of the Universe. We have already seen this in our discussion of the redshift. Since a geodesic (light ray) satisfies $ds \equiv 0$, we have

$$0 = ds^2 = c^2dt^2 - a^2(t)\frac{dr^2}{1-kr^2} \tag{4.12}$$

which we can reorder and integrate along the path from emission to observation (again, we can either define r backward along a light-ray coming toward us or vice versa):

$$\chi_e \equiv \int_{t_e}^{t_0}\frac{c\,dt}{a} = \int_0^{r_e}\frac{dr}{\sqrt{1-kr^2}} = f_k(r_e) \tag{4.13}$$

where we note that the dr integral is a fixed function of r_e (and k):

$$f_k(r_e) = S_k^{-1}(r_e) \equiv \begin{cases} |k|^{-1/2} \sin^{-1}(\sqrt{|k|}\, r_e) & k > 0 \\ r_e & k = 0 \\ |k|^{-1/2} \sinh^{-1}(\sqrt{|k|}\, r_e) & k < 0 \end{cases} \tag{4.14}$$

which just converts to distance along the circle, χ_e, from the "Cartesian coordinate" which r represents as we saw when discussing the geometry of the circle in Section 2.5.1. Recall that we had $k^2 = 1/R$ where R is the radius of curvature (which is the actual radius of the sphere for $k > 0$) so these make dimensional sense.

We can solve this equation for r_e,

$$\begin{aligned} r_e &= \begin{cases} |k|^{-1/2} \sin\left(\sqrt{|k|}\, \chi_e\right) & k > 0 \\ \chi_e & k = 0 \\ |k|^{-1/2} \sinh\left(\sqrt{|k|}\, \chi_e\right) & k < 0 \end{cases} \\ &\equiv |k|^{-1/2} \sin_k\left(\sqrt{|k|}\, \chi_e\right) \\ &= \frac{c}{a_0 H_0} |\Omega_k|^{-1/2} \sin_k\left(\sqrt{|\Omega_k|} a_0 H_0 c^{-1}\, \chi_e\right) , \end{aligned} \tag{4.15}$$

which we write in the unified form defining $\sin_k$ for simplicity (and taking $\sin_k(x) = x$ and cancelling the $\sqrt{|k|}$ factors when $k = 0$ — we can also think of this as the $k \to 0$ limits of the positive and negative k cases). We have used the definition of $\Omega_k \equiv -kc^2/(a_0 H_0)^2$ to eliminate k. We can also put the χ_e integral into a form we can use more simply:

$$a_0 H_0 c^{-1} \chi_e = a_0 H_0 \int_{t_e}^{t_0} \frac{dt}{a} = a_0 H_0 \int_{t_e}^{t_0} \frac{dt}{da}\frac{da}{a} = a_0 H_0 \int_0^{z_e} \frac{dz}{1+z}\frac{a}{\dot{a}}\frac{1}{a} = \int_0^{z_e} \frac{dz}{E(z)} \tag{4.16}$$

[Note: very often, cosmologists will absorb R into the definition of the r coordinate or the scale factor a and take $k = \pm 1$ or $k = 0$. Dimensionally, the form of the FRW metric dictates that kr^2 has no units (from the $1 - kr^2$ factor) and that ar has units of length. This can be realized by giving a units of length and making both k and r unitless, or by making a unitless, giving r units of length, and k units of length^{-2}. There is further freedom of scaling (i.e., units): very often you will see the convention $a_0 = 1$ so that distances are measured as compared to the present day. Note that finally we write these equations in terms of Ω_k, which is unambiguously defined.]

4.3.1 Luminosity Distance

MRR 7.6

From this light-cone coordinate distance we can make our distance measures yet more physical by relating them to a possible observation. For example, imagine that we have an object whose luminosity (energy emitted per unit time) is given by L. On Earth, we measure a flux $\mathcal{F}$ from this object which in (flat, non-expanding) Euclidean space is given

by $\mathcal{F} = L/(4\pi d^2)$ for an object at distance d. We can use this to define the "luminosity distance", d_L, in a curved and expanding spacetime:

$$\mathcal{F} = \frac{L}{4\pi d_L^2} \tag{4.17}$$

where d_L can only depend upon the redshift once we have fixed the spacetime. To calculate it, consider a spherical coordinate system with the origin at the source, $\mathcal{S}$, at radial coordinate r_e (so that our observation point is on the sphere), emitting at time t_e, observed at t_0. The total area of the sphere that goes through our observation point at t_0 is

$$A = \int a_0^2 r_e^2 d\Omega = a_0^2 r_e^2 \int d\cos\theta \, d\phi = 4\pi a_0^2 r_e^2 \; . \tag{4.18}$$

This equation uses the scale factor *today*, a_0, because this gives the proper, physical, area of the sphere at t_0. Because of the expansion of the Universe and the redshift, the total rate of energy coming through the sphere is modified from its rate at emission by two effects. First, the number of photons coming through the sphere per unit time is decreased by $a_0/a_e = 1 + z_e$. Second, each of those photons is doppler shifted so its energy is decreased by an *additional* factor of $a_0/a_e = 1+z_e$. Note that these are *different* effects, and hence the energy per unit time is decreased by $(1 + z)^2$. (Another way to think of this is that the total energy, integrated over all time, would be decreased by one factor of $1 + z$ due to redshift, but that energy is spread over a time that is longer by another factor of $1 + z$.) Hence,

$$\mathcal{F} = \frac{L}{4\pi a_0^2 r_e^2 (1+z)^2} \tag{4.19}$$

so that

$$d_L = a_0 r_e (1 + z) \; . \tag{4.20}$$

Now, we can combine this with our solution from above for r_e (Eqs. 4.15-4.16) to give

$$d_L(z_e) = cH_0^{-1}(1 + z_e)|\Omega_k|^{-1/2} \sin_k \left(\sqrt{|\Omega_k|} \int_0^{z_e} \frac{dz}{E(z)} \right) \quad . \tag{4.21}$$

Note in particular that for $k = 0$,

$$d_L = cH_0^{-1}(1 + z_e) \int_0^{z_e} \frac{dz}{E(z)} \qquad (k = 0) \; . \tag{4.22}$$

These expressions do not depend on the characteristics of the source (L) nor on the measured flux, but do depend on the cosmological parameters through the factors of H_0 and $E(z)$ in the integral, as well as k itself.

4.3.2 Angular-Diameter Distance

MRR 7.8

Now, consider a slightly different measurement. Instead of an object of known luminosity, consider one of known *size* (for example, perhaps there is some galaxy that we know is 10 kiloparsecs in diameter). Of course from here we measure not the size, but the angle that it subtends on the sky. In Euclidean geometry, the diameter D is related to the observed angle θ if we know its distance, d, by $\theta = D/d$, which motivates the definition of the *angular-diameter distance* in more general circumstances:

$$d_A = \frac{D}{\theta} \, . \tag{4.23}$$

From the form of the FRW metric (consider the $d\theta$ term), the proper distance along a circle (assuming $\theta \ll 1$ so we need not distinguish between the chord and the arc) at radius r_e should be $D = r_e a_e \times \theta$, so

$$d_A = r_e a_e = a_0 r_e \frac{a_e}{a_0} = \frac{a_0 r_e}{1+z_e} \, . \tag{4.24}$$

Note that this equation uses the scale factor at *emission*, a_e, since it is the proper circumference of the circle at t_e that matters (compare the luminosity-distance calculation above, Eq. 4.18, in which the proper area of a sphere at t_0 entered). As above, we substitute in our calculation of r_e to get

$$d_A(z_e) = cH_0^{-1}(1+z_e)^{-1}|\Omega_k|^{-1/2} \sin_k \left(\sqrt{|\Omega_k|} \int_0^{z_e} \frac{dz}{E(z)} \right) \, . \tag{4.25}$$

Note that $d_A = d_L/(1+z)^2$, independent of the cosmological parameters. Further, the observable quantity, the *surface brightness*, defined by the flux per solid angle on the sky

$$\mathcal{S} = \frac{\mathcal{F}}{\theta^2} = \frac{L}{4\pi D^2} \frac{1}{(1+z)^4} \tag{4.26}$$

is independent of the cosmology (for fixed physical size D and luminosity L).

In Figure 4.2 we show the behaviour of the luminosity and angular-diameter distances as a function of redshift. Note that the angular diameter distance actually turns over: this means that objects actually can start getting bigger as they get further away! Unfortunately, due to the dimming of the surface brightness (Eq. 4.26), they become harder and harder to see, very rapidly, as the energy is spread out over a wider and wider area of sky.

4.3.3 The Extragalactic Distance Ladder

See the in-class presentation.

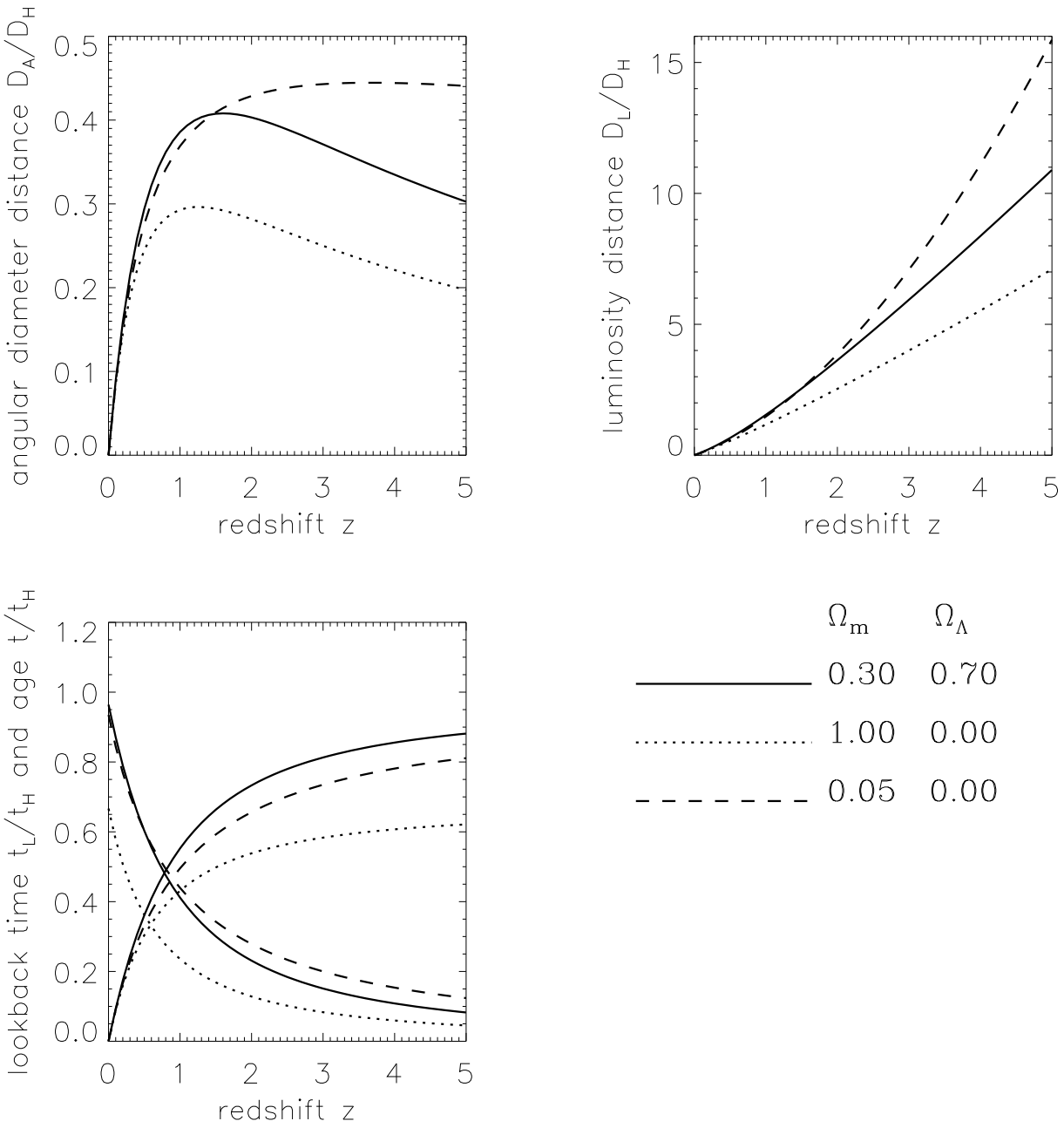

Figure 4.2: Distance measures and the age of the Universe in different cosmologies. Distances are measured in units of c/H_0 and times in units of $1/H_0$.

Chapter 5

Thermodynamics and Particle Physics

MRR 5.1–5.2

Today, the evolution of the Universe is transitioning from being dominated by nonrelativistic matter to dark energy, while the temperature of the Cosmic Microwave Background is $T \simeq 2.73$ K (there are more energetic photons produced by astrophysical processes as well). Because of this cold temperature, particle energies are low and interactions are rare (except in places like stars, or CERN, of course). As we trace the evolution of the Universe backwards, however, it heats up and the mean particle energy increases, so interactions become more and more likely. Hence, we need to understand the relationship between thermodynamics and particle physics in the early Universe, sometimes called the "primeval fireball".

5.1 Radiation Domination

We have already seen that the density of radiation scales as $\rho_r \propto a^{-4}$ whereas that of nonrelativistic matter scales as $\rho_m \propto a^{-3}$ as the universe expands with scale factor $a(t)$ [and recall that $a/a_0 = 1/(1+z)$ relates redshift to the scale factor]. Today, the radiation density is roughly 10^{-4} that of matter, but in the past it would have been higher, and at some point the Universe must have been radiation dominated.

In the absence of any processes interconverting matter and radiation, this would have occurred at a time $t_{\rm eq}$ defined by

$$\rho_m(t_{\rm eq}) = \rho_r(t_{\rm eq}) \tag{5.1}$$

or

$$\begin{aligned} \rho_m(t_0)\,(1+z_{\rm eq})^3 &= \rho_r(t_0)\,(1+z_{\rm eq})^4 \\ \Omega_m \rho_c\,(1+z_{\rm eq})^3 &= \Omega_r \rho_c\,(1+z_{\rm eq})^4 \\ 1+z_{\rm eq} &= \frac{\Omega_m}{\Omega_r} \simeq \frac{0.3}{8\times 10^{-5}} \simeq 3,700\ . \end{aligned} \tag{5.2}$$

(This number is probably uncertain to 10-20%.) At redshifts higher than $z_{\rm eq}$, the Universe was radiation dominated. At this time, we have $T_{\rm eq} = (1+z_{\rm eq})T_0 \sim 10^4$ K or $kT_{\rm eq} \sim 1$ eV.

5.2 Black Body radiation and equilibrium statistics

The cosmic microwave background (CMB) is almost a perfect black body, as we see from the spectrum in Figure 5.1. As we will see in a problem sheet, the effect of redshifting upon a black body distribution of photons is just a black body at the new temperature $T(z) = (1 + z)T_0$. That is, if we observe photons from a black body at temperature T_0 today, they would have been at $(1 + z)T_0$ at redshift z.

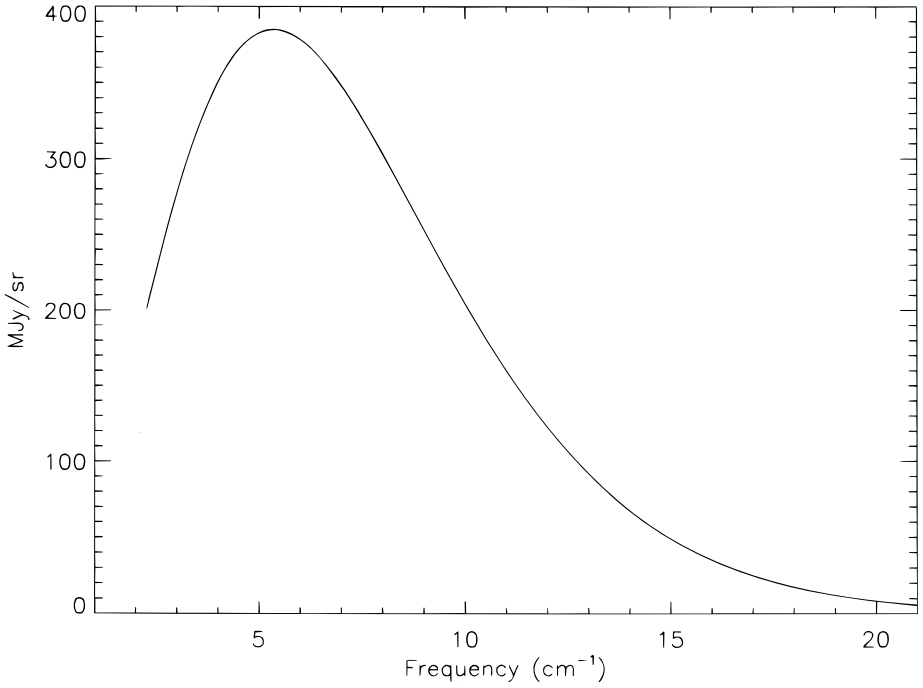

Figure 5.1: The measured spectrum of the Cosmic Microwave Background, plotted along with a black body curve with $T = 2.728$ K. Uncertainties are a small fraction of the line thickness. From Fixsen et al, *Astrophys. J.*, **473**, 576 (1996).

The blackbody distribution of photons of frequency ν is given by the Planck function:

$$n(\nu)\, d\nu = \frac{8\pi}{c^3} \frac{\nu^2\, d\nu}{\exp(h\nu/kT) - 1} \,. \tag{5.3}$$

Here, h is Planck's constant so $h\nu$ is the photon energy, and the energy distribution is therefore given by $\varepsilon(\nu)\, d\nu = (h\nu)n(\nu)\, d\nu$.

The total number density (in a *proper* volume) of CMB photons is

$$\begin{aligned} n_\gamma &= \int n(\nu)\, d\nu = \left(\frac{kT}{hc}\right)^3 8\pi \int_0^\infty \frac{y^2\, dy}{e^y - 1} \\ &= \left(\frac{kT}{hc}\right)^3 16\pi\zeta(3) \\ &\simeq 413.\ \mathrm{cm}^{-3} \end{aligned} \tag{5.4}$$

where $\zeta(3) \simeq 1.202$ is the Riemann zeta function, and we have evaluated the expression at $T_0 = 2.73\,\mathrm{K}$ in the last equality. Note that the number density is proportional to $T^3 = (1+z)^3 T_0^3$ — as we already knew, the proper number density scales with the expansion of the Universe, and hence the comoving density is constant.

We can compare the present-day density to the average number density of baryons in today's Universe, using the measured value of $\Omega_b h^2 \simeq 0.023$:[1]

$$n_b = \Omega_b \rho_c / m_b \simeq \frac{0.023 \times 1.9 \times 10^{-29}\ \mathrm{gcm}^{-3}}{1.67 \times 10^{-24}\ \mathrm{g}} \simeq 2.6 \times 10^{-7} \mathrm{cm}^{-3}\ . \tag{5.5}$$

There are

$$\eta \equiv \frac{n_b}{n_\gamma} \simeq \frac{3 \times 10^{-7}}{400} \sim 10^{-9} \tag{5.6}$$

baryons for every photon. This small number — which as we will see has been roughly constant since the first minutes after the big bang — will have many important effects on the present-day abundances of the elements.

Massive Particles More generally, of course, the equilibrium distribution of a particle's energy is given by the Bose-Einstein or Fermi-Dirac distribution for integer or half-integer spins, respectively. The number density of particles per mode is given by

$$n(\mathbf{p}) d^3p = \frac{1}{e^{(\epsilon(p)-\mu)/kT} \pm 1} \frac{d^3p}{h^3} \tag{5.7}$$

with the plus sign for fermions and the minus sign for bosons. The function tells us the distribution of particle momenta, related to the energy per particle $\epsilon^2 = m^2c^4 + p^2c^2$. For photons, to convert this into the Planck distribution function, Eq. 5.3, $4\pi\nu^2\, d\nu = d^3\nu = (c/h)^3\, d^3p$ from the density of states — the accessible volume of three-dimensional momentum space—and the remaining factor comes from the $g = 2$ polarization (helicity) states available.

To use this to calculate the overall number density, it is easiest to convert this expression into an integral over energy:

$$n = g \int n(\mathbf{p})\ d^3p = 4\pi g \int n(\mathbf{p}) p^2\ dp = \frac{4\pi g}{c^3 h^3} \int \frac{\sqrt{\epsilon^2 - m^2c^4}}{e^{(\epsilon-\mu)/kT} \pm 1} \epsilon\ d\epsilon \tag{5.8}$$

[1] In this expression, h is the Hubble constant in the form $H_0 = 100h\mathrm{km/\,sec\,/Mpc}$. You will very often see densities expressed in the form $\Omega_i h^2$ since $\Omega_i = 8\pi G\rho_i/(3H_0^2)$ and this form lets us take into account our remaining ignorance of the Hubble constant.

where g is the number of degrees of freedom for the particle species (e.g., $g = 2$ for spin-1/2 fermions). We can similarly calculate the energy density

$$\varepsilon = g \int \epsilon(p) n(\mathbf{p})\, d^3p \tag{5.9}$$

as well as quantities like the average pressure (and hence the equation of state relating density, temperature and pressure).

Note that if $\epsilon \gg mc^2$, the energy is dominated by the momentum term, just like for photons. However, the majority of particles in either the BE or FD distribution have energy $\epsilon \sim kT$. Hence, if $kT \gg mc^2$, the mass of the particles are irrelevant — they behave like radiation, and this is true for both bosons and fermions. (Strictly speaking, we are assuming the chemical potential is much less than kT as well.)

In this relativistic limit, when we also assume $\mu \ll m$ (non-degenerate), we find

$$\begin{aligned}
\varepsilon &= g\frac{\pi^2}{30(\hbar c)^3}(kT)^4 \times \begin{cases} 7/8 & \text{Fermions} \\ 1 & \text{Bosons} \end{cases} \\
n &= g\frac{\zeta(3)}{\pi^2(\hbar c)^3}(kT)^3 \times \begin{cases} 3/4 & \text{Fermions} \\ 1 & \text{Bosons} \end{cases} \\
p &= \frac{\varepsilon}{3}
\end{aligned} \tag{5.10}$$

where the last line gives the pressure, and we use $\hbar = h/(2\pi)$. The ratio of the energy density to the number density gives the average energy per particle,

$$\langle E \rangle = \varepsilon/n \simeq \begin{cases} 3.15\, kT & \text{Fermions} \\ 2.70\, kT & \text{Bosons} \end{cases} \tag{5.11}$$

which we will very often just take to be $\langle E \rangle \sim 3\ kT$.

This has important cosmological consequences. At earlier times, the temperature was higher by the usual redshift factor $T = (1+z)T_0$. For any particle of mass m, there is a redshift such that $kT > mc^2$, before which the particle essentially behaves like radiation. Hence, as we go backwards in time, the number density — and hence energy density — of particles behaving like radiation (for example, having equation of state $p = \varepsilon/3$) increases. Because the expansion of the Universe depends on the density of radiation-like and non-relativistic-matter-like particles separately, we need to take this into account when describing the expansion of the Universe at early times. The total energy density requires summing Eq. 5.10 for ε over all relativistic particle species i, which as we will see may each have a different temperature T_i, giving

$$\varepsilon_r = g_*(T)\frac{\pi^2}{30(\hbar c)^3}(kT)^4 \tag{5.12}$$

where we define the effective degrees of freedom as

$$g_* = \sum_{i \in \text{bosons}} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{i \in \text{fermions}} g_i \left(\frac{T_i}{T}\right)^4 \tag{5.13}$$

(It is conventional to use $T = (1+z)T_0$, the temperature of the photons.)

Once the temperature drops below $kT = mc^2$, however, the behaviour of the particle changes. In this limit, we can calculate the number density:

$$n_P = g\left(\frac{mkT}{2\pi\hbar^2}\right)^{3/2} e^{-(mc^2-\mu)/kT} \tag{5.14}$$

This is the number density in a *proper* volume element. If we want to know the number density in a *comoving* volume element, we must rescale by the ratio of the comoving to proper volume, proportional to $a^3 \propto (1+z)^{-3} \propto T^{-3}$.

In Figure 5.2 we combine these and show the comoving density as a function of temperature (and hence redshift, since $T = (1+z)T_0$) for such a particle, assuming $\mu = 0$ (no degeneracy). For $mc^2 \ll kT$, this is a constant — the particle behaves like radiation. For $mc^2 \gg kT$ the proper density decreases exponentially as Eq. 5.14. (Numerical integration of the Fermi-Dirac or Bose-Einstein distributions is required to interpolate between the two limits.)

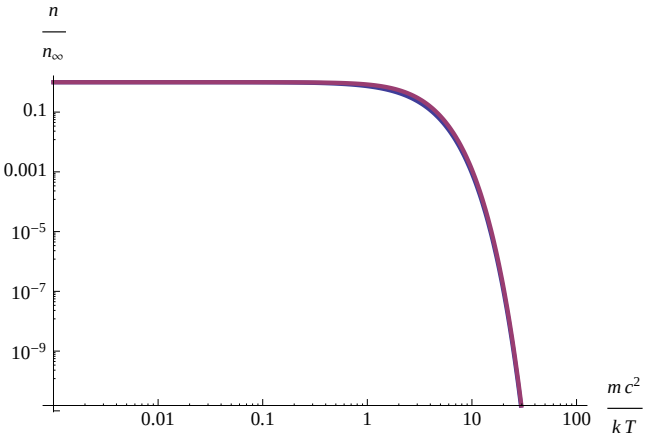

Figure 5.2: The equilibrium comoving number density of a non-degenerate particle species as a function of temperature. We plot the density relative to the number density at $T \to \infty$ (more properly, the $m \to 0$ limit of Eq. 5.8) to remove the overall dependence upon the particle properties and temperature. Both BE and FD densities are shown, but they are nearly indistinguishable.

Note that these distributions apply to particle species in equilibrium; hence we must take into account the interactions that these particles experience. Indeed, this will become crucial as we discuss the relics of the hot early Universe: the light elements, the microwave background itself, and the dark matter. If all of the particles in the Universe were in equilibrium, the abundances of, say, Hydrogen atoms, would be *much* less than that observed – using $m = m_p$ and $T = T_0$ in Eq. 5.14 gives $n_P \sim 10^{-10^{12}}$ cm^{-3}!

5.3 The Hot Big Bang

5.3.1 Interactions

As the universe expands, its state is determined by the temperature, $T = (1+z)T_0$, and the expansion rate, $H = \dot{a}/a$. In a radiation-dominated Universe ($a \propto t^{1/2}$), the latter is

$$H = \frac{\dot{a}}{a} = \frac{1}{2}\frac{1}{t} \propto a^{-2} \propto (1+z)^2 \,. \tag{5.15}$$

Any interactions between particle species are governed by

- Interaction rates, Γ (with units 1/time), which may depend on the temperature, particle masses, and cross sections; and
- Particle masses, which determine the equilibrium abundances.

From above, we know that when $kT \gg mc^2$, the particle acts like radiation, so $E \simeq pc$. If the interaction with photons is sufficiently strong, it can interconvert with photons and will have the same temperature.

The interaction rate Γ tells us the mean time between interactions, $1/\Gamma$, or the mean free path $\lambda \sim v/\Gamma \sim c/\Gamma$ (since the particles behave like radiation with $v = c$). Our calculation of the horizon distance, $d_H \sim ct \sim c/H$ gives us the maximum distance between particles in causal contact, hence when $\lambda \gg d_H$ we expect that interactions will be unable to maintain thermal equilibrium.

To see this another way, the number of interactions that a particle species has from time t onward is

$$N_{\rm int} = \int_t^\infty \Gamma(t')\, dt' \,. \tag{5.16}$$

If we take the typical case that $\Gamma \propto T^n$, then, when we also have $a \propto t^m$ ($m = 1/2$ for RD and $m = 2/3$ for MD), we find that

$$N_{\rm int} = \frac{1}{n - 1/m}\left(\frac{\Gamma}{H}\right)_t \tag{5.17}$$

(Exercise: Show this.) The average number of interactions experienced by a particle after time t is therefore less than one if $\Gamma < H$, as long as $n > m$.

By both of these arguments, or by our more detailed discussion of the Boltzmann equation below, when $\Gamma \gg H$, each particle will experience interactions, and can plausibly remain in thermal equilibrium. However, if $H \gg \Gamma$, the Universe expands away more rapidly than the particles can interact, and most particles do not experience the interaction, and hence fall out of equilibrium. We call this departure from equilibrium *freeze-out*, as the comoving abundance of a species that has $\Gamma \ll H$ is frozen (in the absence of any further interactions, of course). Note that a species can freeze out either when it is behaving like radiation ($kT \gg mc^2$) or matter ($kT \ll mc^2$). In the former (relativistic freeze-out) we say that the outcome is a *hot relic*, and in the latter (massive freeze-out) a *cold relic*.

5.3.2 Thermal history of the Universe

Even before taking detailed account of the interactions between different sorts of particles, we have enough information to put together a timeline for the Universe. As we go further *back* in time, the Universe gets hotter and hotter as $T = (1 + z)T_0$. As the temperature rises, the average energy of particles increases, allowing more and more kinds of interactions to take place (just as higher energies are needed in particle accelerators to access more interactions). Hence, at the earliest times all possible particles were relativistic, and if interactions were strong enough, they remained in thermal equilibrium. As the Universe cools (i.e., thinking forward in time), particles become non-relativistic and possibly fall out of equilibrium, possibly leaving around relics that we see today.

kT (GeV)	t (s)	
10^{19}	10^{-43}	The Planck Era: Classical GR breaks down
$10^{14\text{–}16}$	$10^{-(35\text{–}38)}$	Thermal Equilibrium established; GUT transition? ($M_X \sim 10^{15}$ GeV)
10^{2}	10^{-11}	Electroweak phase transition (M_W)
0.1–0.5	$10^{-(5\text{–}6)}$	Quark confinement ("chiral symmetry breaking")
?	?	Baryogenesis
10^{-1}	10^{-4}	$\mu^+\mu^-$ annihilation (and freeze-out)
?	?	Dark Matter interactions freeze out?
10^{-3}	1	ν decoupling
5×10^{-4}	3–4	e^+e^- annihilation, leaves mainly γ, ($\nu\bar{\nu}$) separately in equilibrium
10^{-4}	180	Nucleosynthesis $\Rightarrow$ He^4, D, T, Li
$10^{-(8\text{–}10)}$	$10^{10\text{–}11}$	Matter Domination
$10^{-(10\text{–}11)}$	$10^{11\text{–}13}$	H recombination ($e + p \rightarrow \mathrm{H} + \gamma$)
		Universe becomes neutral and transparent

Table 5.1: The history of the Universe.

In Table 5.3.2 we show some of the most important "events" in the history of the Universe. Some terms, such as "chiral symmetry breaking" and "GUT Transition" come from particle physics and will not be discussed much further here. Others, such as **decoupling** and **freeze-out** refer to what happens when interactions fall out of equilibrium and are crucial to the history of the Universe. I have left a few spaces filled with question marks, because they depend to some extent on the as-yet unknown physics of the dark matter and of the creation of the present-day matter/antimatter asymmetry.

5.4 Relic Abundances

As we noted above, if all of the particles in the Universe were in equilibrium, it would be very boring indeed. But there is ample evidence that many interactions are no longer in equilibrium:

- Different particle species have different temperatures;

- There is considerably more matter than antimatter;
- There are nucleons rather than quarks;
- There are atoms rather than ions; and
- There is structure, rather than a smooth featureless gas at a single temperature.

How did the Universe get to filled with clumpy, baryonic, neutral, matter? To put the question another way: Why is the Universe interesting?

5.4.1 Baryogenesis and the Sakharov Conditions

One of the most basic of these questions regards the difference between matter and antimatter. If there were copious amounts of antimatter in the Universe, we would see it as a background of gamma rays from electron-positron annihilation (as well as more massive particles and antiparticles). The lack of such a strong background puts stringent limits on the fraction of antimatter in the Universe, many orders of magnitude below a 1:1 ratio.

So how do we achieve this asymmetric state? Of course, one possibility is just "initial conditions": the asymmetry is somehow built into the big bang. We would prefer a dynamical solution, one that starts from a more symmetric state but somehow ends up with the highly asymmetric Universe in which we live.

But in fact, the Universe is not as asymmetric as the complete lack of antimatter makes it seem. Consider again our determination earlier of the baryon-to-photon ratio, $\eta \sim 10^{-9}$. For every billion photons, baryon, there is roughly one baryon or, equivalently (from the overall charge-neutrality of the Universe) roughly one electron. When the Universe was hot enough that electrons and positrons were considered radiation (so $kT \gg m_e c^2 \simeq 0.511$ MeV), photons and electrons/positrons would have been in equilibrium with roughly equal number densities. If there were exactly equal numbers of positrons and electrons, when the Universe cooled enough so that electrons and positrons annihilated into photons, the net electron number (and hence baryon number) would have been essentially zero. But if instead there were just one extra electron for every 10^9 positrons, that electron (and its associated baryon) would have been left over, with each of the matching electron/positron pairs creating a pair of photons via $e^+e^- \to \gamma\gamma$. Hence, the complete present-day asymmetry between matter and antimatter, is actually only due to a one-in-a-billion asymmetry at earlier times.

The question of how to arise at even this small asymmetry was considered by the Soviet physicist Andrei Sakharov.[2] He realized that there were three conditions that must hold in order to convert a matter/antimatter-symmetric Universe into one dominated by matter:

[2]Sakharov was known as the "father of the Soviet nuclear bomb", but in later life was a dissident, protesting against the regime. More importantly for our purposes, he also dabbled in astrophysics and cosmology.

1. Baryon number violation. This is the most obvious condition: particle physics must allow interactions which change the number of baryons. So, for example, we must allow interactions like
$$X + \bar{Y} \to B \tag{5.18}$$
where X is some particle, $\bar{Y}$ some antiparticle (so the left hand side has net baryon number $B = 0$), and B represents some state of particles with net baryon number $B \neq 0$.

2. CP violation. If CP is not violated, then in quantum field theory for any interaction, the related interaction with all particles replaced by antiparticles and vice versa will have exactly the same rate. That is, the rates of the following two interactions must be *different*:
$$\begin{aligned} X + \bar{Y} &\to B \\ \bar{X} + Y &\to \bar{B} \, . \end{aligned} \tag{5.19}$$
Any net change in baryon number (allowed by condition 1) will otherwise be cancelled by a change in antibaryon number.

3. Finally, there must be a departure from thermal equilibrium. In thermal equilibrium, the rate of a reaction going forward ($X + \bar{Y} \to B$) will be the same as the rate of the reaction going backward ($B \to X + \bar{Y}$), again cancelling out any change in baryon number.

These are the **Sakharov conditions**.

In fact, we know that all three of these conditions do occur. First, although the standard model of particle physics conserves baryon number at so-called "tree level" (Feynman diagrams) there are non-perturbative quantum effects that can violate baryon number and change the amount of matter relative to antimatter. However, it is not clear that this idea, known as *electroweak baryogenesis* actually gives us a Universe with the correct baryon number. More generally, most extensions to the standard model (e.g., supersymmetry, grand unified theories) violate baryon number in other ways, such as by the decay of new, massive particles.

Second, it has been known since the 1950s that the standard model also exhibits CP violation. In this case, it seems clear that on its own this is insufficient to account for the present-day matter/antimatter asymmetry.

Finally, the expansion of the Universe itself provides the opportunity for departures from thermal equilibrium. As we discussed above, and in more detail very soon, when the mean time between interactions becomes long compared with the expansion rate of the Universe, interactions cannot happen often enough to maintain equilibrium.

The specifics of these conditions are needed to produce the observed baryon asymmetry. However, there are equivalent conditions for producing any of the other "interesting" features of the Universe above: the physics must allow the asymmetry to be produced, to not be cancelled out, and to stay away from the naive abundance dictated by thermal equilibrium. In fact, baryogenesis is probably the most difficult case, as we are not

completely sure where to get the correct baryon number and CP violation, whereas the physics underlying the preference of, say, nuclei over quarks, and neutral atoms over ions, is very well understood.

5.4.2 Interaction rates and the Boltzmann Equation

Even when such conditions are satisfied, in order to calculate the abundance of some particle species in more generality, we need to consider its interactions in detail. We start with the distribution function, $f(t, \mathbf{x}, \mathbf{p})$, the time-dependent density of particles in position and momentum (there is a relativistic generalization of this, of course). We have already written down the momentum distribution for particles in equilibrium in Section 5.2 above, but now we want to calculate it more generally. The tool for this is the **Boltzmann Equation**. In its most general form, the Boltzmann equation merely says that the only interactions can change the distribution function:

$$\frac{Df_i}{Dt} = C\left[\{f_j\}\right] \tag{5.20}$$

where Df_i/Dt is the Liouville operator, just the total time derivative of the distribution function for particle species i, and the quantity $C[\{f_j\}]$ represents the effect of interactions, which can possibly depend on the distribution function of *all* other species, j. In general relativity, the d/dt operation can actually be taken to include the effects of gravity, more or less automatically by correctly accounting for the coordinates that we use to describe the manifold. Even with Newtonian gravity, it is still easiest to account for its effects on the left-hand side, through the Liouville operator. Indeed, we already know how to do this for the FRW metric.

We will still assume an FRW Universe homogeneous and isotropic and any given time. Hence, we must take into account the expansion of the Universe. Basically, the equation must account for the fact that, in the absence of interactions, the *comoving* number density of particles is conserved. In fact we already know an equation that describes this: it is just our non-relativistic matter conservation equation (the fluid equation), Eq. 2.19:

$$\frac{dn}{dt} + 3\frac{\dot{a}}{a}n = 0 \tag{5.21}$$

where n gives the number density. That is, in this context, the Liouville operator $D/Dt = d/dt + 3\dot{a}/a$. What about the right-hand side, the collision operator C? Obviously, it depends on the details of the interactions that we're considering.

In thermal equilibrium, we know that the number density should obey the equilibrium distributions discussed in Section 5.2.

Consider a two-particle to two-particle interaction (such as annihilation $x + \bar{x} \to y + \bar{y}$ or recombination $e + p \to H + \gamma$). In equilibrium, the number of forward interactions is the same as inverse interactions. The forward rate will be $\sigma|v|n^2$, where σ gives the cross section and $|v|$ the relative velocity of the particles. Via *detailed balance*, the inverse rate must be such that the collision term is zero when $n = n_{\rm eq}$, the number density

in equilibrium. Hence the inverse rate must be proportional to $\sigma|v|n_{\rm eq}^2$. Thus we have heuristically shown that we can write the whole equation as

$$\frac{dn}{dt} + 3\frac{\dot{a}}{a}n = -\langle\sigma|v|\rangle\left(n^2 - n_{\rm eq}^2\right) \tag{5.22}$$

where $\langle\sigma|v|\rangle$ is the so-called thermally-averaged value of the cross section times the velocity, and (so if $n = n_{\rm eq}$ it is equivalent to having no interactions). We can rewrite this as

$$\frac{\dot{n}}{n_{\rm eq}} + 3\frac{\dot{a}}{a}\frac{n}{n_{\rm eq}} = -\langle\sigma|v|\rangle n_{\rm eq}\left(\frac{n^2}{n_{\rm eq}^2} - 1\right) \tag{5.23}$$

or

$$\frac{\dot{n}}{n_{\rm eq}} = -\Gamma\left(\frac{n^2}{n_{\rm eq}^2} - 1\right) - 3\frac{\dot{a}}{a}\frac{n}{n_{\rm eq}} \tag{5.24}$$

where the interaction rate is $\Gamma \equiv \langle\sigma|v|\rangle n_{\rm eq}$. The time or temperature dependence of Γ depends on the interaction under consideration, so the solution to this equation usually needs to be calculated numerically. We show an example in Figure 5.3. In general, higher Γ (higher $\langle\sigma|v|\rangle$) results in *later* departure from equilibrium (*freeze-out*). As we noted above, a particle that freezes out when $kT \gg mc^2$ is called a *hot relic*, and one that freezes out when $kT \ll mc^2$ is a *cold relic*.

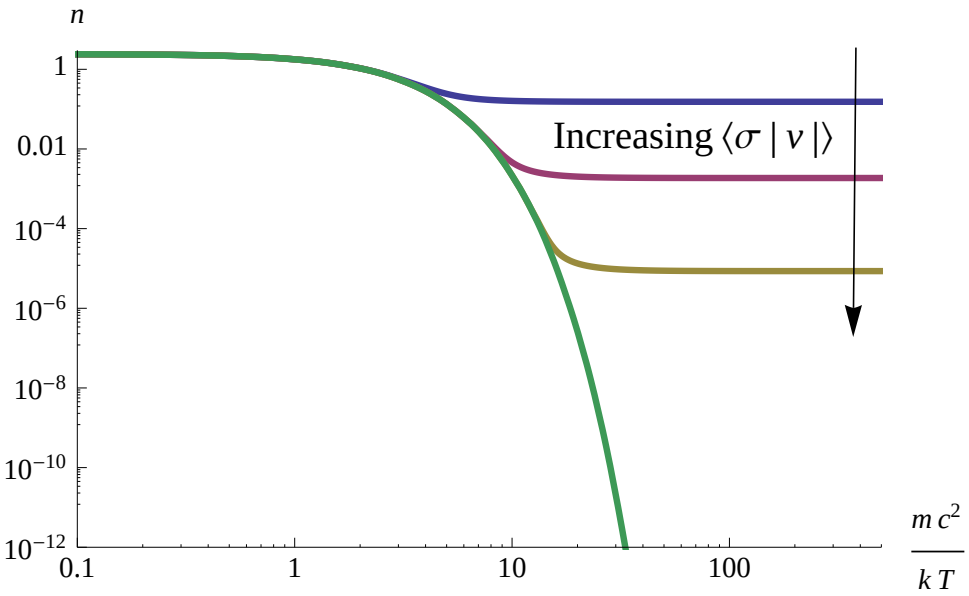

Figure 5.3: Freeze-out of particle species as a function of interaction strength, $\langle\sigma v\rangle$. The exponentially-decaying curve is the equilibrium abundance; the lines that "peel off" from equilibrium correspond to increasing values of $\langle\sigma v\rangle$ for lower values of the final abundance and the freeze-out temperature (so later freeze-out time).

As an example, we show the numerical solution to this equation for the case where $\langle\sigma v\rangle \propto (kT/mc^2)^n$ for $mc^2 \gtrsim 3kT$ for various values of the cross-section. We see that

higher values of the cross section (higher Γ) correspond to *later freeze-out times* and *lower abundances*. Examining the numerical solution also shows that defining the freeze-out temperature from $\Gamma(T_F) = H(T_F)$ and the final abundance from $n_F = n_{\rm eq}(T_F)$ is a very good approximation.

Hot relics: Neutrinos

An excellent case study is given by neutrinos. They were certainly relativistic over the course of many of the events listed in Table 5.3.2, although over the last decade it has become evident that neutrinos have a nonzero mass, likely to be considerably less than 1 eV. *Exercise: what mass would a neutrino need in order to still be relativistic today, when $T = 2.73$ K?* Such a mass (assuming the cross sections of the standard model) means that they freeze out when still relativistic and so are a hot relic.

Neutrinos have very important differences from other matter particles such as electrons and positrons on the one hand, and radiation like photons on the other, however. Like electrons and positrons, they are fermions and moreover have both particles and antiparticles. Unlike electrons, however, they interact very weakly and hence came out of equilibrium much earlier, before they could annihilate into other light particles. Hence, the total neutrino plus antineutrino density is higher than that of electrons plus positrons. (Also, if they have sufficiently low mass, they could *still* behave like radiation.) Because of their very weak interactions with normal matter, however, when the electron-positron pairs annihilated, they produced photons, but not neutrinos. Hence, there was considerable extra energy dumped into the photon background — the CMB — but not the neutrino background, and therefore the CMB actually has a higher temperature than the neutrino background. As you will see in a problem set, the temperature of the neutrino background relative to that of the photons can be calculated, and is

$$T_\nu = \left(\frac{4}{11}\right)^{1/3} T_{\rm CMB} \ . \tag{5.25}$$

Their abundance will then be the value of the equilibrium number density, Eq. 5.10, evaluated at the freeze-out temperature $T = T_F$, and then converted to a comoving density, which is then constant from then on. Note that the conversion to comoving density requires multiplying by $a^3/a_0^3 = (1+z_f)^{-3} = T_0^3/T_F^3$, which cancels out the T_F^3 factor in n;[3] the comoving density depends on g but nothing else. If these particles have sufficient mass that they are non-relativistic today we can simply calculate their energy density by multiplying the number density by the mass. This gives

$$\Omega_\nu h^2 \simeq 10^{-2} \left(\frac{m_\nu}{1\ {\rm eV}}\right) \ . \tag{5.26}$$

Since we know that $\Omega \lesssim 1$ today, this translates into a cosmological bound on the neutrino mass, or more precisely, the sum of all the neutrino masses: $\sum m_\nu \lesssim 100$ eV.

[3]There is an additional factor due to any further relativistic particles that have contributed to the present-day photon abundance and thus T_0 but not to T_F, as in the discussion of e^+e^- annihilation.

Although the freeze out temperature of neutrinos turns out not to be critically important for their present day abundance it is straightforward to estimate. Neutrinos interact via the weak force, which typically gives an interaction cross section $\sigma \sim G_F^2 T^2$ ($G_F = \alpha_W / M_W \simeq 10^{-5}\text{GeV}^{-2}$ is the Fermi energy, $\alpha_W \simeq 1/30$ is the weak fine structure constant and $M_W \simeq 100$ GeV is the weak scale). This gives $\Gamma \simeq n\sigma v \sim G_F^2 (kT)^5$ (since relativistic particles have $n \propto T^3$), so $\Gamma/H \simeq G_F^2 (kT)^3 m_{\rm Pl}^{-1} \simeq (T/\text{MeV})^3$ ($m_{\rm Pl}$ is the Planck mass). This gives $\Gamma/H = 1$ at $kT_F \sim$ MeV — neutrinos freeze out near $kT \sim 1$ MeV.[4]

Cold Relics: Weakly Interacting Dark Matter

Another example is given by so-called Weakly Interacting Massive Particles, or WIMPs, which might be the dark matter. These are massive, stable particles that interact very weakly with normal matter, and are in fact expected to exist in extensions of the standard model such as supersymmetry. Their cross sections are likely to be similar to those of neutrinos, but their masses would be much higher. Hence, they would behave like *cold relics*.

Supersymmetric theories usually feature a stable, massive particle, the Lightest Supersymmetric Partner, which is often electrically neutral, a *neutralino* (typically a linear superposition of the Higgsino and the "bino"). The calculation of the relic abundance for a cold relic is typically more complicated than for a hot relic, since freeze out occurs in the regime that the equilibrium abundance is falling exponentially where the details of the interaction are important. For a specific SUSY model, we could solve the Boltzmann equation numerically to predict the relic abundance. Alternatively, a simple estimate turns out to be quite accurate.

Assuming that freeze out occurs when $\Gamma = n\langle\sigma v\rangle n \sim H(T_F)$ we can write the freeze out abundance as $n_F \sim H(T_F)/\langle\sigma|v|\rangle \sim g_*^{1/2} T_F^2/\langle\sigma|v|\rangle m_{\rm pl}$, where we use $H \sim g_*^{1/2} T^2/m_{\rm pl}$ during radiation domination. To evaluate this, we need the freeze out temperature T_F, which really requires solving the Boltzman equation. However, such calculations typically find that the freeze out temperature is $T_F \sim M_X/10$ where M_X is the WIMP mass, and is fairly insensitive to the details of the interaction. This key piece of information is enough to estimate the relic abundance.

As for the neutrinos, we can map the number density of WIMPs at freeze out to the present day $n_X(a_0) = n_X(T_F)(T_0/T_F)^3$, but this time a dependence on T_F remains. For typical weak scale masses $M_X \sim M_W \sim 100$GeV the WIMP will freeze out when $g_* \sim 100$ and the annihilation of various particle species after WIMP decoupling will heat the photons relative to the WIMP so that $T_X \sim (1/100)^{1/3} T_0$. Taking all this into account, we find that the energy density today $\rho_X = m_X n_X$ can be written as

$$\Omega_X h^2 \sim \left(\frac{m_X}{10 T_F}\right) \left(\frac{g_*}{100}\right)^{1/2} \frac{10^{-27}\text{cm}^3\,\text{s}^{-1}}{\langle\sigma v\rangle}. \tag{5.27}$$

[4]Some of the equations in this paragraph are given in natural units with $\hbar = c = 1$, to be discussed in more detail in future chapters.

This has the expected dependence that the energy density in WIMPs decreases for larger interaction cross-section.

This result is extraordinary, although it might not seem it at first glance. We have assumed that are WIMPs interact via the weak force and so will have a cross-section $\sigma \sim G_F^2 T^2$, evaluating this for T_F and assuming that $v \sim c$ gives $\langle \sigma_W v \rangle \sim 10^{-26} \mathrm{cm}^3\,\mathrm{s}^{-1}$. Which would predict $\Omega_X h^2 \sim 0.1$. This is almost exactly the observed abundance of dark matter today! Remember that supersymmetry was invented to solve particle physics problems, not cosmological ones, so this could be just a coincidence. Nonetheless, it is intriguing that without fine tuning a WIMP could produce the required amount of dark matter, so much so that this result is often described as the "WIMP miracle".

We have not gone into too much detail about baryogenesis, neutrinos, and WIMPs, but in the following lectures, we will discuss two important transitions in the early Universe using these ideas. First, the creation of neutral hydrogen from the ionized plasma of electrons and protons at about 400,000 years after the big bang, and then the synthesis of the light elements from free neutrons and protons at about three minutes.

Chapter 6

Hydrogen Recombination and the Cosmic Microwave Background

MRR 5.2

6.1 Introduction and Executive Summary

In this chapter we will discuss how the Universe went from being an ionized plasma to a neutral gas of (mostly) hydrogen. The basic story is simple: the early Universe was hot enough, $kT \gg 1$ Ry $= 13.6$ eV, to keep the Universe ionized, and therefore for the thermal bath of photons to remain tightly coupled to the ions because of Thomson scattering. When the Universe cooled sufficiently, electrons and photons could (re-)combine to form neutral hydrogen, which also freed the photons from their interactions with matter. We see these photons as the Cosmic Microwave Background, essentially as the surface of an opaque "cloud" at redshift $z \sim 1100$.

As this story involves all well-understood processes in atomic physics, we can actually calculate the ionization history of the Universe in some detail. The most important interactions will be hydrogen recombination,

$$p^+ + e^- \to H + \gamma \tag{6.1}$$

and Thomson scattering

$$e^- + \gamma \to e^- + \gamma \tag{6.2}$$

where the appropriate cross section is

$$\sigma_T = \frac{8\pi}{3}\frac{\alpha\hbar}{m_e c} \simeq 6.7 \times 10^{-25}\ \mathrm{cm}^2 \tag{6.3}$$

and the binding energy is

$$(m_e + m_p - m_H)c^2 = 13.6\ \mathrm{eV} = 1\ \mathrm{Ry} = B. \tag{6.4}$$

The interaction rate is just $\Gamma = n_e c\sigma_T \propto 1/a^3 \propto (1+z)^3 \propto T^3$. Compare this to the expansion rate, which is either $H \propto T^2$ (RD) or $H \propto T^{3/2}$ (MD), so that we know that

at some time in the past, we must have gone from $\Gamma > H$ to $H > \Gamma$ more recently, so this reaction would have *frozen out*.

Another consequence of these interactions freezing out is that the photons, once tightly coupled to the ionized plasma (which is opaque due to Thomson scattering, $e^-\gamma \to e^-\gamma$), are able to stream freely through the now-neutral, and hence transparent, hydrogen gas. Thus we see these photons as if freed from a cloud; this is the **last scattering surface**, and it forms the cosmic microwave background, which looks to us (using microwave telescopes) as the surface of a cloud: further away (higher redhsifts) is opaque, nearer is transparent.

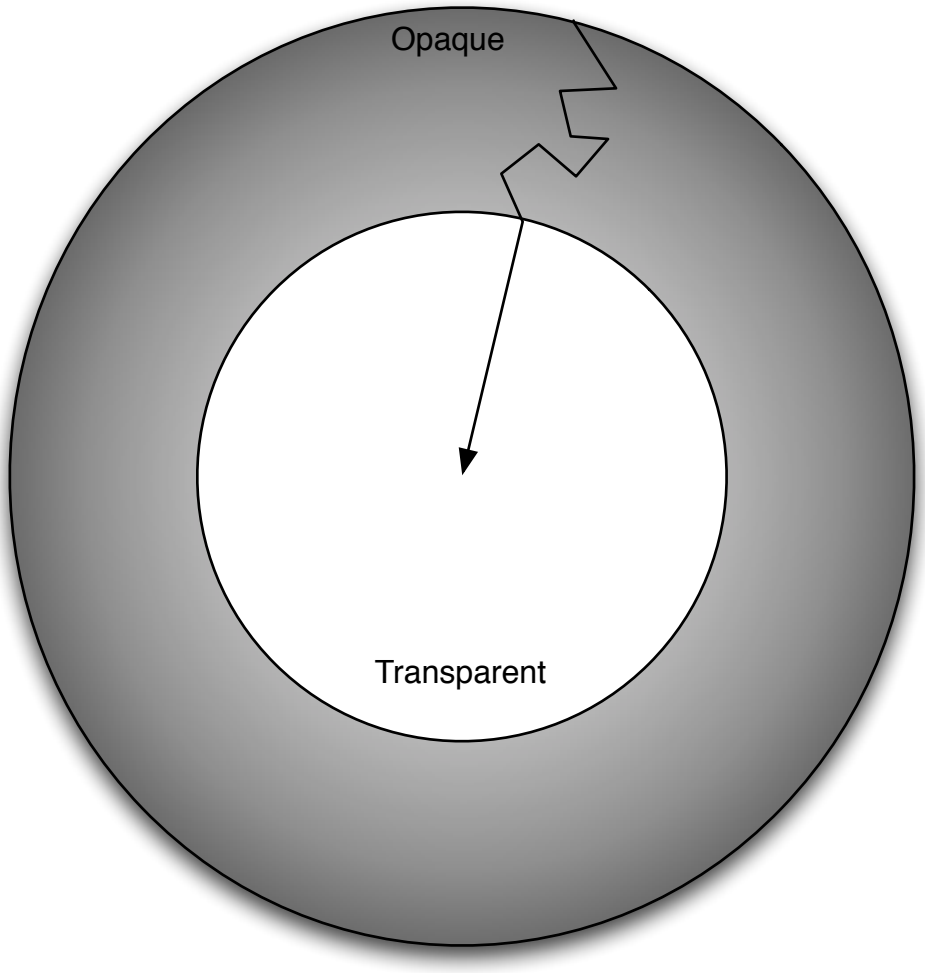

Figure 6.1: The surface of last scattering. Redshift increases outward. At early times, the photons are tightly coupled to the charged plasma. At later times they stream freely through the neutral gas.

6.2 Equilibrium ionization

First. we calculate the ionization state of the hydrogen under the assumption of equilibrium. In addition to the above numbers describing the interaction, we will need to explicitly assume charge neutrality ($n_p = n_e$) and baryon number conservation ($n_B = n_p + n_H$)[1]

In thermal equilibrium, we will assume a Maxwell-Boltzmann (non-relativistic) distribution for protons, electrons and Hydrogen atoms,

$$n_i = g_i \left(\frac{m_i k T_i}{2\pi\hbar^2}\right)^{3/2} \exp\left[\frac{(\mu_i - m_i c^2)}{kT_i}\right] \tag{6.5}$$

and a massless Bose-Einstein distribution for the photons. The total number density of photons is just the integral of the distribution as in Eq. 5.4:

$$n_\gamma = \left(\frac{kT}{hc}\right)^3 16\pi\zeta(3) \ , \tag{6.6}$$

which also gives the overall baryon density, $n_B = \eta n_\gamma$, where $\eta = 2.7 \times 10^{-8}\Omega_b h^2$ is the baryon-to-photon ratio. Statistical chemical equilibrium of Eq. 6.1 requires $\mu_p + \mu_{e^-} = \mu_H + \mu_\gamma = \mu_H$, and the degrees-of-freedom factors are $g_\gamma = g_p = g_e = 2$, $g_H = 4$.

Combining the expressions for the equilibrium abundances of the protons, electrons, and hydrogen atoms gives us the formula

$$n_H = \frac{g_H}{g_e g_p} n_e n_p \left(\frac{m_e m_p k T}{2\pi m_H \hbar^2}\right)^{-3/2} e^{B/(kT)} \tag{6.7}$$

If we define the *ionization fraction*,

$$X_e = \frac{n_p}{n_B} = \frac{n_p}{n_p + n_H} \ , \tag{6.8}$$

we get the **Saha equation**

$$\frac{1 - X_e}{X_e^2} = \frac{4\sqrt{2}\zeta(3)}{\sqrt{\pi}} \eta \left(\frac{kT}{m_e c^2}\right)^{3/2} e^{B/(kT)} \tag{6.9}$$

which looks complicated but is really just a quadratic equation for the equilibrium ionization fraction as a function of temperature T, or equivalently of redshift by $T = (1+z)T_0$. *Exercise: derive Eqs. 6.7-6.8 from the equations given in the previous paragraph.* We show $X_e(z)$ in Figure 6.2.

In the figure, we see that ionization fraction goes from one (fully ionized) to zero (neutral) around redshift 1300-1400. This corresponds to a temperature $kT \sim 0.3$eV Compare this to the redshift at which $kT = 13.6$ eV, calculated in one of the problems: $z(13.6\text{eV}) \sim 60,000$. Why the large discrepancy? It is the very small value of η: there are

[1] We know from our discussion in Section 5.4.1 that Baryon number is not always conserved, but we assume that this happens much earlier than this epoch.

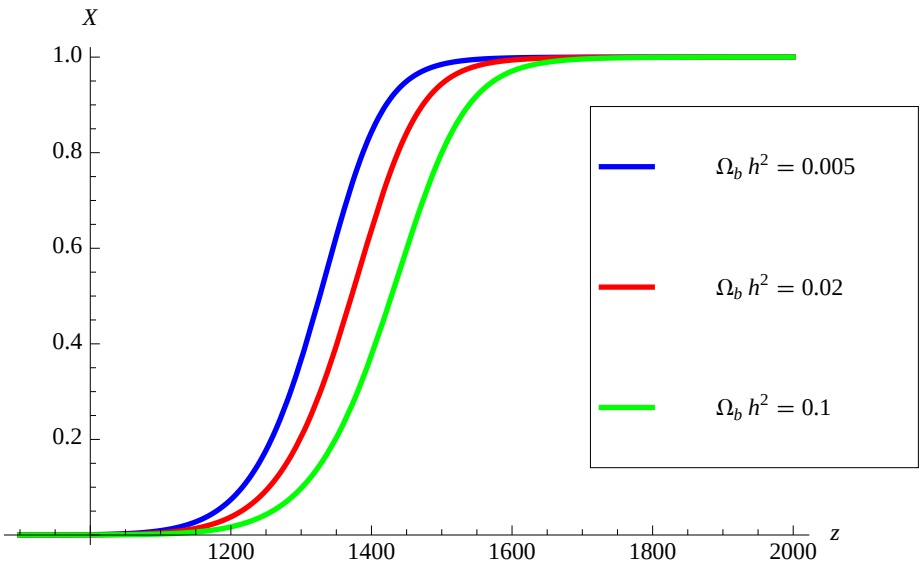

Figure 6.2: The equilibrium ionization fraction, X as a function of redshift, for different values of the Baryon density, $\Omega_b h^2$.

so many photons for every baryon that even the very small number in the exponentially-suppressed tail of the BE distribution of photons is enough to keep the hydrogen ionized until much later. Indeed, we could have derived this just by finding the temperature at which there is approximately a single ionizing photon with $h\nu > 13.6$ eV in the photon's Bose-Einstein distribution. *Exercise: show this.*

6.3 Freeze-Out

In the previous section we calculated the *equilibrium* ionization fraction. However, because the interaction rate (i.e., the recombination rate) depends upon the density of free electrons and protons, as more and more neutral hydrogen is formed, the rate of photon scattering off of electrons goes down and down, and will eventually freeze out.

Consider Thomson scattering,

$$e^- + \gamma \to e^- + \gamma \,, \tag{6.10}$$

which has a rate

$$\Gamma \simeq n_e \sigma_T c \,. \tag{6.11}$$

The number density of electrons depends upon the ionization fraction X and the baryon density:

$$n_e = X n_B = X \eta n_\gamma = X \eta \frac{3\zeta(3)}{2\pi^2} \left(\frac{kT_0}{\hbar c}\right)^3 (1+z)^3 \tag{6.12}$$

This is to be compared to $H = H_0 E(z)$. In Figure 6.3 I show (separately), the expansion timescale, $1/H$, and the mean time between interactions, $1/\Gamma$. When they cross, $\Gamma \sim H$

and for lower temperatures the interaction is frozen out. Note that this freeze-out occurs again at $z_F \sim 1100$ when the equilibrium ionization fraction is $X \ll 1$ – despite the very large number of photons per baryon, it takes a very small number of electrons to make the mean free path much smaller than the Hubble scale, and hence to make the Universe opaque. However, once this happens, the recombination interaction $ep \to H\gamma$ *also* freezes out, leaving a final ionization fraction $X = X_{eq}(z_F) \sim 10^{-4}$, which, although small, is much larger than the exponentially-reduced equilibrium ionization fraction that would have otherwise been predicted from Section 6.2.

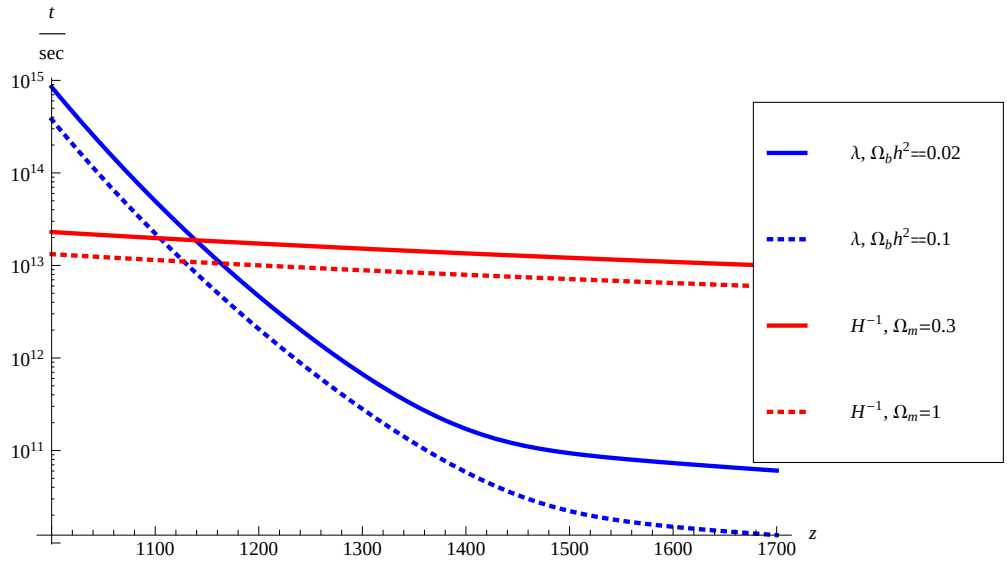

Figure 6.3: The expansion timescale ($H^{-1} = a/\dot{a}$) and the mean time between interactions, ($\lambda = 1/\Gamma$) for different values of $\Omega_b h^2$ and Ω_m. For redshifts less than the crossing point of the curves, the photons and baryons/electrons are no longer interacting.

In fact if we calculate this more carefully using the Boltzmann equation formalism of the previous lectures, we get the same result (as would be expected): freeze-out of Thomson scattering at $kT \sim 0.25$ eV.

It is important to realize that there are three somewhat distinct events:

- **Recombination**: The equilibrium ionization fraction, X goes from nearly one to nearly zero.
- **Last Scattering**: The freeze-out of $e^-\gamma$ Thomson scattering; this is also known as the **decoupling** of photons from the baryons.
- The freeze-in of residual ionization at a higher value than equilibrium, i.e., the freeze-out of the recombination reaction, Eq. 6.1. (This calculation requires more of the full Boltzmann equation apparatus to do accurately and we will not do it in detail here.)

Chapter 7

Big-Bang Nucleosynthesis

MRR 5.3

This material is covered in several textbooks, including Rowan-Robinson and Liddle, but there is an especially classic discussion in S. Weinberg, "The First Three Minutes".

In this chapter, we will discuss the synthesis of the light nuclei from the initial primordial mixture of protons and neutrons (along with the ubiquitous photons) — **Big Bang Nucleosynthesis** (BBN). In broad strokes, this is similar to our discussion of Hydrogen recombination and the formation of the CMB, but with quite a few extra complications. At early times, protons, neutrons, and light nuclei such as deuterium, tritium and Helium, are in equilibrium (specifically called "nuclear statistical equilibrium" in this case, just as we discussed "ionization equilibrium" in the context of the CMB). As the Universe expands and cools, different nuclear reactions freeze out, leaving us with relic abundances of the stable nuclei.

BBN has a few extra complications, however. First, we have to track the abundance of not just one end product (neutral hydrogen) but of several different nuclei. Second, one of the starting constituents, neutrons, are unstable when not in a nucleus, with a half-life of about 11 minutes. Finally, several of the possible end-states (light nuclei) have binding energies that are small or comparable to kT and so freeze-out can be delayed.

With our calculation of hydrogen recombination, we found in Eq. 6.7 that the equilibrium Hydrogen density is proportional to $\exp(B/kT)$ where $B = 13.6$eV is the binding energy of the hydrogen atom. Similarly, the most strongly-bound light nucleus is ^{4}He, with binding energy $B_4 = 28.5$ MeV, so we essentially expect most of the nucleons to end up in Helium in equilibrium, with other species suppressed exponentially in comparison. Moreover, the lack of any stable nuclei at all of mass 5 or 8 makes it very difficult to get to higher-mass nuclei beyond Helium. However, we will see that freeze-out of the interactions will mean that not all the species will remain in equilibrium so a more careful calculation is required.

Indeed, a full calculation of BBN requires the solution of many coupled differential equations, but the main results can be calculated from simple principles such as those we have applied to the CMB.

7.1 Initial Conditions for BBN: neutron-proton freeze-out

We start at temperatures $kT \gg 1$ MeV, times $t \ll 1$ s (compare the neutron-proton mass difference, $\Delta m = m_n - m_p = 1.29\ \mathrm{MeV}/c^2$). At this time, weak interactions maintain "nuclear statistical equlibrium" (NSE). The crucial (weak) interactions are

$$n \leftrightarrow p + e^- + \bar{\nu}_e \qquad \nu_e + n \leftrightarrow p + e^- \qquad e^+ + n \leftrightarrow p + \bar{\nu}_e \ . \tag{7.1}$$

We are late enough that we can assume non-relativistic (Maxwell-Boltzmann) statistics, so in equilibrium

$$n_i = g_i \left(\frac{m_i kT}{2\pi\hbar^2}\right)^{3/2} e^{(\mu_i - m_i)c^2/kT} \tag{7.2}$$

so

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{3/2} \exp\left[\frac{(\mu_n - \mu_p) - (m_n - m_p)c^2}{kT}\right] \ . \tag{7.3}$$

We can relate the chemical potentials to our interactions as before, e.g., $\mu_\nu + \mu_n = \mu_p + \mu_{e^-}$, but in fact we have $\mu \ll \Delta m$ so we can ignore that term in the exponential. If we further assume $m_n/m_p \simeq 1$ in the prefactor,

$$\frac{n_n}{n_p} = \exp\left[-\frac{\Delta m c^2}{kT}\right] \ . \tag{7.4}$$

But we had better check that we are in equilibrium. The rate for these interactions is related to the weak interaction scale, $\Gamma = n\sigma v \sim G_F^2 T^2 T^3$, which can be calculated in more detail giveing, e.g.,

$$\Gamma(\nu_e n \leftrightarrow p e^-) \simeq 2.1 \left(\frac{kT}{\mathrm{MeV}}\right)^5 \mathrm{s}^{-1} \tag{7.5}$$

In an RD universe, this must be compared to

$$H = \frac{\dot{a}}{a} = \frac{1}{2t} \simeq \left(\frac{kT}{\mathrm{MeV}}\right)^2 \mathrm{s}^{-1} \tag{7.6}$$

Putting these together, near $kT \sim 1$MeV,

$$\frac{\Gamma}{H} \simeq \left(\frac{kT}{0.8\ \mathrm{MeV}}\right)^3 \tag{7.7}$$

Hence, we freeze-out at $T_F \simeq 0.8$ MeV. At this point, we have a neutron-proton ratio

$$\frac{n_n}{n_p} = \exp\left[-\frac{\Delta m c^2}{kT_F}\right] \simeq \exp[-1.29/0.8] \simeq 0.2 \tag{7.8}$$

(a more careful calculation gives 0.17-0.18, so this is good to about 10-15%).

It is worth pointing out that the freeze out of these interactions also means that the neutrinos will have frozen out more generally — they decouple from their interactions with matter and photons, and stream freeely thereafter, just as the photons themselves later decouple from the baryons and leptons, as with the formation of the CMB discussed in Chapter 6.

7.2 Helium production

After the freeze-out of the reactions interconverting neutrons and protons, the number density of each is fixed (except for neutron decay which we will account for later). Now, we consider various chains of reactions, both starting with

$$p + n \to d + \gamma \; ; \tag{7.9}$$

(d is deuterium, the hydrogen isotope made of a neutron and a proton, which has a binding energy $B_2 \simeq 2$ MeV). From here, there are two possible chains:

$$d + d \to {}^3\mathrm{He} + n \qquad {}^3\mathrm{He} + d \to {}^4\mathrm{He} + p \tag{7.10}$$

and

$$d + d \to t + p \qquad t + d \to {}^4\mathrm{He} + n \tag{7.11}$$

(t is tritium, the hydrogen isotope made of two neutrons and a proton). Either way, we first have to make deuterium, but there are two factors that make this difficult. First, as in the case of Hydrogen recombination, the equilibrium deuterium fraction is controlled by an equation similar to the Saha Equation, Eq. 6.8, with factors of $(kT/m_p c^2)^{3/2}\eta \ll 1$. Just as in the case of recombination, the large baryon-to-photon ratio keeps the interaction $pn \to d\gamma$ from proceeding forward (i.e., the photons dissociate the deuterium) well below the temperature $kT = B_2 \simeq 2$ MeV binding energy. Instead, the deuterium fraction remains small until $kT \simeq 0.1\ MeV$. Even at this point, the interactions cannot proceed fully, due to the interaction rates:

$$\frac{\Gamma(pn \to d\gamma)}{H} \simeq \left(\frac{kT}{0.05\ \mathrm{MeV}}\right)^5 \frac{n_p}{n_p + n_n} \left(\frac{\Omega_B h^2}{0.02}\right) \tag{7.12}$$

This has $\Gamma/H > 1$ for $kT > 0.05$ MeV $\simeq kT_F = kT_{\mathrm{BBN}}$, equivalent to a time $t_{\mathrm{BBN}} = (kT_{\mathrm{BBN}}/\mathrm{MeV})^{-2}/2$ s $\simeq 200$ s, which we will take to be about 3 minutes. Therefore only a very small number of deuterium nuclei are formed from the baryons, and hence further there is a very low rate for either of the dd interactions (since both have rates $\Gamma(dd) \propto n_d^2$). We essentially never make it to equilibrium abundances of helium. This is called the **deuterium bottleneck**, lasting until $kT \sim 0.05$ MeV.

Once we reach 0.05 MeV, we can form deuterium and both of these pathways can proceed; there is rapid burning of almost all of the remaining protons and neutrons into Helium. To calculate the final state, we first define the *mass fraction* of a helium nucleus:

$$Y = \frac{A_{He} n_{He}}{n_B} \tag{7.13}$$

where $A = 4$ is the atomic mass of Helium, and so the numerator is proportional to the mass density of Helium, and the denominator to the total mass density of baryons. We assume that essentially of the neutrons go into He nuclei, so the number density of Helium will be half of the number density of neutrons. Hence

$$Y = \frac{4n_{He}}{n_n + n_p} = \frac{4n_n/2}{n_n + n_p} = \frac{2n_n/n_p}{1 + n_n/n_p} \tag{7.14}$$

where now we have related the helium mass fraction to the neutron-proton ratio calculated above. At np freeze-out $T_F \simeq 1$ MeV we had $n_n/n_p \simeq 0.17 \simeq 1/6$. By the time the deuterium bottleneck is broken, however, it is now $t_{\rm BBN} \sim 3$ min, which means that a non-negligible fraction of the neutrons will have decayed into protons, $n \to p^+ e^- \bar{\nu}_e$, with a half-life of 10.5 minutes. Hence

$$n_n(T_{\rm BBN}) \simeq n_n(T_F) 2^{-3/10.5} \,. \tag{7.15}$$

This decreases the neutron-to-proton ratio to

$$\left.\frac{n_n}{n_p}\right|_{T_{\rm BBN}} \approx 0.13 \tag{7.16}$$

and thus

$$Y \simeq \frac{2 \times 0.13}{1 + 0.13} \simeq 0.24 \tag{7.17}$$

We can also define the mass fraction of other species $X_i = A_i n_i / n_B$ which can be found by considering the freeze-out of the reactions we have considered so far, as well as further reactions that can build higher-mass nuclei, giving final abundances slightly higher than the equilibrium abundances. Lithium is produced by the chain

$$^4\mathrm{He} + t \to {}^7\mathrm{Li} + \gamma \qquad {}^3\mathrm{He} + {}^4\mathrm{He} \to {}^7\mathrm{Be} + \gamma \qquad {}^7\mathrm{Be} + e^- \to {}^7\mathrm{Li} + \nu_e \,. \tag{7.18}$$

7.2.1 Dependence upon the cosmological parameters

The details of the production of helium and the other light elements depend upon the state of the Universe over the epochs in time and temperature that we have been discussing. This is most obvious for the dependence upon the baryon density, $\Omega_B h^2$, which comes in directly in the equilibrium distribution of deuterium (as well as in the rate of the deuterium fusion reaction $\Gamma[pn \to d\gamma]$, Eq. 7.12). Just as a higher baryon density results in a higher density of hydrogen atoms (via $pe \to H\gamma$), a higher baryon density also implies a higher density of deuterium (via $pn \to d\gamma$). This, in turn, implies a higher density of Helium, with an even stronger dependence upon the baryon density. We show the dependence upon the baryon density for several light elements in Figure 7.1

It is more difficult to synthesize elements beyond Helium. First, there are no stable elements with $A = 5$ or $A = 8$, so we cannot do so by adding just a single nucleon to Helium. Second, the way to create any such elements would therefore be to fuse several light nuclei together, but each of those would have positive charge, and hence the reaction rates are strongly suppressed by the **Coulomb Barrier**. By the time deuterium is produced in sufficient abundances to form Helium, the temperature has decreased low enough that few nuclei have enough energy to breach the barrier.

Nonetheless, some ^{7}Li is formed via the reactions of Eq. 7.18. For baryon density $\eta \lesssim 3 \times 10^{-10}$, this occurs via the tritium path:

$$^3t + {}^4\mathrm{He} \to {}^7\mathrm{Li} + \gamma \,; \tag{7.19}$$

however, lithium is very easy to destroy via collisions with protons, so increasing η actually means decreasing final lithium abundance. Whereas for higher values of η, there is more ^{3}He around and we can first produce beryllium via the Helium-3 path,

$$^3\mathrm{He} + {}^4\mathrm{He} \rightarrow {}^7\mathrm{Be} + \gamma \ , \tag{7.20}$$

which then inverse β-decays via electron capture to ^{7}Li. Because beryllium is actually more strongly bound than lithium, once this pathway is opened the final lithium abundance increases with η.

Furthermore, as we increase η, more and more d and ^{3}He is burnt into helium the reactions Eqs. 7.10-7.11, which freeze out later and later, since the reaction rates are proportional to η.

Although a full calculation requires the simultaneous solution of coupled differential equations, it is usually straightforward to perform a perturbation analysis to see the effect of making a small change in the other cosmological parameters. To give a flavor of this analysis consider the effect of changing the time (or equivalently, temperature, T_{BBN}) that BBN occurs. If the expansion was faster, freeze-out, breaking the deuterium bottleneck, would have occurred earlier. At this earlier time, neutrons would have had less time to decay, and the neutron-proton ratio would have been higher, and hence the Helium mass fraction, Y would have been higher. But why would the expansion rate be higher? One way is to increase the density of radiation in the universe, since $H^2 \propto \rho$, so any (relativistic) species beyond those we know of in the standard model would increase the Helium abundance — indeed for quite a while observations of the Helium abundance (and that of other elements) were the strongest constraint on additional light particles, in particular neutrinos.

7.3 Observations of primordial abundances

See classroom presentation.

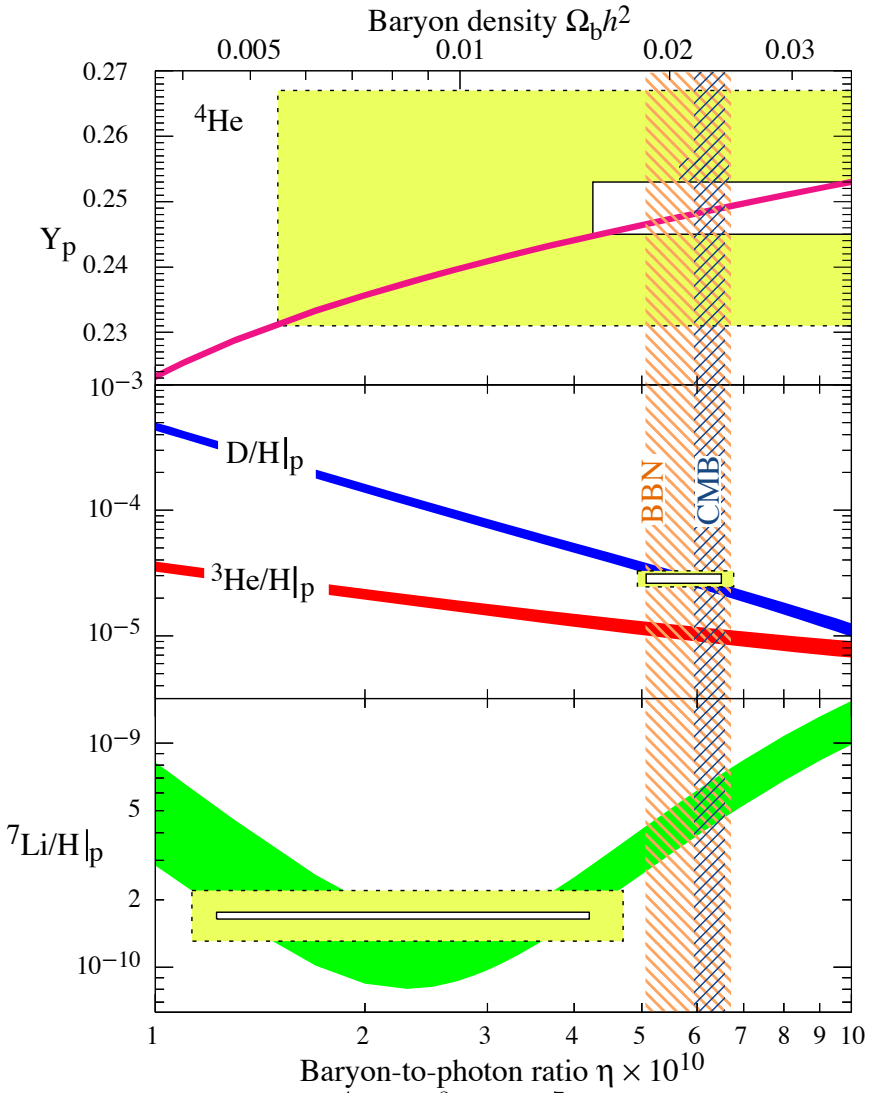

Figure 7.1: The mass fraction of various species as a function of the baryon density. Bands show the 95% confidence range. Boxes indicate observed light element abundances (smaller boxes: $\pm 2\sigma$ statistical errors; larger boxes: $\pm 2\sigma$ statistical and systematic errors). The narrow vertical band indicates the CMB measure of the cosmic baryon density, while the wider band indicates the concordance range of direct measurements of the light element abundances. From Fields and Sarkar, in Amsler et al., PL B667, 1 (2008, 2009) `http://pdg.lbl.gov`

Chapter 8

Interlude

8.1 Natural Units

The major theories of modern physics developed since the time of Newton come armed with a set of physical constants: formulas from relativity (and electromagnetism) always use c, the speed of light. In thermodynamics, temperature always appears in the energy combination kT where k is Boltzmann's constant. Quantum mechanics, of course, uses Planck's constant, $\hbar$. Finally, gravitation (GR or Newtonian) uses Newton's constant, G. It further turns out that it is impossible to express the units of any of these in terms of any other. *Exercise: Show this.*

Therefore, it makes sense when doing calculations in relativity and electrodynamics to measure all speeds with respect to c (indeed we often do this by considering the quantity $\beta \equiv v/c$, but now we are just using the shorthand of writing v instead of β). Moreover, we can also decide to measure masses in energy units: instead of saying that the electron has mass $m_e = 0.511$ MeV$/c^2$, we can just use the symbol m_e for the quantity that we used to call $m_e c^2$.

With these conventions, many equations become simpler and their meaning more manifest. For example, the relationship between relativistic energy, mass and momentum is simply:

$$E^2 = m^2 + p^2 \ . \tag{8.1}$$

Without the speed-of-light factors, the equation is much more symmetric amongst the quantities. This is even more manifest in the version of this equation applicable to particles at rest, arguably Einstein's most famous equation: $E = m$. We now see this as the statement that energy and mass are just different words for exactly the same thing. Similarly, in a cosmological context we have had to worry about both energy density ε and matter density ρ but in our $c = 1$ units we just see $\varepsilon = \rho$.

Similarly, in quantum mechanics calculation, we similarly measure all quantities with units of angular momentum or action in terms of $\hbar = h/(2\pi)$, and we can decide to use the symbol ω for the quantity $\hbar\omega$ which has energy units. In this case, Heisenberg's uncertainty relation takes the simple form

$$\Delta x \ \Delta p \geq \frac{1}{2} \tag{8.2}$$

and Schrodinger's equation is

$$i\frac{\partial\psi}{\partial t} = H\psi \tag{8.3}$$

with the recipe $p \to i\partial/\partial x$ in the Hamiltonian.

Finally, we can use T for the quantity kT in thermodynamics calculations. But we can do this more generally: in shorthand, we say we are taking $\hbar = c = k = 1$ and are using *natural units*. In normal units

$$[c] = LT^{-1}\,, \qquad [\hbar] = ML^2T^{-1}\,, \text{and} \qquad [k] = ML^2T^{-2}K^{-1} \tag{8.4}$$

where M is mass, T is time, L is length and K is temperature (since T is already taken). In cosmology and particle physics, it is traditional to *keep* Newton's constant explicit in calculations (although in relativity and fields like string theory, $G = 1$, or occasionally $8\pi G = 1$, is often used).

In these examples, it is clear that quantities with units of energy have a special place. Indeed, we can't make an energy out of any combination of $\hbar$ and c (k is obvious since we know kT has energy units and there is no way to get a temperature out of $\hbar$ and c):

$$\begin{aligned} ML^2T^{-2} &= [c]^a[\hbar]^b \\ &= L^aT^{-a}M^bL^{2b}T^{-b} \\ &= M^bL^{a+2b}T^{-a-b} \end{aligned} \tag{8.5}$$

which is equivalent to

$$b = 1 \qquad a + 2b = 2 \qquad -2 = -a - b \tag{8.6}$$

which are three equations in two unknowns and do not have a solution (a similar argument holds for lengths or times). So, as long as we keep units on G, we still need to choose some units in which to report physical quantities, and it is traditional in cosmology and particle physics to use energy units, in particular electron-volts (or keV, MeV, GeV as appropriate). Note that *any* physical quantity can be expressed as an energy to some power this way. For example, we know that we can relate a length λ to a frequency c/λ and hence to an energy $\hbar c/\lambda$ — so length can be thought of as having units [energy^{-1}]. How do we do the conversion explicitly? We need to find a combination of a length, ℓ with $\hbar$ and c that gives the appropriate units:

$$\begin{aligned} [\text{Energy}] &= [\ell^\alpha\hbar^\beta c^\gamma] \\ ML^2T^{-2} &= L^\alpha M^\beta L^{2\beta}T^{-\beta}L^\gamma T^{-\gamma} \\ &= M^\beta L^{\alpha+2\beta+\gamma}T^{-\beta-\gamma} \end{aligned} \tag{8.7}$$

which is three equations in three unknowns, giving $\beta = \gamma = -\alpha = 1$ so the combination $\hbar c/\ell$ has energy units as expected. Since we can also convert from length to time using factors of c, we can also express times in energy units.

As a specific example, consider the Hubble constant,

$$H_0 = 100h\ \text{kms}^{-1}\text{Mpc}^{-1} = 3.24h \times 10^{-18}\text{s}^{-1} = 2.1h \times 10^{-33}\text{eV} \tag{8.8}$$

or the current CMB temperature

$$T_0 = kT_0 = 2.4 \times 10^{-4}\text{eV} \ . \tag{8.9}$$

Since we are keeping G, what units does it have? We could work it out by brute force, but an easier way is to first recall that the gravitational potential is given by the combination GM^2/L and from the previous discussion we know that a length L has units of $[\text{Energy}]^{-1} = [E]^{-1}$. So

$$[GM^2] = [G] \times [E^2] = 1 \tag{8.10}$$

or

$$[G] = [E]^{-2} = [M]^{-2} = [L]^2 = [T]^2 \tag{8.11}$$

The associated energy, mass, length, time etc. are known as the *Planck* energy, *Planck* mass, *Planck* length, *Planck* time, etc.:

$$E_{\rm Pl} = G^{-1/2} = M_{\rm Pl} \tag{8.12}$$

which in normal units is

$$E_{\rm Pl} = \sqrt{\frac{\hbar c^5}{G}} = 1.22 \times 10^{19}\ \text{GeV} \ . \tag{8.13}$$

Similarly the Planck length and time are

$$l_{\rm Pl} = t_{\rm Pl} = G^{1/2} \tag{8.14}$$

which in normal units is

$$t_{\rm Pl} = \sqrt{\frac{\hbar G}{c^5}} = 5.4 \times 10^{-44}\ \text{s} \tag{8.15}$$

(and we just need another factor of c to convert to the Planck length). If we do set $G = 1$ then all quantities can be expressed as pure numbers, measured with respect to these quantities: these are called *Planck units*.

In natural units, we can also do a lot of "back-of-the-envelope" physics. Consider, for example, the energy density of a blackbody. Except for the temperature, there are no physical constants with units that could appear in such a problem beyond the fundamental physical constants we have discussed. There are no particle masses or fundamental lengths that might matter. Moreover, we know that gravitation doesn't enter into the problem, so the Planck length/mass/time/energy should not be relevant. All we have to work with is the temperature. So how do we make an energy density? We need to get units of energy per volume. Volume is $[L]^3$ and we know that $[L] = [E]^{-1}$, so we know $[\varepsilon] = [E]^4$. Since all we have to work with is the temperature, this *must* be of the form

$$\varepsilon = \mathcal{O}(1) \times T^4 \tag{8.16}$$

where $\mathcal{O}(1)$ refers to a constant, expected to be of order one.[1] Indeed, the full expression, as we have seen, is

$$\varepsilon = \frac{\pi^2}{15(\hbar c)^3}(kT)^4 \tag{8.17}$$

[1] Of course, there are rare occasions when these considerations may go awry — sometimes the expected $\mathcal{O}(1)$ constant may for some reason be much larger or smaller than one.

so the constant is $\pi^2/15 \simeq 0.66$.

Now consider the Friedmann Equation, first in a flat, $\Lambda = 0$ universe:

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho \\ H^2 &\sim \frac{8\pi}{3}\frac{T^4}{m_{\rm Pl}^2} \end{aligned} \tag{8.18}$$

where in the second line we have assumed radiation domination, so $\rho = \varepsilon \sim T^4$ and used $G^{-1/2} = M_{\rm Pl}$. This gives the simple formula

$$H \sim T^2/m_{\rm Pl} \qquad \text{in RD.} \tag{8.19}$$

8.1.1 The Cosmological Constant problem

Now, let's consider a more general universe, the one that seems to obtain, with a cosmological constant. First, let's try to work out what units the cosmological constant should have. From the Λ term in the Friedmann equation

$$[H]^2 = [\Lambda/3] \tag{8.20}$$

or

$$[T]^{-2} = [\Lambda] \tag{8.21}$$

so

$$[\Lambda] = T^{-2}\,, \tag{8.22}$$

Which means that in our natural units,

$$[\Lambda] = [T]^{-2} = [L]^{-2} = [M]^2 = G^{-1}\,, \tag{8.23}$$

Now, we have measured the cosmological constant to be

$$\Omega_\Lambda = \frac{\Lambda}{3H_0^2} \simeq 0.7 \tag{8.24}$$

so

$$\Lambda \simeq 0.7 \times 3H_0^2 \simeq (9.4{\rm Gyr})^{-2} \simeq \left(10^{26}{\rm m}\right)^{-2}\,. \tag{8.25}$$

Another way to think about the value of the cosmological constant is to consider its energy density:

$$\begin{aligned} \rho_\Lambda c^2 &= \Omega_\Lambda \rho_{\rm crit} = \Omega_\Lambda \times \frac{3H_0^2}{8\pi G} \\ &\sim 6\times 10^{-10}\ {\rm J\ m^{-3}} \sim 10^{-47}\ {\rm GeV}^4 \simeq (10^{-12}\ {\rm GeV})^4 \end{aligned} \tag{8.26}$$

where in the last we use natural [energy] units. Do we have any fundamental theories that could *predict* the value of the cosmological constant? When thought of as an energy

density, we have seen that the cosmological constant has the unique property that it does not change with time, despite the expansion of the Universe — it is the energy of the *vacuum*. We also know that our best description of particle physics, quantum field theory, gives a generic prediction for the value of the vacuum energy density: it should be $\rho_{\rm vac} \sim E^4$ (in natural units), where E is some energy scale that describes the high-energy ("ultraviolet" in particle physics parlance) cutoff of our theory. That is, E is the scale at which our effective quantum field theory breaks down. Unfortunately, the observed value of $E_\Lambda \sim 10^{-12}$ GeV $= 0.001$ eV is much lower than any such expected cutoffs — it is well below the energy scale of everyday physics. Indeed, if Λ is from a true quantum-gravity theory, we would expect $E = E_{\rm Pl}$, in which case we would have

$$\Omega_\Lambda \sim \frac{1}{GH^2} \sim (t_{\rm Pl}^2 H^2)^{-1} \sim \left(\frac{10^{17}}{10^{-44}}\right)^2 \sim 10^{122} \,, \tag{8.27}$$

which is much, much, much greater than the observed value of 0.7. Even if the cosmological constant were due not to quantum gravity but to supersymmetry, we would still expect $\Omega_\Lambda \sim 10^{60}$, not much better.

There are various solutions, all somewhat unsatisfactory, proposed to solve the cosmological constant problem. The first is that the "real" cosmological constant is $\Lambda = 0$, but that there is some other physical mechanism — "dark energy" — that can provide an energy density with equation-of-state $w = p/\rho = -1$, not a true vacuum energy. In fact, a *scalar field* can provide this. Indeed, we will see that inflation is thought to be driven by such a scalar field, although the energy scales of inflation and dark energy are so different that no one has been able to come up with a single mechanism for producing both epochs of $w = -1$.

Another possibility is that the properties of the vacuum depend upon the details of the fundamental theory. In string theory, for example, there are a huge number (possibly 10^{200} or greater!) ways to compactify down to 3+1 dimensions, and the cosmological constant could be different in each of them. Then, we may need to use something like the anthropic argument to find the cosmological constant. The basic idea is due to Weinberg ("Anthropic Bound on the Cosmological Constant", Phys. Rev. Lett. 59 (22): 2607–2610. doi:10.1103/PhysRevLett.59.2607): generically, the theory predicts a distribution of Λ, but most of the distribution is with considerably greater values than observed. However, with much greater values, the Universe would look very different than it does today. In particular, it seems that it would be very difficult to form any structures like galaxies (and likely the stars within them) at all, as the Universe would start exhibiting accelerated expansion before the structures could form. (We shall see this in more detail when we discuss large-scale structure). Hence, we wouldn't be here to observe the cosmological constant if it were much larger. Thus, the prediction is that we should observe the largest possible value of Λ consistent with structure formation. This seems to be, roughly, true. (It should be pointed out that many cosmologists are deeply troubled by anthropic arguments!)

For the rest of these notes, we will usually use $\hbar = c = k = 1$ natural units.

8.2 Open Questions in the Big Bang Model

The Big Bang model — i.e., the FRW metric — is amazingly successful: it predicts an expanding Universe, a background of relic photons (the CMB), and detailed relations amongst the light element abundances. But the initial conditions are left unspecified: why do the densities of the various components have their observed values? Why is the Universe approximately homogeneous and isotropic?

We can make these questions more precise, and see that the initial conditions do not appear to be very generic at all. Rather, the Universe appears to be very finely tuned.

8.2.1 The Flatness Problem

Consider our definition of the contribution of curvature to the energy density as a function of redshift:

$$\Omega_k(z) \equiv 1 - \Omega_{\rm tot}(z) = \frac{-k}{a^2H^2} = \Omega_k(z=0)\frac{(1+z)^2}{E^2(z)} \tag{8.28}$$

with

$$E^2(z) = \Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_\Lambda + \Omega_k(1+z)^2 \tag{8.29}$$

(since there is no actually density associated with curvature, it is often better to just think of this as related to the total density as in the first equality). If the universe is truly flat, then $\Omega_k(z) = 0$ for all time. But otherwise, in the early matter-dominated era,

$$\Omega_k(z) = \frac{(1+z)^2}{\Omega_m(1+z)^3} = \frac{1}{\Omega_m(1+z)} \tag{8.30}$$

whereas in the radiation-dominated era

$$\Omega_k(z) = \frac{(1+z)^2}{\Omega_r(1+z)^4} = \frac{1}{\Omega_r(1+z)^2}\ . \tag{8.31}$$

In both cases, $|\Omega_k(z)|$ is an increasing function of time (i.e., a decreasing function of redshift). No matter how close the Universe is to flat today, it was even closer in the past. We know that $|\Omega_k| \lesssim 0.1$ today. What does this imply for the curvature at some early epochs?

- At hydrogen recombination, $z \sim 1000$, we have $\Omega_k \lesssim 10^{-4}$;
- At matter-radiation equality ($z \sim 10^4$), we have $\Omega_k \lesssim 10^{-5}$;
- at nucleosynthesis, ($z \sim 10^8$), we have $\Omega_k \lesssim 10^{-13}$;

and at earlier and earlier epochs (electroweak symmetry breaking, the Planck epoch), the requirement gets stronger and stronger. So if the Universe is *not* flat today, it had to start out remarkably close to — but not quite — flat. This is not a very generic condition at all.

8.2.2 The Horizon Problem

When we talked about our picture of the Universe today, we discussed our observations of the Cosmic Microwave Background. At a level of even better than one part in 1000, the CMB is smooth (although we will see that the tiny fluctuations at level of 10^{-5} are another crucial clue to the origin and evolution of the Universe). What does this mean?

Recall the angular diameter distance

$$d_A(z) = cH_0^{-1}(1+z)^{-1}|\Omega_k|^{-1/2}\sin_k\left(\sqrt{|\Omega_k|}\int_0^{z'}\frac{dz'}{E(z')}\right) \quad . \tag{8.32}$$

which is *defined* by the relation between angular size, θ and physical size L: $\theta = L/d_A$. We expect that the largest distance that physics ought to be able to act is the horizon size,

$$d_H(z) = \frac{cH_0^{-1}}{1+z}\int_z^\infty \frac{dz'}{E(z')} \, . \tag{8.33}$$

so we can combine these to find the angular size of the horizon, $\theta_H = d_H/d_A$. If we assume a flat Universe, we have

$$\theta_H = \frac{d_H}{d_A} = \frac{\int_z^\infty dz'/E(z')}{\int_0^z dz'/E(z')} \, . \tag{8.34}$$

If we further assume a matter-dominated Universe, the integrals are easy, since $E(z) = \Omega_m^{1/2}(1+z)^{3/2}$:

$$\theta_H = \frac{(1+z)^{-1/2}}{1-(1+z)^{-1/2}} \sim (1+z)^{-1/2} \ll 1 \qquad \text{for } z \gg 1. \tag{8.35}$$

In particular, $\theta_H \simeq 1.7^\circ$ at $z \simeq 1000$, and this is about right even if we take the details of the density of matter, radiation and cosmological constant into account.

This means that any two patches of the CMB sky more than a couple of degrees apart should not have been in causal contact — so how did they get to be the same temperature to at least one-tenth of one percent? In fact, we don't need to use the CMB to see the problem: even at redshifts of a few, different regions of the sky are in different horizon volumes, and yet we observe the matter density (as seen in the galaxy density) to be roughly similar.

Another way to see this problem is to consider the scale of a structure in the Universe, say, that of a galaxy. The *physical* scale of a galaxy, $\lambda_{\rm gal}$, grows along with the scale factor, $\lambda_{\rm gal} \propto a \propto t^{2/3}$ (MD) or $\lambda_{\rm gal} \propto t^{1/2}$ (RD). But (in a Universe dominated by matter, radiation, or curvature) the horizon scale grows as $d_H \sim t \sim H^{-1}$. Today, $\lambda_{\rm gal} < d_H$ and we say that the scale is "inside the horizon" but because they grow at different rates, at some point in the past we must have had $\lambda_{\rm gal} > d_H$, "outside the horizon"; the time at which the $d_H = \lambda_{\rm gal}$ is called **horizon-crossing** for that scale (see Figure 8.1). We expect that structures can only grow when they are inside the horizon (due to causality) but we must further be able to set up initial conditions so that this can happen. And

for astrophysically-interesting scales, this happens at a time for which we thing we have a good understanding of the physics (e.g., the galaxy-size scale of $\lambda \sim 10$ Mpc enters at approximately the epoch of matter-radiation equality).

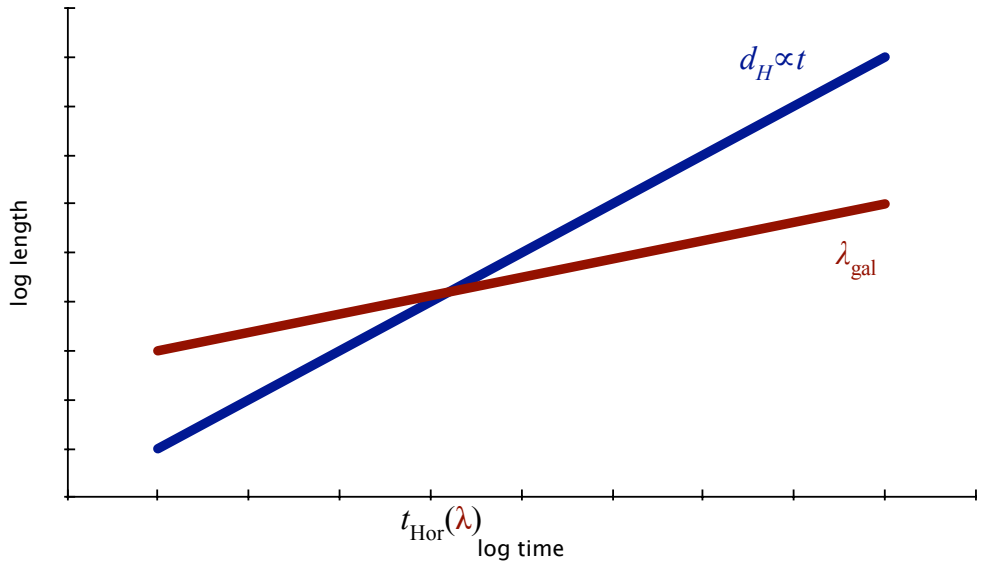

Figure 8.1: *Physical* scales inside and outside the horizon. Prior to (i.e., to the left of) horizon-crossing at $t_{\rm Hor}(\lambda)$, the scale is outside the horizon; afterwards it is inside the horizon and causal physics can act.

Note that we could have equally made this argument in terms of *comoving* scales, $l_{\rm gal} = \lambda_{\rm gal}/a = {\rm const}$, which now must be contrasted with the comoving horizon distance $\chi_H = d_H/a$. In a universe with matter, radiation and curvature, $\chi_H \sim (aH)^{-1}$ and the comoving Hubble length always grows with time: the horizon grows to encompass larger and larger scales, so any scale now inside the horizon was once outside.

8.2.3 The relic particle problem

The final puzzle has to do with leftover relics from the early Universe. We have already seen that the CMB itself is such a relic, as are the light elements such as helium. These come from epochs of transition from one state to another as various interactions freeze out. Are there any other relics we might expect?

It has been hypothesized that there is a grand-unified theory (GUT) that combines the strong and electroweak forces — in particle physics parlance, this means finding the appropriate single group in which to embed the standard model SU(3) × SU(2) × U(1). If so, this theory will almost inevitably have a very massive electromagnetic **monopole**. Because of the way it interacts, we would expect it to be formed with a number density of about one monopole per horizon volume at some high temperature $T_{\rm GUT}$. (The mechanism is akin to the formation of defects in phase transitions in solids, extended to particle physics by Higgs, and cosmology by Tom Kibble of Imperial. A similar mechanism may

also produce cosmic strings[2] and similar cosmological topological defects.)

There would therefore be approximately one monopole per Hubble volume, giving a *physical* number density of monopoles at this early time of $n(T_{\rm GUT}) \sim H_{\rm GUT}^3$, with $H_{\rm GUT} \sim T_{\rm GUT}^2/m_{\rm Pl}$ appropriate for an RD universe. This density would have been diluted by a factor of $(1+z_{\rm GUT})^{-3} = T_0^3/T_{\rm GUT}^3$ between the GUT time and now, so the present-day mass density of monopoles of mass $M_{\rm mon}$ would be

$$\rho_{\rm mon}(t_0) \sim M_{\rm mon}\frac{T_0^3 T_{\rm GUT}^3}{m_{\rm Pl}^3} \ , \tag{8.36}$$

We can work out $\Omega_{\rm mon}$ by dividing by $\rho_{\rm crit} = 3H_0^2/(8\pi G) \sim H_0^2 m_{\rm Pl}^2$:

$$\Omega_{\rm mon} \sim \frac{T_{\rm GUT}^3 T_0^3 m_{\rm mon}}{m_{\rm Pl}^5 H_0^2} \sim 10^{11}\frac{m_{\rm mon}}{10^{16}\ {\rm GeV}}\left(\frac{T_{\rm GUT}}{10^{14}\ {\rm GeV}}\right)^3 \tag{8.37}$$

where we have used $H_0 \sim 10^{-42}$ GeV, $T_0 \simeq 2\times10^{-13}$ GeV, $m_{\rm Pl} \simeq 1.2\times10^{19}$ GeV and put in typical values for the GUT scale and the monopole mass.

Since we know that $\Omega \sim 1$ we know that this is not possible. (It is often said that these relics would "overclose" the Universe, but a better interpretation would be to say that such a density at this early time would have caused the Universe to be closed and collapse again on a very short timescales, rather than continue expanding for the 14 billion years — and counting! — that we observe.) Moreover, our observations of, say, the light-element abundances also imply that the Universe must have been radiation-dominated at least prior to about three minutes after the big bang.

This problem was thought to be such an important issue that it was the primary motivation for Guth's original model of inflation — it is often referred to more specifically as the **monopole problem**. In fact, we do not know if a GUT transition actually happened in the early Universe, but there are various other transitions that may have happened, resulting in a high density of very massive particles.

[2]Not to be confused with the superstrings of string theory!

Chapter 9

Inflation

Liddle Ch. 13

In the last chapter, we discussed the way in which the Universe in which we live started out in a very special state: nearly flat, nearly homogeneous, and dominated by radiation. It is of course possible that these initial conditions are just a raw fact that we have to learn to deal with, but we would prefer to find a causal mechanism to enforce these conditions. In about 1980, Alan Guth in the USA (and independently Alexei Starobinsky and Andrei Linde in the former USSR) came up with a mechanism that takes a broad range of initial conditions and makes them all look like a flat, homogeneous, radiation-dominated Universe — inflation. Furthermore, it was quickly realized that inflation also provided a mechanism to generate density fluctuations of just the right character to grow into the large-scale structure we observe in today's distribution of galaxies, as well as in the fluctuations in the CMB which were first observed in the early 1990s.

9.1 Accelerated Expansion

The basic idea of inflation is that a period of accelerated expansion takes a very small volume of the early Universe and blows it up so much and so quickly that any inhomogeneities or curvature in this volume are smoothed out, and the density of nonrelativistic particles is diluted. At the same time, any quantum fluctuations are blown up to macroscopic size, providing the seeds for large-scale structure.

How does accelerated expansion do all this? Note first that both the flatness and horizon problems are a result of the behavior of the quantity $(aH)^{-1}$, approximately equal to the comoving Hubble length. In a Universe with matter, radiation, and curvature this is also approximately equal to the horizon size and which (it seems) must always grow with time. But, of course, $(aH)^{-1} = 1/\dot{a}$, so that this quantity grows is exactly the statement that $\ddot{a} < 0$ (as long as we are in an expanding, not contracting, Universe)—the expansion of the Universe is decelerating.

Let us start by recalling our definition of the comoving horizon distance:

$$\chi_H = d_H/a = \int_0^t \frac{dt'}{a(t')} = \int_0^a \frac{da'}{a'} \frac{1}{a'H(a')} \,. \tag{9.1}$$

As we have seen, usually this quantity behaves like ("is approximately proportional to") the comoving horizon distance $1/(aH)$, and both of these are increasing functions of time. But for a fixed present day value $1/(a_0H_0)$ we can make χ_H as large as we want if we can make $1/(aH)$ decrease with time — because that makes it *increase* back towards $a = 0$ and we can therefore increase the value of the integral. If we can do this then the Hubble scale grows more slowly than any fixed comoving distance, as we show in a cartoon in Figure 9.1. After an accelerating expansion, the Hubble length has shrunk with respect to the comoving coordinates.

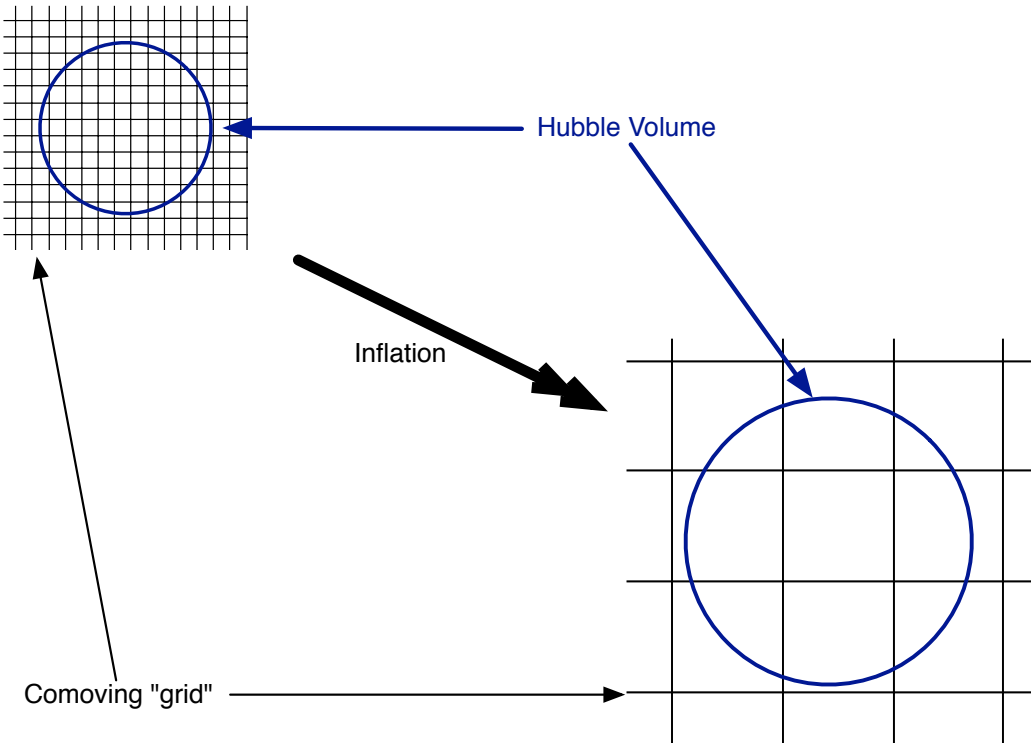

Figure 9.1: The Universe before and after inflation. The gridlines represents the comoving coordinate system, and the circle the Hubble volume $1/H$, which has shrunk with respect to the comoving coordinates after inflation.

We see how this solves the Horizon problem in Figure 9.2. The accelerating expansion means that the *Hubble* scale remains constant, but comoving scales increase much more rapidly. Note that in this case the *true horizon* scale is now much larger than the Hubble length.

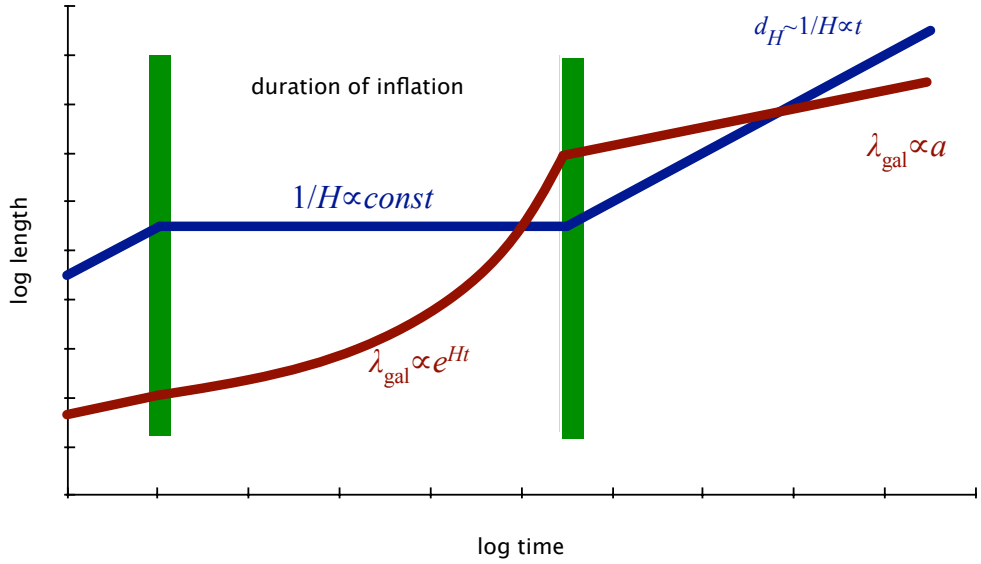

Figure 9.2: Scales entering and leaving the Hubble scale, which is the *apparent* (but not actual) horizon in a Universe with a period of accelerating expansion.

Thus, accelerating expansion should be able to solve the horizon problem. It is clear that it can also solve the flatness problem as it means that the value of $|\Omega_k| = |k|/(aH)$ gets driven closer and closer to zero while acceleration is occuring. Heuristically, this makes sense: if we rapidly expand a curved surface (relative to our coordinate system) it looks more and more flat in those coordinates.

Finally, acceleration solves the relic (monopole) problem in much the same way: it dilutes the number of massive relic particles in a given (physical) volume much faster than "ordinary" decelerating expansion. In fact, it is a little more complicated than this, because we still have to find a way to stop the accelerated expansion and make the Universe radiation-dominated after the period of accelerated expansion — this is called *reheating*.

After we discuss what kind of matter is necessary in order to make the Universe accelerate, we will return to these issues and calculate just how long inflation needs to last. But the idea of inflation is very simple: starting from a fairly generic state, the Universe undergoes accelerated expansion, which increases the scale factor by many orders of magnitude. However, this cools the Universe down proportional to the scale factor, so we must find a mechanism for reheating: stopping the accelerated expansion and converting the energy density of the Universe into radiation, at which point the Universe looks like a hot, radiation-dominated, flat big-bang model. We show these steps in Figure 9.3.

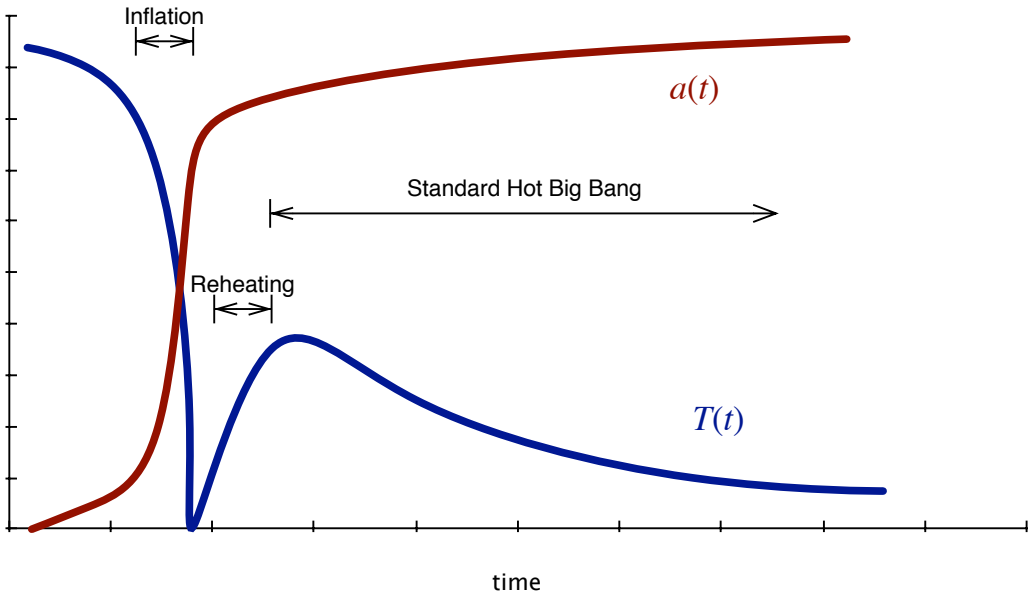

Figure 9.3: The scale factor and temperature in inflation and reheating, leading to a Universe that looks like a standard radiation-dominated hot big bang.

9.1.1 Acceleration and negative pressure

How do we realize an accelerating Universe in a physical system? Recall the second-order Friedmann equation for the acceleration:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{1}{3}\Lambda \,. \tag{9.2}$$

We want the right-hand side of this equation to be *positive*. Obviously the Λ term is sufficient for this — we have already seen that a Λ-dominated Universe is accelerating. But we have also seen that in the Universe as we have observed, Λ-domination has begun only recently, and this would not have solved the problems as outlined above — we need acceleration in the *early universe*.

So let us instead ignore the Λ term and concentrate on the pressure-density terms. For acceleration, we need

$$\rho + 3p < 0 \tag{9.3}$$

or an equation-of-state parameter

$$w = p/\rho < -1/3 \,. \tag{9.4}$$

If we assume positive density, we need *negative pressure*. In fact, a cosmological constant is exactly equivalent to matter with $w = -1$, and we will discuss models of inflation in which this is the case, but really any sufficiently negative equation of state will do.

For this $w = -1$ case, we have already seen how it affects the expansion, back in Chapter 3. It gives a de Sitter Universe with exponential expansion:

$$a \propto \exp\left[\sqrt{\Lambda_{\rm eff}/3}\, t\right] \propto \exp\left[\sqrt{8\pi G\rho_{\rm inf}/3}\, t\right] \propto e^{Ht}\ , \tag{9.5}$$

where in the last equality we use

$$H = \frac{\dot{a}}{a} = \sqrt{\Lambda_{\rm eff}/3} = \sqrt{8\pi G\rho_{\rm inf}/3} \tag{9.6}$$

and have also defined the effective density $\rho_{\rm inf} = \Lambda_{\rm eff}/(8\pi G)$. Here, we are not talking about a true cosmological constant but just some effective $\Lambda_{\rm eff}$, realized as some substance (field or particle), the **inflaton**, with density $\rho_{\rm inf}$ and pressure $p_{\rm inf} = -\rho_{\rm inf}$, but any

We can now make a timeline for inflation. At very early times the Universe is radiation-dominated (although it is possible that we could have curvature or matter domination). Just as the late-time Universe becomes dominated by Λ, eventually the very early Universe becomes dominated by the inflaton, at or around the GUT scale, 10^{15} GeV, corresponding to $t \sim H^{-1} \sim 10^{-34}$ s. This transition (which is similar to a phase transition) causes inflation to start, lasting until something like $t \sim 10^{-32}$ s (a factor of 100 in age — in the next section, we shall discuss in more detail how much inflation is needed), during which $H \sim$ const, driving exponential expansion.[1] This is similar to a phase transition in solid-state physics, which can also build in long-range correlations, exactly as we are trying to do here. This would give something like $a(t_f) = \exp(100)a(t_i)$ where i and f refer to the initial and final times. Over this period, therefore, the volume increases by $\exp 300 \sim 10^{130}$ and the temperature decreases by $\exp(-100)$. We therefore need to reheat the Universe to a high temperature $T_{RH} \lesssim T_i$, which converts the entire energy density into radiation particles. Before inflation, a single Hubble volume contained approximately 10^{14} radiation particles (using the black-body density); after inflation and reheating, that same comoving region has 10^{130+14} particles of radiation, vastly increasing the entropy density of the Universe.

9.1.2 The duration of Inflation

How long does inflation need to last in order to solve these big bang problems?

Before inflation, we have scale factor $a = a_i$, $t = t_i \sim H_i^{-1}$, with this scale corresponding to the physical size of the pre-inflationary causal horizon. During inflation, we have $a \propto \exp(H_i t)$, so right after inflation at $t_f = t_i + \delta t \gg t_i$, that pre-inflationary horizon has now expanded to physical size $d_f = \exp(H\delta t)H_i^{-1}$, and the Universe has a post-inflationary reheating temperature T_{RH}. Since then, the universe has expanded by $T_{RH}/(3\ {\rm K})$, so today, the original causal horizon now has size

$$d = \exp(H\delta t)H_i^{-1}\frac{T_{RH}}{3\ {\rm K}} = \exp(N)H_i^{-1}\frac{T_{RH}}{3\ {\rm K}}\ . \tag{9.7}$$

[1] There are also models of inflation where the expansion is still accelerating, but only as a power-law, known as "extended inflation".

where we have set $N = H\delta t$ as the "number of e-folds" that inflation lasts. We have plentiful evidence that the universe is isotropic at or near the current Hubble length, so we need

$$\begin{aligned} d &> H_0^{-1} \\ e^N &> H_0^{-1} \frac{3\ \mathrm{K}}{T_{RH}} \frac{T_{RH}^2}{m_{\mathrm{Pl}}} \\ &> \frac{3\ \mathrm{K}}{H_0} \frac{T_{RH}}{m_{\mathrm{Pl}}} \end{aligned} \tag{9.8}$$

where we have used $H \sim T^2/m_{\mathrm{Pl}}$ appropriate for an RD universe. With $H_0 \sim 10^{-33}$ eV and $T_0 \sim 10^{-4}$ eV, this gives

$$N > N_{\mathrm{min}} = 68 + \ln(T/M_{\mathrm{Pl}}) = 55 + \ln(T/10^{14}\ \mathrm{GeV})\ . \tag{9.9}$$

This is what is needed to solve the horizon problem. In order to solve the flatness problem we can make a similar analysis. Before inflation, we start with some initial value of the curvature parameter, $\Omega_k(t_i) = -k/(a_i H_i)^2$. Today, the scale factor is $a_0 = a_i e^N T_{RH}/T_0$, accounting for inflationary and ordinary expansion. So the current curvature

$$\begin{aligned} \Omega_k(t_0) &= \frac{-k}{(a_0 H_0)^2} = \frac{-k}{(H_0 a_i e^N T_0/T_{RH})^2} = \frac{-k}{(a_i H_i)^2} \left(e^{-N} \frac{T_0}{T_{RH}} \frac{H_i}{H_0} \right)^2 \\ &= \Omega_k(t_i) \left(e^{-N} \frac{T_0}{H_0} \frac{T_{RH}}{m_{\mathrm{Pl}}} \right)^2 = \Omega_k(t_i) e^{2N_{\mathrm{min}} - 2N} \end{aligned} \tag{9.10}$$

where N_{min} is the minimum number of e-folds required to solve the horizon problem, Eq. 9.9. Thus, if $\Omega_k(t_i)$ is of order one, the same amount of inflation that solves the horizon problem will solve the flatness problem; increasing the initial curvature parameter by orders of magnitude requires only increasing the number of e-folds logarithmically. In fact, realistic inflation models tend to last much longer than N_{min}, so this is easily satisfied.

Finally, inflation decreases the relic abundance down to acceptable levels from $\Omega_{\mathrm{mon}}(t_0) \sim 10^{11}$ (Eq. 8.37). The monopoles are produced at T_{GUT} with number density $n(t_{\mathrm{GUT}}) \sim H_{\mathrm{GUT}}^3$ which is diluted by a factor $(T_i/T_{\mathrm{GUT}})^3$ by the beginning of inflation, a factor of $\exp(-3N)$ during inflation, and subsequently by another factor of $(T_0/T_{RH})^3$. All together, this is decreased by a factor of $\exp(-3N) T_i^3/T_{RH}^3$ from the no-inflation prediction of Eq. 8.37. We must have $T_i > T_{RH}$ (or otherwise you would inflate again after reheating), but typically the two are comparable, so the main effect is from the exponential. We need $3N \gtrsim 11 \ln 10 \sim 25$, which is easily satisfied if we already solve the horizon and flatness problems.

Thus, we see that generically we need something like 60 e-folds to solve the various problems. To put this another way, with at least that much expansion, very generic initial conditions (inhomogeneous, curved, lots of heavy particles) are funnelled into what seems to be a very special state: smooth, flat, and radiation-dominated.

9.2 Inflation via a scalar field

Now, we'll consider a specific model of inflation, one that will further let us see how inflation not only solves our initial-condition problems, but also provides the seeds for the formation of density perturbations. It is important to realize that in order to get acceleration, *classical* (albeit relativistic) field theory suffices; only when we consider the end of inflation — reheating and the seeding of density perturbations does quantum mechanics enter.

The Lagrangian density of a scalar field, φ with potential $V(\varphi)$ is

$$\mathcal{L}(\varphi) = \frac{1}{2}\partial^\mu\varphi\partial_\mu\varphi - V(\varphi) \tag{9.11}$$

and the stress-energy tensor is then given by

$$T^{\mu\nu} = \partial^\mu\varphi\partial^\nu\varphi - \mathcal{L}g^{\mu\nu} \tag{9.12}$$

where $g^{\mu\nu}$ is the metric. For comparison, recall that we wrote down the stress-energy tensor of a perfect fluid with pressure p and density ρ:

$$T^{\mu\nu} = \mathrm{diag}(\rho, -p, -p, -p)\ . \tag{9.13}$$

If we assume that the field is spatially homogeneous, in fact the stress-energy takes on the perfect fluid form, with

$$\begin{aligned}\rho_\varphi &= \frac{1}{2}\dot\varphi^2 + V(\varphi)\\ p_\varphi &= \frac{1}{2}\dot\varphi^2 - V(\varphi)\ ;\end{aligned} \tag{9.14}$$

the correction terms for spatial inhomogeneities are of order $\nabla\varphi^2/a^2$ where ∇ is the comoving-coordinate gradient. Because of the $1/a^2$ factor, the effect of any spatial gradient is quickly rendered irrelevant by exponential accelerated expansion.

We can derive the equation of motion in several ways. For a general scalar field, we could vary the action, $S = \int d^4x\sqrt{-g}\mathcal{L}$, where g is the determinant of the metric tensor or we could use the covariant conservation of the stress-energy tensor $\nabla_\nu T^{\mu\nu} = 1$. But both of these require a bit more relativity than we've developed here. Instead, we can just plug in these expressions for the pressure and density into the fluid equation

$$\dot\rho + 3H(\rho + p) = 0 \tag{9.15}$$

which gives

$$\ddot\varphi + 3H\dot\varphi + V'(\varphi) = 0\ . \tag{9.16}$$

If it were not for the friction term proportional to H ("Hubble drag") note that this is exactly the same as the equation of motion for the position of a particle in a one-dimensional potential V, so we can use our intuition from that situation.

During inflation, we are looking for solutions that look like $p \approx -\rho$, which means that the time-derivative terms must be negligible, so $\rho \approx +V$, $p \approx -V$. In order to do this, the velocity $\dot{\varphi}$ must be negligible — the potential must be very flat. However, for inflation to end, the potential cannot be completely flat — it must eventually fall into a potential well where it can oscillate, which corresponds to a mass in quantum mechanics — a term in the potential like

$$V(\varphi) = \frac{1}{2} m_\varphi^2 \varphi^2 + \cdots \tag{9.17}$$

gives a mass, so nearly any concave potential gives mass to the field. Hence we expect that a potential of the form given in Figure 9.4 will have the right properties.[2] In fact we further expect that the φ field should couple to other particles, which would induce another "friction" term $\Gamma\varphi$ on the left-hand side of the equations of motion, where Γ represents the decay rate of φ into other particles. If these particles are light, we end up with a radiation-dominated Universe.

When the field is sitting on the approximately flat part of the potential, we say that it is in the "slow-roll" regime, with $V'(\varphi)$ very small.

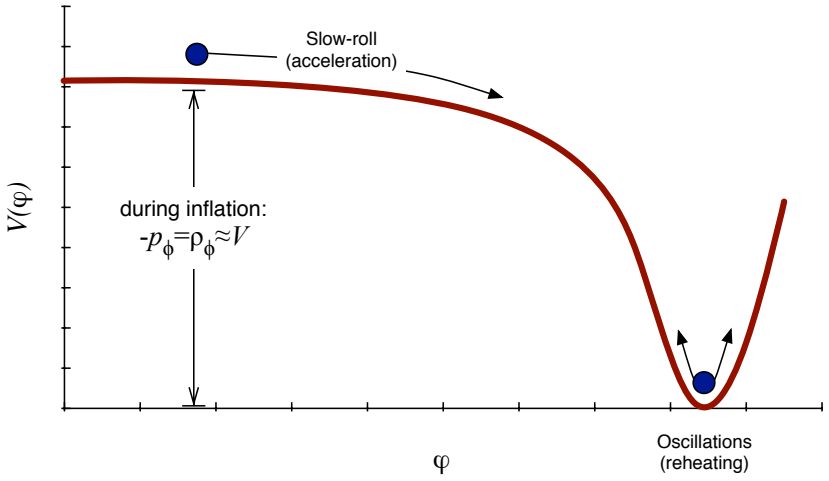

Figure 9.4: The inflationary potential. The slow-roll regime is when the potential is relatively flat, and the reheating regime is when it strongly curved.

9.2.1 Density Perturbations

Very soon after inflation was invented, it was realized that it gives a mechanism to answer yet another open question about the initial conditions for our hot big bang Universe: how

[2] Guth's original model of inflation had a potential barrier in between the accelerating and reheating regimes, which required the nucleation of bubbles in a phase transition to end inflation, but the nucleation rate was too low to end inflation properly.

were the initial fluctuations which have subsequently grown into the large-scale structure we observe in the Universe today generated?

The basic mechanism is very simple, and is a relatively generic consequence of the combination of inflation's accelerated expansion with the randomness of quantum mechanics. Any system will exhibit quantum fluctuations on very small scales. Inflation takes these fluctuations — initially on very small scales — and blows them up very rapidly. Once they are larger than the apparent horizon (the Hubble length, $H^{-1} \sim ct$), they are frozen in and behave as completely classical fluctuations. (Even though the *actual* horizon is much larger, the behavior of fluctuations at any given time is still controlled by the speed of light.) Outside the horizon, fluctuations can only evolve in a very simple way, due to the finite speed of light, so only when a scale re-enters the Horizon do overdensities (lumps) begin to collapse.

Without quantum field theory it is difficult to do the calculation precisely, but we can at least make a plausible argument using dimensional analysis. We start with our scalar field, which we will split into

$$\varphi = \varphi_{\rm cl} + \delta\varphi_{\rm QM} \tag{9.18}$$

where "cl" labels the classical evolution, and "QM" are the quantum fluctuations. We wrote down the Lagarangian density for our scalar field, $\mathcal{L} = \partial^\mu\varphi\partial_\mu\varphi/2 - V(\varphi)$ (Eq. 9.11), so the units on the scalar field are

$$[\varphi] = (\text{length})^{-1} = (\text{time})^{-1} \ . \tag{9.19}$$

If we are really in the slow-roll regime, then there is only one quantity with these units in our problem, and that is the (exponential) expansion rate, $H = \text{const}$. Since we know that the average of our quantum fluctuations should be zero, this implies that we can use this to fix the variance of our fluctuations, $\langle\delta\varphi^2\rangle \sim H^2$. In fact, a more careful analysis gives

$$k^3 P_\varphi(k)/(2\pi^2) = \langle\delta\varphi^2\rangle_k \simeq \left(\frac{H}{2\pi}\right)^2 \ . \tag{9.20}$$

In this equation, $P_\varphi(k)$ is the power spectrum of the φ field at spatial frequency k, using Fourier-transform conventions to be defined in the next chapter. The factor of 2π on the right-hand side arises because this is actually the so-called Gibbons-Hawking temperature associated with the horizon in a de Sitter Universe (the equivalent of the Hawking temperature associated with the horizon of a black hole), $T_{\rm GH} = H/(2\pi)$.

Because H is approximately constant during inflation, we can integrate this expression to get the total mean-square fluctuation in φ, integrated over all frequencies (which are sometimes called "modes"), which ends up giving

$$\langle\delta\varphi^2\rangle_k \simeq N \times \left(\frac{H}{2\pi}\right)^2 \ . \tag{9.21}$$

where N is the number of e-folds of inflation from before. In the limit of exactly exponential (de Sitter) expansion, this would result in an initial *power spectrum of density fluctuations* (which we will define more precisely in the next chapter) with the so-called

Harrison-Zel'dovich spectrum, $P(k) \propto k$. Of course, these equations can only be approximate: inflation has to end at some point, so we cannot really have $H = \text{const}$. This will result in a more realistic initial spectrum,

$$P(k) \propto k^{n_s} \,, \tag{9.22}$$

where n_s is the scalar spectral index (another name for density perturbations are *scalar* perturbations, corresponding to actual fluctuations in the curvature of the spacetime manifold, described by a scalar number), and we usually have n_s just below one, with details depending on $V(\varphi)$. The amplitude of the spectrum depends on the coupling constants present in $V(\varphi)$, and are constrained to be quite small in most models by the relatively small amplitude of temperature and density fluctuations observed in the CMB and on large scales today.

We will see in the next chapter that it is straightforward to understand the evolution of the power spectrum of such fluctuations once they exist.

Moreover, for a weakly-coupled scalar field (and we have just noted that observations seem to require weak coupling), the distribution of these fluctuations will be very close to Gaussian (in field theory, a *free* field is exactly Gaussian). Hence, once we have described these second moments, there is no more information available to us.

In a similar manner, inflation also creates a background of gravitational radiation (gravitons, or "tensors"). Gravitational radiation does not directly create lumps and voids (it does not couple directly to the density of matter) but the movements it induces are indeed visible, although as yet undetected, as patterns in the polarization of the CMB. The gravitational radiation is described by a separate power spectrum,

$$P_T(k) \propto k^{n_t} \tag{9.23}$$

where now n_t is usually just below *zero*, and the amplitude is governed by the value of $V(\varphi)$ during inflation, i.e., by the energy scale of inflation.

The observation of these tensor modes via CMB polarization is one of the main goals of the next generation of CMB experiments (beyond the Planck Satellite).

Chapter 10

Structure Formation

Schneider Ch. 7

10.1 Notation and Preliminaries

So far, we have been largely discussing a homogeneous Universe, with matter density $\rho_m = \bar{\rho}_m = \Omega_m \rho_{\rm crit} \simeq 10^{-29}{\rm g/cm}^3$, where we use an overbar to refer to the mean matter density. In the galaxy, the average density is about one million times higher than the average, and here on earth, the average density if, of course, even higher than that of water (1 g/cm^3). We clearly need to understand how the deviations from the mean grow in different circumstances and on different scales.

We will need some notation that will help us separate the mean density from the fluctuation. We define the *density contrast*, or *fractional density perturbation*,

$$\delta(\mathbf{x}, t) \equiv \frac{\delta\rho}{\rho}(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \tag{10.1}$$

If the Universe really is homogeneous, it is easy to define what we mean by $\bar{\rho}$, but as soon as we allow fluctuations (as we must in order to describe the real Universe), it becomes more complicated. In relativity, this is related to the fact that we are free to use any set of coordinates that we wish and the physics will stay the same — but the equations can look very different. One way to see this is to notice that (at least in the early Universe when fluctuations were small) it is possible to define a coordinate system in which $\rho = \bar{\rho}$ is a function of time, but *not* position. Fluctuations generically grow with time, so if, in one coordinate system, two nearby points have different densities $\rho(\mathbf{x}_1, t_1) < \rho(\mathbf{x}_2, t_1)$ at time t_1, then there is some time $t_2 < t_1$ at which $\rho(\mathbf{x}_1, t_1) = \rho(\mathbf{x}_2, t_2)$ (note the *different* times). But in GR we are free to redefine our time coordinate and give these time the *same* time label $\hat{t}$. If we use $\hat{t}$ to define our averages $\bar{\rho}(\hat{t})$ then of course we now find that $\delta = 0$, different than we would have found using the t time coordinate. Although the description is very different, it turns out that this freedom is mathematically very similar to the gauge freedom in field theories like general relativity.

Just as in electromagnetism, there are certain conventional choices of background coordinates ("gauge") that are useful in different circumstances. One is called the *syn-*

chronous gauge in which all locations use the proper time of an observer located their as the time coordinate. Another is the *Newtonian gauge* in which the equation of motion for the perturbations looks most like the Poisson equation of nonrelativistic gravity.

It turns out that well inside the apparent horizon, well away from regions of very strong gravity such as black holes, and at speeds considerably lower than c, most reasonable choices of the gauge agree. It is mostly outside the horizon that the choice matters. However, we shall be careful to use only physical "gauge-invariant" descriptions on such large scales and not use equations that refer to specific coordinate systems.

It is clear that our understanding of any physical theory will only be such that it will predict the *statistics* of the density perturbation, rather than the actual values as a function of our time and space coordinates. Hence, we need to consider suitable averages of δ and functions of δ.

Given our definition of $\delta(\mathbf{x}, t)$ it is obvious that the spatial average at a given time requires

$$\langle \delta(\mathbf{x}, t) \rangle = 0 \ . \tag{10.2}$$

(Actually, there is a subtlety, as there are three different averages that we could contemplate here. The first is a over all space at a given time; the second is an average with respect to the probability distribution of δ. (The equivalence of these two is mathematically similar to ergodic theory in thermodynamics.) Of course, neither of these two average can be performed in the real Universe, in which case averages are over observable parts of the Universe. We shall assume no distinction between these possibilities in practice.)

The next moment, the variance, will be nonzero:

$$\langle \delta(\mathbf{x}, t)\delta(\mathbf{y}, t) \rangle = \xi\left(\mathbf{x}, \mathbf{y}, t\right) = \xi\left(\left|\mathbf{x} - \mathbf{y}\right|, t\right) \ . \tag{10.3}$$

We have defined ξ, the density *correlation function*, which is also the second moment of the density distribution. It is only a function of the distance between the points $\mathbf{x}$ and $\mathbf{y}$, which is a statement about the *statistical isotropy* of the Universe. Our underlying description of the Universe should be invariant if we shift the origin and orientation. (This is the spatial equivalent to a stationary random process in time.)

The correlation structure is much simpler if we consider the Fourier Transform of δ (as are the equations of motion, as we will see soon):

$$\tilde{\delta}(\mathbf{k}, t) = \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}} \delta(\mathbf{x}, t) \tag{10.4}$$

and the inverse

$$\delta(\mathbf{x}, t) = \int \frac{d^3k}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{x}} \tilde{\delta}(\mathbf{k}, t) \ . \tag{10.5}$$

With these definitions, we can calculate the correlation function of our Fourier-Transformed

quantities, all evaluated at a single time, so we suppress our time coordinate:

$$\begin{aligned}\left\langle \tilde{\delta}(\mathbf{k})\tilde{\delta}(\mathbf{k}')\right\rangle &= \left\langle \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}}\delta(\mathbf{x}) \times \int d^3x' e^{i\mathbf{k}'\cdot\mathbf{x}'}\delta(\mathbf{x}')\right\rangle \\ &= \int d^3x \int d^3x' e^{i(\mathbf{k}\cdot\mathbf{x}+\mathbf{k}'\cdot\mathbf{x}')}\langle\delta(\mathbf{x})\delta(\mathbf{x}')\rangle \\ &= \int d^3x \int d^3x' e^{i(\mathbf{k}\cdot\mathbf{x}+\mathbf{k}'\cdot\mathbf{x}')}\xi\left(|\mathbf{x}-\mathbf{x}'|\right) \\ &= \int d^3y \int d^3x' e^{i(\mathbf{k}\cdot\mathbf{y}+\mathbf{k}\cdot\mathbf{x}'+\mathbf{k}'\cdot\mathbf{x}')}\xi(y) \\ &= \int d^3x' \, e^{i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}'} \int d^3y \, e^{i\mathbf{k}\cdot\mathbf{y}}\xi(y) \\ &= (2\pi)^3\delta_D(\mathbf{k}+\mathbf{k}')P(k) \, , \end{aligned} \tag{10.6}$$

where δ_D is the Dirac delta function. Because δ is real (i.e., not complex), we could also use $\delta(-\mathbf{k}) = \delta^*(\mathbf{k})$ and write this as

$$\left\langle \tilde{\delta}(\mathbf{k})\tilde{\delta}^*(\mathbf{k}')\right\rangle = (2\pi)^3\delta_D(\mathbf{k}-\mathbf{k}')P(k) \, . \tag{10.7}$$

These equations define the *power spectrum* of density fluctuations,

$$P(k) = \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}}\xi(x) = 4\pi \int dx \, x^2 \frac{\sin kx}{kx}\xi(x) \tag{10.8}$$

which turns out to be the three-dimensional Fourier transform of the correlation function considered as a function of only one variable. Because $\xi(x)$ is only a function of the magnitude of the length vector, $x = |\mathbf{x}|$, $P(k)$ is similarly only a function of the magnitude of the wavenumber, $k = |\mathbf{k}|$ (since we can actually do the angular part of the integral). The correlation function and power spectrum are the second-order moments of the density. If the density is described by a Gaussian distribution, then these second-order moments (along with the means $\langle\delta\rangle = 0$) completely describe the matter density distribution.

Note that we are talking about the distribution of the *function* $\delta(\mathbf{x})$, which we can think of as an infinitely-multivariate probably distribution $\Pr[\delta(\mathbf{x}_1), \delta(\mathbf{x}_2), \delta(\mathbf{x}_3), \ldots]$ where the $\mathbf{x}_i$ enumerate the uncountably infinite number of possible positions. At least for a Gaussian, this functional distribution is just described by our power spectrum or correlation function and no more. In Fourier space, this is especially simple: the δ function in Eq. 10.7 means that the values are uncorrelated, and so we can just write

$$\Pr\left[\tilde{\delta}(\mathbf{k})\right] = \frac{1}{\sqrt{2\pi P(k)}} \exp\left[-\frac{1}{2}\frac{\left|\tilde{\delta}(\mathbf{k})\right|^2}{P(k)}\right] \tag{10.9}$$

We can also calculate the statistics of some more physically relevant quantities. Consider the density fluctuation measured not at a point, but in spheres of some radius R. At any point, we can measure the mass in that sphere

$$\frac{\delta M_R}{M_R}(\mathbf{x}) = \frac{1}{4\pi R^3/3}\int_{y<R} d^3y \, \delta(\mathbf{x}+\mathbf{y}) = \int d^3u \, W_R(\mathbf{u}-\mathbf{x})\delta(u) \tag{10.10}$$

where in the second equality we define the *window function*, $W_R(\mathbf{x})$, which is $1/V$ within the volume, and 0 outside. We can now calculate the second moments of the M_R distribution from the second moments of δ. First note that Eq. 10.10 is a convolution, so in Fourier space the transform of $\delta M_R/M_R$ is just the product of the transforms of δ and the window function. Hence

$$\frac{\delta M_R}{M_R}(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\, e^{i\mathbf{k}\cdot\mathbf{x}} \tilde{\delta}(\mathbf{k}) \tilde{W}_R(\mathbf{k}) \tag{10.11}$$

with

$$\begin{aligned}
\tilde{W}_R(\mathbf{k}) &= \frac{1}{4\pi R^3/3} \int d^3x\; e^{i\mathbf{k}\cdot\mathbf{x}} W_R(x) \\
&= \frac{1}{4\pi R^3/3} \int_{x<R} dx\; x^2 d\phi\; d\cos\theta\; e^{ikx\cos\theta} \\
&= \frac{3}{(kR)^2}\left(\frac{\sin kR}{kR} - \cos kR\right) \\
&\equiv \tilde{W}(kR)
\end{aligned} \tag{10.12}$$

(in particular, $\tilde{W}(0) = 1$, so the zero-radius sphere is equivalent to just using δ itself), so

$$\begin{aligned}
\left\langle \left(\frac{\delta M}{M}\right)^2 \right\rangle_R &= \int \frac{d^3k}{(2\pi)^3} \frac{d^3k'}{(2\pi)^3} \left\langle \left| \tilde{\delta}(\mathbf{k}') \tilde{W}_R(\mathbf{k}') \right|^2 \right\rangle \\
&= \int \frac{d^3k}{(2\pi)^3} \left|\tilde{W}(kR)\right|^2 P(k) = \int \frac{dk}{k} \left|\tilde{W}(kR)\right|^2 \frac{k^3}{2\pi^2} P(k) \\
&= \int \frac{dk}{k} \left|\tilde{W}(kR)\right|^2 \Delta^2(k)
\end{aligned} \tag{10.13}$$

where we define $\Delta^2(k) = k^3 P(k)/(2\pi^2)$, which is the contribution per logarithmic integral to the mean-square fluctuation, and which has no units. (Recall that we encountered the similar combination $k^3 P_\varphi(k)/(2\pi^2)$ related to the inflaton field in Eq. 9.20). Roughly speaking, $\Delta^2(k)$ gives the density fluctuation on a length scale $L \sim 2\pi/k$.

10.2 Spherical Collapse

To get a feel for the way perturbations evolve, let's consider an idealized situation: a perfectly spherical perturbation in an otherwise homogeneous, flat, $\Omega_m = 1$, Universe with density $\bar{\rho}(t)$. We can do this by picking some point which we will take to be $r = 0$ and compressing all of the matter within $r < r_2$ to a higher uniform density, so it now has a radium $r_1 < r_2$. The average density within r_2 is the same as before, $\bar{\rho}$, but within r_1, the density is $\rho > \bar{\rho}$.

But we saw near the beginning of the course in Section 2.3 that the mass inside a uniform density sphere is unaffected by the mass outside it. Hence, we know that the mass inside the sphere must behave like a Universe with a higher density — and hence

a higher Ω. But our mean density is $\Omega = 1$, and hence our overdense sphere will behave like an $\Omega_m > 1$ Universe — it will expand and recollapse.

Outside of our carved-out sphere, $r > r_2$ we just have the usual equations (with an overbar to represent the mean):

$$\left(\frac{\dot{\bar{a}}}{\bar{a}}\right)^2 = \frac{8\pi G}{3}\bar{\rho} \tag{10.14}$$

which, for nonrelativistic matter gives

$$\bar{a} \propto t^{2/3} \qquad \text{and } \bar{\rho} = \frac{1}{6\pi G t^2} \propto \bar{a}^{-3} \, . \tag{10.15}$$

For $r < r_1$, we instead have

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{10.16}$$

and consider some early time t_i at which $a = a_i$, $\Omega = \Omega_i$ and $\rho = \rho_i$ such that $\rho_i \gg k/a_i^2$ so that within the perturbation $\Omega_i \simeq 1$ and $\rho_i \simeq \bar{\rho}(t_i)$ — i.e., we start with a small perturbation.

We saw in a problem sheet that we could calculate the maximum scale factor before the recollapse of a closed Universe, and the time at which this happens:

$$\frac{a_{\rm max}}{a_i} = \frac{\Omega_i}{\Omega_i - 1} \qquad H_i t_{\rm max} = \frac{\pi}{2}\frac{\Omega_i}{(\Omega_i - 1)^{3/2}} = \frac{\pi}{2}\left(\frac{a_{\rm max}}{a_i}\right)^{3/2} \Omega_i^{-1/2} \tag{10.17}$$

so

$$\rho_{\rm max} = \left(\frac{a_i}{a_{\rm max}}\right)^3 \rho_i = \frac{3\pi}{32 G \Omega_i^{1/2} t_{\rm max}^2} \tag{10.18}$$

which we can compare to $\bar{\rho}(t_{\rm max}) = 1/(6\pi G t_{\rm max}^2)$ so

$$\frac{\rho_{\rm max}}{\bar{\rho}(t_{\rm max})} = \frac{9\pi^2}{16\Omega_i^{1/2}} \simeq \frac{9\pi^2}{16} \simeq 5.55 \, . \tag{10.19}$$

This is equivalent to $\delta_{\rm max} = 4.55$. If instead we concentrate not on the full collapse, but just on the very earliest times after t_i, we find that $\delta = 1 - 1/\Omega \propto a(t) \propto t^{2/3}$. These small overdensities are called the "linear regime" for reasons that will be more obvious in the next section. We can usually calculate things much more readily in the linear regime, and so it is common to compare this nonlinear (but idealized) case with what would happen if linear evolution continued. Between t_i and $t_{\rm max}$, we would have $\delta_{\rm Lin}/\delta_i = \bar{a}_{\rm max}/\bar{a}_i = (t_{\rm max}/t_i)^{2/3} = (3\pi/4)^{2/3}(a_{\rm max}/a_i) \simeq 1.77/\delta_i$. So, irrespective of the starting conditions, maximum density (a.k.a. "turnaround") occurs when the *linear* density contrast would have been $\delta \simeq 1.77$

In a perfectly uniform universe, this overdensity would just collapse down to a point. But in a more realistic scenario, it will *virialize* and convert its gravitational energy into random kinetic energy (by a process known as "violent relaxation"!). At the maximum, with $KE = 0$, the overdensity has total energy $PE = E_{\rm grav} \simeq -3GM/(10 r_{\rm max})$. After virialization to form a sphere of radius $r_{\rm vir}$, we have $KE = -PE/2$ and therefore a total energy $E_{\rm grav}/2 \simeq -3GM/(20 r_{\rm vir})$. For the energy to remain constant, we need $r_{\rm vir} \simeq r_{\rm max}/2$ or $\rho_{\rm vir} \simeq \rho_{\rm max}/8$.

10.3 Linear Perturbations

Now we will consider a more general situation allowing an arbitrary $\delta(\mathbf{x}, t)$, although we will find that we cannot solve the equations for all possible times and values of the density — we will have to use perturbation theory.

10.3.1 Newtonian Theory—non-expanding

As a warm-up, let us consider the case of a non-expanding medium; this was first considered by Jeans in the early 20th Century. We will describe our fluid by its density, ρ, pressure, p and velocity $\mathbf{v}$; we also need to consider the gravitational potential Φ. All of these quantities are functions of position, $\mathbf{x}$ and time, t. First, we consider the *continuity equation*. The change in the mass in some volume V with surface A must be equal to the flux through the surface:

$$\frac{dM}{dt} = -\int_A \rho\mathbf{v}\cdot\mathbf{d}A$$
$$\int_V \frac{\partial\rho}{\partial t}\,dV = -\int_V \nabla\cdot(\rho\mathbf{v})\,dV \tag{10.20}$$

where we use Gauss' theorem, so, in differential form

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0 \qquad \textit{continuity}\,. \tag{10.21}$$

Next, we consider $\mathbf{F} = m\mathbf{a}$, which in this context is known as the *Euler equation*, with a gravitational force and a pressure force. The gravitational force is $F_g = -m\nabla\Phi$ and the pressure force is $F_p = -\int_V(\nabla p)dV$. By the chain rule, $\mathbf{a} = \dot{\mathbf{v}} = \partial\mathbf{v}/\partial t + (\mathbf{v}\cdot\nabla)\mathbf{v}$.

$$\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v} = -\frac{\nabla P}{\rho} - \nabla\Phi \qquad \textit{Euler}\,. \tag{10.22}$$

We also need the *Poisson equation* relating gravitational potential and matter density,

$$\nabla^2\Phi = 4\pi G\rho \qquad \textit{Poisson}\,, \tag{10.23}$$

and an *equation of state* linking the pressure and density, usually given in the form of an expression for the [adiabatic] speed of sound,

$$c_s^2 = \left(\frac{\partial p}{\partial\rho}\right)_{\rm adi} \qquad \textit{equation of state}\,. \tag{10.24}$$

These equations are too complicated to solve in complete generality. Instead, we will use perturbation theory. It is easy to see that a zeroth-order solution is given by $\mathbf{v} = 0$, $\rho = \rho_0 = \text{const}$, $p = p_0 = \text{const}$ (assuming no spatial variation in the equation of state) and $\nabla\Phi_0 = 0$[1] We will the write all of our variables as

$$\rho = \rho_0 + \rho_1 \qquad p = p_0 + p_1 \qquad \mathbf{v} = \mathbf{v}_1 \qquad \Phi = \Phi_0 + \Phi_1 \tag{10.25}$$

[1] Actually, this isn't true, as it contradicts the Poisson equation! This is sometimes called the *Jeans swindle*, and it is nonetheless a good approximation for what actually goes on. Moreover, the same issues do not arise in the expanding case we will discuss next.

where 0 refers to the unperturbed solution, and 1 to our small perturbations. To linear order (i.e., any products of first-order quantities, such as $\rho_1 \times \mathbf{v}_1$, are neglected), and subtracting off the zeroth-order solution, our equations become

$$\frac{\partial \rho_1}{\partial t} + \rho_0 \nabla \cdot \mathbf{v}_1 = 0 \tag{10.26a}$$

$$\frac{\partial \mathbf{v}_1}{\partial t} + c_s^2 \frac{1}{\rho_0} \nabla \rho_1 + \nabla \Phi_1 = 0 \tag{10.26b}$$

$$\nabla^2 \Phi_1 = 4\pi G \rho_1 \tag{10.26c}$$

If we take the divergence of (b) and substitute in (c) and the time derivative of (a), we get a single, second-order differential equation:

$$\frac{\partial^2 \rho}{\partial t^2} - c_s^2 \nabla^2 \rho_1 = 4\pi G \rho_0 \rho_1 \ . \tag{10.27}$$

This equation is straightforward to solve by considering plane-wave solutions (or, equivalently, writing these equations in terms of the spatial Fourier transform of ρ). In terms of our density perturbation $\delta = \rho_1/\rho_0$, we will try

$$\rho_1 = A \exp(i\mathbf{k} \cdot \mathbf{x} + i\omega t)\rho_0 \ . \tag{10.28}$$

Substituting this into Eq. 10.27 gives

$$\begin{aligned} -\omega^2 \rho_1 + k^2 c_s^2 \rho_1 &= 4\pi G \rho_0 \rho_1 \\ -\omega^2 + k^2 c_s^2 &= 4\pi G \rho_0 \end{aligned} \tag{10.29}$$

giving the dispersion relation

$$\omega^2 = c_s^2 k^2 - 4\pi G \rho_0 = c_s^2 (k^2 - k_J^2) \tag{10.30}$$

defining the *Jeans wavenumber*,

$$k_J = \frac{\sqrt{4\pi G \rho_0}}{c_s} \ . \tag{10.31}$$

If $k > k_J$, $\omega^2 > 0$ and the solution is oscillatory; if $k < k_J$, $\omega^2 < 0$ and the solution is exponentially growing or decaying (there is usually one of each as this is a second-order equation). In the limit $k \ll k_J$, $\rho \propto e^{\pm t/\tau}$ with timescale $\tau \simeq (4\pi G\rho)^{-1/2}$.

Basically, small-scale perturbations oscillate as *sound waves*, whereas large-scale perturbations can grow. Another way to see how the behavior changes is to compare the gravitational timescale $\tau_g \sim (G\rho)^{-1/2}$ with the pressure timescale ($\tau_p \sim \lambda/c_s \simeq (kc_s)^{-1}$). When $\tau_g < \tau_p$, collapse has time to occur before pressure can act to respond. We will see very similar behavior in the case of the expanding universe.

Note that we have found another reason why our Fourier analysis of Section 10.1 is valuable: individual Fourier modes evolve independently.

(I am ignoring the fact that we have not actually solved for the velocity field, and in fact there is some subtlety: from Eq. 10.26a, the density only allows us to solve for $\nabla \cdot \mathbf{v}$, which leaves $\nabla \times \mathbf{v}$ undetermined — this is just conservation of angular momentum, which, to linear order, doesn't couple to the density field.)

10.3.2 Perturbation theory in an expanding Universe

The generalization of the Jeans analysis to an expanding Universe is technically difficult, but conceptually not much different from the above case. We have to account for a few new features:

- The background evolution is not time-independent, but $\rho_m \propto a^{-3}$, $\rho_r \propto a^{-4}$ for NR matter and radiation;
- The continuity, Euler, and Poisson equations must be updated to take account of the possible presence of relativistic matter (e.g., radiation); and
- It is easiest to express the final differential equations in comoving coordinates.

In fact, we will simplify a bit from the most complex possible situation of multiple components (e.g., matter plus radiation), each possibly with its own fluctuations. Instead, we will largely concern ourselves with the fluctuations in the (NR) *matter* component, even when the Universe is radiation dominated.

We have already seen how to modify our equations to account for relativistic matter. In the continuity equation, we replace ρ with $\rho + p = \rho(1 + w)$ as in the homogeneous fluid equation, Eq. 2.42:

$$\dot\rho + (1+w)\nabla\cdot(\rho\mathbf{v}) = 0 \ . \tag{10.32}$$

In this case, the density and pressure refer to the *total* from all components. Each component will also satisfy its own separate conservation equation (e.g., the number of dark matter particles is conserved as long as we are long after freeze-out for them).

Similarly, we must replace ρ with $\rho + 3p$ in the Poisson equation, as in the second-order Friedmann equation Eq. 2.41 (which is, after all, an equation for the acceleration, so it should not be surprising that is this combination that is the relativistic source for gravitational acceleration):

$$\nabla^2\Phi = 4\pi G\rho(1+3w) \ . \tag{10.33}$$

In the expanding universe, we will write our perturbations as $\rho = \bar\rho(1+\delta)$, $p = p_0 + \delta p = \bar\rho(w + c_s^2\delta)$, $\mathbf{v} = \mathbf{v}_0 + \mathbf{u} = H\mathbf{x} + \mathbf{u}$ and $\Phi = \Phi_0 + \phi$.

We also need to change from fixed coordinates $\mathbf{x}$ to comoving (also called Lagarangian) coordinates $\mathbf{q}$. These are related by $\mathbf{x} = a\mathbf{q}$, with the unperturbed velocity $\mathbf{v}_0 = H\mathbf{x}$. We must calculate the time derivative of some function $f(\mathbf{x}, t)$,

$$\begin{aligned}\left(\frac{\partial f}{\partial t}\right)_{\mathbf{q}} &= \left(\frac{\partial f(a\mathbf{q},t)}{\partial t}\right)_{\mathbf{q}} \\ &= \left(\frac{\partial f}{\partial t}\right)_{\mathbf{x}} + \sum_i \left(\frac{\partial f}{\partial x_i}\right)\frac{d(aq_i)}{dt} \\ &= \left(\frac{\partial f}{\partial t}\right)_{\mathbf{x}} + (\mathbf{v}_0\cdot\nabla_{\mathbf{x}})\, f \end{aligned} \tag{10.34}$$

and the spatial derivatives are related by a simple change of variables,

$$\nabla_{\mathbf{x}} f = \frac{1}{a}\nabla_{\mathbf{q}} f \ . \tag{10.35}$$

We also need the specific cases $\nabla_{\mathbf{x}} \cdot \mathbf{x} = 3$, $\nabla_{\mathbf{x}} \cdot \mathbf{v}_0 = 3H$ and $(\mathbf{u} \cdot \nabla_{\mathbf{x}})\mathbf{v}_0 = H\mathbf{u}$.

With all of these developments, we can derive the equations for the NR matter perturbation in comoving coordinates, again to linear order,

$$\dot{\delta} + \frac{1}{a}\nabla \cdot \mathbf{u} = 0 \tag{10.36a}$$

$$\dot{\mathbf{u}} + \frac{\dot{a}}{a}\mathbf{u} + \frac{c_s^2}{a}\nabla\delta + \frac{1}{a}\nabla\phi = 0 \tag{10.36b}$$

$$\nabla^2\phi = 4\pi G a^2 \bar{\rho}_m \delta \ , \tag{10.36c}$$

which can be compared to the non-expanding case, Eq. 10.26. Note that in these equations $\delta = \delta\rho_m/\rho_m$ where ρ_m is the average NR matter density (not the total density). We allow for the possibility of a non-zero sound speed for the matter (e.g., for baryons).

We can again combine these to form a single second-order differential equation

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - \frac{c_s^2}{a^2}\nabla^2\delta - 4\pi G\bar{\rho}_m\delta = 0 \ . \tag{10.37}$$

Compared to the non-expanding, matter-only case, Eq. 10.27, this equation has a few new features. First, note the explicit presence of the factors of w from the continuity and Poisson equations. Second, there is a factor of $1/a^2$ in the spatial gradient term. Finally, there is a new term, $2H\delta$, occasionally referred to as "Hubble Drag", which will serve to slow the growth of perturbations compared to the exponential form in the non-expanding case.

Nonetheless, we can analyze this equation in much the same manner as before. By substituting in a plane-wave solution (or, equivalently, by Fourier transforming δ), we transform $\nabla^2 \to -k^2$ and get

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} + c_s^2\left[\frac{k^2}{a^2} - \frac{4\pi G\bar{\rho}}{c_s^2}\right]\delta = 0 \ . \tag{10.38}$$

Recall that we are in comoving coordinates, so k is measured in comoving coordinates; the *physical* wavenumber is k/a. Hence, as before, we can identify the physical Jeans wavenumber, defined by

$$k_J = \frac{\sqrt{4\pi G\bar{\rho}}}{c_s^2} \ . \tag{10.39}$$

Alternately, we can define the Jeans length $\lambda_J = 2\pi/k_J$, again in physical coordinates, or the Jeans mass $4\pi\lambda_J^3\bar{\rho}/3$. The analytic form of the solutions to this equation are somewhat more complicated than in the non-expanding case, but their character is the same. For small scales, $k \gg k_J$, we again have oscillatory solutions, with frequency approximately given by $\omega \simeq c_s k/a$. The exact form of this solution is a linear combination of Bessel functions, not a sinusoid, corresponding to a mildly damped sound wave, with the damping on a the scale of the Horizon.

In particular, the qualitative features of this analysis also apply to the perturbations to the dominant component in a radiation-dominated Universe. In this case, the Jeans length is comparable to the Horizon scale (since the sound-speed is comparable to the

speed of light). Hence, perturbations in a radiation-dominated Universe do not grow inside the horizon.

To examine the large-scale limit, we will take $k \to 0$ (in which case the equation becomes the correct linear equation even in the relativistic, superhorizon case). With vanishing spatial derivatives, we can write the solution as

$$\delta(\mathbf{q}, t) = \delta_0(\mathbf{q}) D(t) \tag{10.40}$$

where δ_0 gives the spatial part of the solution at some particular time (i.e., the initial conditions), and $D(t)$ is known as the *growth factor*. In the matter-dominated epoch, the equation for D is just

$$\ddot{D} + 2\frac{\dot{a}}{a}\dot{D} = 4\pi G\bar{\rho} D = \frac{3}{2}\left(\frac{\dot{a}}{a}\right)^2 D \ , \tag{10.41}$$

where we have used the first-order Friedmann equation to eliminate $\bar{\rho}$. This is a second-order equation, hence with two solutions. Often, one of these will grow in time, and the other will decay, and these are known, respectively, as the *growing mode* and *decaying mode*. In both matter- and radiation-dominated Universes, this has power law solutions. First, matter-domination, with $w = 0$, $a \propto t^{2/3}$, $\dot{a}/a = 2/(3t)$, has

$$D = C_1 t^{2/3} + C_2 t^{-1} \ . \tag{10.42}$$

There is a growing mode $D_+ \propto t^{2/3} \propto a$ and a decaying mode $D_- \propto t^{-1} \propto H$. Since the Jeans length is infinite, this growth applies on all scales when the pressure is negligible.

The more general equation, valid during radiation domination and on superhorizon scales (which can nonetheless be derived from the above Newtonian equations with the appropriate special-relativistic generalizations to include the effects of pressure), is

$$\ddot{D} + 2\frac{\dot{a}}{a}\dot{D} = \frac{3}{2}\left(\frac{\dot{a}}{a}\right)^2 (1+w)(1+3w) D \ , \tag{10.43}$$

In a radiation-dominated universe, now with $w = 1/3$, $a \propto t^{1/2}$, $\dot{a}/a = 1/(2t)$, so the solution to this equation is

$$D = C_1 t + C_2 t^{-1} \ , \tag{10.44}$$

with growing mode $D_+ \propto t \propto a^2$ and a decaying mode $D_- \propto t^{-1} \propto H$. We saw that the Jeans length in an RD universe is comparable to the Hubble scale, so only perturbations outside the (apparent) horizon grow in this way.

Multiple Components, Curvature and Dark Energy

In the previous analysis, we have assumed a single component in the Universe with $w = c_s^2$. We very often care about a more complicated case. For example, even in a universe dominated by curvature, radiation, or a cosmological constant, we often care about the behavior of the (pressureless) matter. In this case, each component is separately

conserved, and each contribute separately to the gravitational potential and to the force on any particle. The result of this is a slightly more complicated second-order differential equation

$$\ddot{\delta}_i + 2\frac{\dot{a}}{a}\delta_i + \left[\frac{c_i^2 k^2}{a^2}\delta_i - 4\pi G \sum_j \rho_j \delta_j\right] = 0 \ . \tag{10.45}$$

where the subscript labels the species and the sum is over all species, and $\delta_j = \delta\rho_j/\rho_j$ where ρ_j is the average of particle species j. (As written, this equation ignores the factors of w since we will use it below to calculate the evolution of the $w = c_s^2 = 0$ matter perturbations.)

In particular, we can use this for the common case of matter perturbations in a radiation-dominated universe. In this case, we care about $i = m =$ matter, so $c_m = 0$ and we can ignore the $4\pi G\rho$ term since $\sum_j \rho_j\delta_j = \rho_m\delta_m + \rho_r\delta_r$; the first term is suppressed since $\rho_m \ll \rho_r$ and the second since $\delta \simeq 0$ since (as we saw above) there are only sound waves, not growing perturbations, to the radiation itself. Hence, the differential equation becomes

$$\begin{aligned} 0 &= \ddot{\delta}_m + 2\frac{\dot{a}}{a}\delta_m \\ &= \ddot{\delta} + \frac{1}{t}\dot{\delta} \ . \end{aligned} \tag{10.46}$$

since there are no spatial derivatives, we can consider this as an equation for the growth factor, with solution

$$D_m = C_1 + C_2 \ln t \qquad RD \ . \tag{10.47}$$

The growing mode is no longer a power law, but logarithmic. This is sufficiently slowly that, essentially, matter perturbations do not grow during radiation-dominated periods.

We can similarly calculate the behavior of matter perturbations in an open curvature-dominated universe. We now go back to our original single-component Eq. 10.41, with $w = c_s^2 = 0$, but use the appropriate $a \propto t$. Now, the term $4\pi G\bar{\rho} = (3/2)\Omega_{\rm tot}(t)H^2$, so

$$0 = \ddot{\delta} + \frac{2}{t}\dot{\delta} + \frac{3}{2t^2}\Omega(t) \simeq \ddot{\delta} + \frac{2}{t}\dot{\delta} \tag{10.48}$$

where we ignore the last term since not only do we eventually have $\Omega(t) \ll 1$, but moreover it is falling with time so it must eventually become negligible. The solution is

$$D_m = C_1 + C_2/t \qquad CD \ . \tag{10.49}$$

Again, there is no growing mode.

Finally, we consider the behavior of matter perturbations in a universe dominated by a cosmological constant: $w = -1$, $a = \exp(Ht)$, $\dot{a}/a = H = $ const:

$$0 = \ddot{\delta} + 2H\dot{\delta} - 4\pi G\bar{\rho}\delta \simeq \ddot{\delta} + 2H\dot{\delta} \ , \tag{10.50}$$

where again, we find that we can ignore the final term, this time since $\bar{\rho}$ is quickly diluted by the exponential expansion, and the solution is

$$D_m = C_1 + C_2 e^{-2Ht} \qquad \Lambda D \ . \tag{10.51}$$

Yet again, we find that perturbations do not grow.

Relativistic Perturbation Theory

In the previous sections, we have assumed that we can apply our Newtonian calculation on all scales, even those larger than the apparent horizon. In principle, as discussed in Section 10.1 above, we should not necessarily be able to trust our calculation on those scales. In fact, the calculation we have done *is* applicable on superhorizon scales, at least for some particular choice of coordinates (i.e., gauge).

10.4 The processed power spectrum of density perturbations

10.4.1 Initial Conditions

We can now give another interpretation to the Harrison-Zel'dovich power spectrum, $P(k) \propto k$, defined in the previous chapter.[2] It such that the amplitude of density fluctuations for scales entering the horizon is a constant,

$$\delta_H^2(t) = \Delta^2(k = aH, t) = \frac{D^2(t)}{D^2(t_i)}\Delta_i(k = aH) = \text{const} \; . \tag{10.52}$$

Here, $k = aH$ is the comoving wavenumber corresponding to the comoving length scale $\sim 1/(aH)$ at time t, $D(t)$ is the growth factor calculated in Section 10.3.2, and t_i is some suitably early time (just after inflation, say). Because this combination doesn't change with time, the behavior is said to be scale-free (which is a slightly different and more restrictive use than in other fields of physics).

Using the results of the previous section, we can show that these two definitions are equivalent. In a radiation-dominated Universe, for scales outside the horizon, we have $D \propto a^2$, $a \propto t^{1/2}$, and $aH = \dot{a} \propto t^{-1/2} \propto a^{-1}$ so

$$\begin{aligned} \delta_H^2 &\propto a^4 (aH)^3 (aH)^{n_s} \propto a^4 \times a^{-3-n_s} \\ &\propto a^{1-n_s} \end{aligned} \tag{10.53}$$

which is indeed constant for $n_s = 1$. Similarly, in a matter-dominated Universe, $D \propto a$, $a \propto t^{2/3}$, $aH \propto t^{-1/3} \propto a^{-1/2}$, so

$$\begin{aligned} \delta_H^2 &\propto a^2 (aH)^3 (aH)^{n_s} \propto a^2 \times a^{-3/2-n_s/2} \\ &\propto a^{(1-n_s)/2} \end{aligned} \tag{10.54}$$

which is again constant for $n_s = 1$. In a realistic inflationary model, we expect n_s slightly below one; $n_s \simeq 0.95$ is observed.

[2] It is probably more properly referred to as the Harrison-Zel'dovich-Peebles-Yu spectrum after all the authors who discussed it more or less independently.

Curvature and isocurvature fluctuations

So far the discussion has concerned either a single, dominating, component, or a small admixture of matter in a radiation-dominated universe. In the former case, our density perturbations correspond directly to perturbations of the metric — the curvature of the manifold. These are often referred to as adiabatic perturbations, referring to the fact that the entropy (which resides almost entirely in the radiation) is constant.

But the matter/radiation case points out there are more possibilities. Consider a universe with two components each with a different equation of state, matter and radiation, say. It is possible to perturb the components so that $\delta\rho_m = -\delta\rho_r$; this gives a *net* perturbation $\delta\rho = \delta\rho_m + \delta\rho_r = 0$, so the manifold has no curvature. These are called *isocurvature perturbations*.

Outside the horizon, adiabatic and isocurvature perturbations behave very differently, as the microphysics encoded in the equation of state cannot act. At horizon crossing, the initial conditions are very different from $\delta_H = \text{const}$. Inside the horizon, the different components behave differently and, eventually, the isocurvature pertubations become actual perturbations to the curvature. Because of the different initial conditions, however, the shape of the processed power spectrum can be very different than for adiabatic initial conditions. At present, adiabatic perturbations seem to fit the data much better, as shown in Section 10.4.3 below.

10.4.2 The transfer function

Finally, we can combine all of the these calculations — the initial conditions and the evolution of different kinds of perturbations inside and outside the horizon — and calculate the power spectrum of density perturbations that we expect to observe at a given time. Since inflation (or, most likely, any other theory of the early Universe) only provides a statistical description of the initial conditions, that is the best we can hope for.

Under linear evolution, each mode $\tilde{\delta}(\mathbf{k}, t)$ evolves independently and hence the evolved density (as well as the potential, pressure and velocity) is a linear functional of the initial conditions. If the initial conditions were Gaussian, then so are the evolved quantities, albeit with a different correlation function or power spectrum. (However, we may need more than *only* the power spectrum if the initial conditions are non-Gaussian; similarly, nonlinear evolution would also require a non-Gaussian description.) We further expect that our growing-mode solutions will dominate after sufficient evolution. Hence we expect that we should be able to write our processed power spectrum as

$$P(k, t) = T_i^2(k) P(k, t_i) \ , \tag{10.55}$$

where t_i is some early initial time (i.e., when $P(k) = Ak^{n_s}$ is a good description over all scales). $T_i^2(k)$ is the *transfer function* which, for linear evolution in Fourier space, acts separately on each scale.

We need to combine information for scales inside and outside the horizon, before and after matter-radiation equality. We summarize those results in Table 10.4.2.

	Inside	Outside
RD	-	a^2
MD	a	a

Table 10.1: The growth of perturbations, inside and outside the apparent horizon during matter- and radiation-domination. The entries correspond to the behavior of the growing mode, $D_+(a)$ (there is no growth inside the Horizon during radiation-domination).

The transition from radiation to matter domination occurs at $1 + z_{\rm eq} = \Omega_m/\Omega_r$, at which time comoving Hubble length was $1/(aH)_{\rm eq}$, corresponding to a comoving wavenumber $k_{\rm eq} = a_{\rm eq}H_{\rm eq}$

We say a scale "enters the horizon" when its length is comparable to the Hubble length $1/H$ (note that this is really the *apparent* horizon if inflation has happened). In terms of our comoving wavenumber, this happens $k = (aH)_{\rm enter}$; during the post-inflation decelerating phase of evolution, the Hubble scale grows in comoving coordinates, so new scales are constantly entering the Horizon as we saw in Figures 8.1 and 9.2. We show the time evolution of the different scales in Figure 10.1

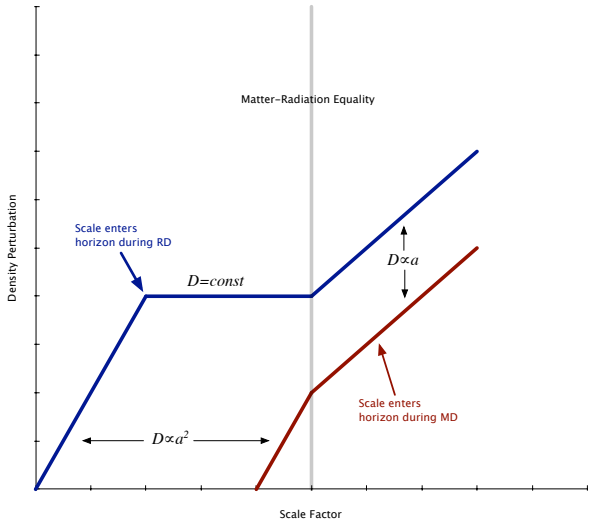

Figure 10.1: Growth of different scales, entering the horizon during radiation domination (blue, upper) and matter domination (red, lower).

Scales larger than the Hubble scale at equality ($k < k_{\rm eq}$) entered the apparent horizon during matter domination. They have always been growing, by a^2 during RD, and by a during MD, independent of their scale. Their growth from t_i well before equality to some

late time t_0 well after equality is

$$\begin{aligned}\delta(t_0) &= \frac{D(t_{\rm eq})}{D(t_i)}\frac{D(t_0)}{D(t_{\rm eq})} \times \delta(t_i) \\ &= \left[\frac{a_{\rm eq}^2}{a_i^2}\frac{a_0}{a_{\rm eq}}\right] \times \delta(t_i) \qquad k < k_{\rm eq}\end{aligned} \tag{10.56}$$

The factors involving D or a on the right hand side are independent of position and hence these scales will preserve their initial power spectrum. In the language of our transfer function,

$$T(k) = \text{const} \qquad k < k_{\rm eq}\ . \tag{10.57}$$

Scales smaller than the Hubble scale at equality ($k > k_{\rm eq}$), however, entered the apparent horizon during radiation domination. Hence they experienced a deficit of growth while they were inside the horizon during radiation domination. A given comoving scale $k = a_{\rm enter}H_{\rm enter}$ will not grow between $t_{\rm enter}$ and $t_{\rm eq}$ and hence

$$\begin{aligned}\delta(t_0) &= \frac{D(t_{\rm enter})}{D(t_i)}\frac{D(t_0)}{D(t_{\rm eq})} \times \delta(t_i) \\ &= \frac{a_{\rm enter}^2}{a_i^2}\frac{a_0}{a_{\rm eq}} \times \delta(t_i) \\ &= \frac{a_{\rm enter}^2}{a_{\rm eq}^2} \times \left[\frac{a_{\rm eq}^2}{a_i^2}\frac{a_0}{a_{\rm eq}}\right] \times \delta(t_i) \qquad k > k_{\rm eq}\ ,\end{aligned} \tag{10.58}$$

where in the last equality we have pulled out the same constant factors as occur for those scales entering during MD (Eq. 10.56). During RD, $a \propto t^{1/2}$ At this time, $k = (aH)_{\rm enter} \propto (t^{1/2}t^{-1})_{\rm enter} \propto 1/a_{\rm enter}$. Hence scales entering during RD experience a relative deficit of growth by the factor

$$\begin{aligned}T(k) &\propto \frac{a_{\rm enter}^2}{a_{\rm eq}^2} \\ &\propto \frac{k_{\rm eq}^2}{k^2} \propto k^{-2} \qquad k > k_{\rm eq}\ .\end{aligned} \tag{10.59}$$

Combining our two cases,

$$T^2(k) = \text{const} \times \begin{cases} 1, & \text{if } k < k_{\rm eq} \\ k^{-4} & \text{if } k > k_{\rm eq}\ . \end{cases} \tag{10.60}$$

Traditionally, the time evolution is handled separately, so we set the initial constant to be one. Thus, an initial spectrum with $P(k, t_i) = Ak^{n_s}$ evolves to $P(k) = Ak^{n_s}$ on large scales and $P(k) = Ak^{n_s-4}$ on small scales. (In practice the turnover is not instantaneous, so the spectrum is smooth near $k = k_{\rm eq}$.) We show how this looks in Figure 10.2 for $n_s = 1$. (For this specific case, this can also be derived by using the fact that the horizon-scale perturbation δ_H, defined in Eq. 10.52 is constant in time, along with the sub-horizon growth functions of Table 10.4.2.)

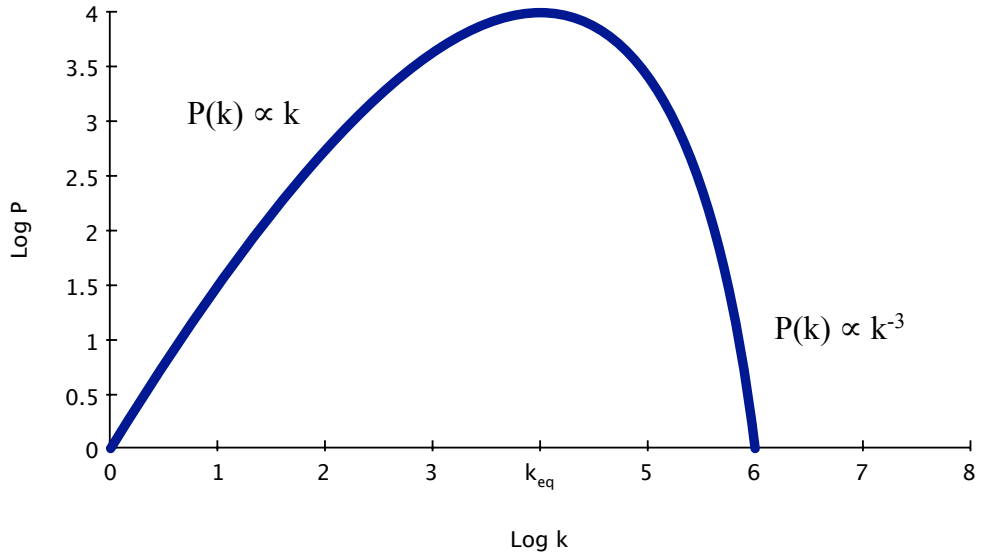

Figure 10.2: The processed power spectrum of density perturbations, $P(k)$, for an initial spectrum $P(k) \propto k$.

The present-day power spectrum thus depends upon various cosmological parameters. First, it depends on the value of $k_{\rm eq}$, and hence the ratio of the matter and radiation densities in the Universe. It obviously depends on the index n_s, and hence on some aspects of inflation. As we will see soon, it also depends on other cosmological parameters through baryonic features. We see a relatively recent compilation of measurements in Figure 10.3.

Hot Dark Matter

So far, we have assumed that the matter component in the Universe is dominated by cold dark matter, with a small admixture of baryons which will be discussed in the following section. This pressureless cold dark matter is moving very slowly and is only affected by gravity. However, another (conceptually) possible form of dark matter is so-called hot dark matter, light particles which are moving at velocities comparable to the speed of light at early times. One possible realization of this would be a light neutrino species, produced in sufficient abundance to have $\Omega_\nu \approx 0.3$. These large speeds are higher than the "escape velocities" of density perturbations, which serves to smooth out small perturbations as they are forming. This would have the effect of a sharp cutoff in the formation of structure — a cutoff that would prevent the formation of the structure that we see in the Universe today. Hence, light neutrinos are no longer considered a viable model for the dark matter. More precisely, they cannot be *all* of the dark matter, although current data allow a small admixture of hot dark matter.

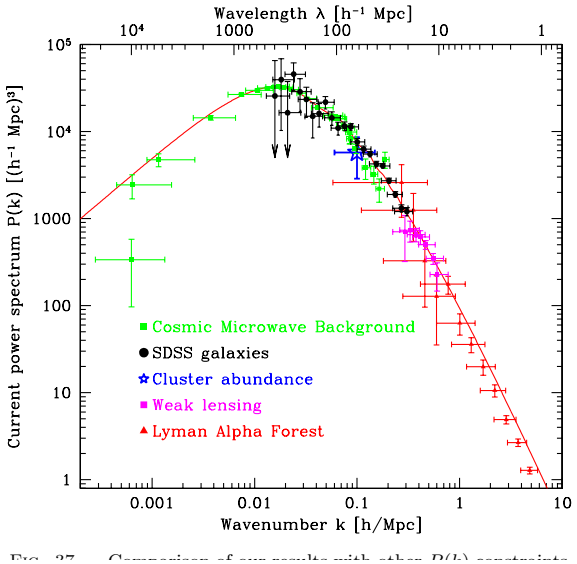

Figure 10.3: The observed power spectrum of density perturbations, as measured from a variety of techniques, from Tegmark et al 2003.

10.4.3 The effect of baryons: BAOs and the CMB

Baryon Acoustic Oscillations

There is one more important complication to consider: the presence of baryons as a separate component of matter. At early times, the baryons have pressure not because of their intrinsic equation of state, but because they are tightly coupled to the radiation. Hence, the "baryon-photon fluid" has an equation of state reduced slightly from $c_s^2 \simeq 1/3$ by the presence of the baryons, until decoupling at $z \simeq 1,100$. Even though this is already the matter-dominated era, perturbations in the baryons do not grow, even though perturbations in the pressureless dark matter are able to collapse. It is exactly the pattern of these baryon-photon sound waves that we see in both the Cosmic Microwave Background and, at a weaker level, in the matter power spectrum as measured by the distribution of galaxies. Here, they are known as *Baryon Acoustic Oscillations* (BAOs).

In fact, the pattern generated by these sound waves is very specific, and they therefore have a characteristic scale of the maximum distance that a sound wave could have travelled before the baryons and photons decouple at $t \simeq 370,000$ or $z \simeq 1,100$, corresponding to a scale of roughly 100 Mpc in the Universe today. Indeed, this is the characteristic scale of spots in the CMB, and more recently we have begun to observe that there are more galaxies than expected at this separation, a phenomenon known as *baryon acoustic oscillations*. Moreover, because this scale is fixed in comoving coordinates by our CMB observations, it can be used as a standard ruler to measure the cosmological parameters.

The presence of baryons thus modifies our transfer function: in real space, it would induce a peak at a separation at approximately 100 Mpc; in Fourier space this shows up as "bumps and wiggles" in $T(k)$ at that scale (roughly speaking, this is because the

Fourier transform of a spike is a sinusoid). These have indeed been observed by the 2DF and Sloan surveys of galaxies on large scales (Figure 10.4), and are beginning to allow to us to determine cosmological parameters from measurements of the galaxy power spectrum.

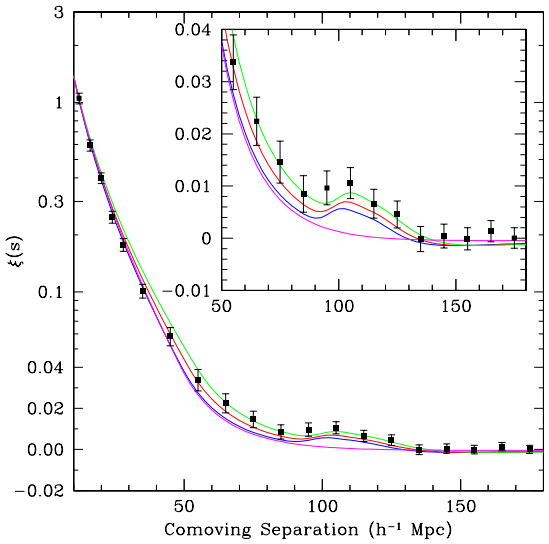

Figure 10.4: Baryon Acoustic Oscillations as measured in the galaxy correlation function by the SDSS Collaboration, Eisenstein et al 2005. The inset is a zoom with expanded axes, and the curves are various models.)

Fluctuations in the CMB

There is similar behavior expected and observed in the CMB. Roughly speaking, the temperature perturbation at a particular position on the sky is given by

$$\frac{\delta T}{T}(\hat{\mathbf{x}}) = \frac{1}{4}\frac{\delta\rho_\gamma}{\rho_\gamma} + \mathbf{v}\cdot\hat{\mathbf{x}} + \int dt a^{-1} h_{ij}\hat{\mathbf{x}}_i\hat{\mathbf{x}}_j \; . \tag{10.61}$$

The first term just comes from the energy density of a black body, $\rho_\gamma \propto T^4$, the second from the relative Doppler shift between an observer and a parcel of matter moving with velocity $\mathbf{v}$ and the final term is due to photons falling into and out of potential wells, in relativistic notation as the perturbation to the metric, h_{ij}. All three of $\delta\rho_\gamma$, $\mathbf{v}$ and h_{ij} can be thought of as small perturbations and related to the perturbation calculations of this chapter, and hence the CMB temperature fluctuation can also be expressed as a linear functional of the initial density perturbations. Because the CMB is represented on the (two-dimensional and spherical) sky, we cannot use Fourier transforms, but rather the spherical harmonic transform:

$$\frac{\delta T}{T}(\hat{\mathbf{x}}) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{+\ell} a_{\ell m} Y_{\ell m}(\hat{\mathbf{x}}) \tag{10.62}$$

and its inverse

$$a_{\ell m} = \int d^2\hat{\mathbf{x}}\, \frac{\delta T}{T}(\hat{\mathbf{x}}) Y^*_{\ell m}(\hat{\mathbf{x}}) \tag{10.63}$$

For a statistically isotropic sky, the correlation function is only a function of the angular distance between points:

$$\left\langle \frac{\delta T}{T}(\hat{\mathbf{x}}) \frac{\delta T}{T}(\hat{\mathbf{y}}) \right\rangle = C(\theta) \qquad \hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = \cos\theta \tag{10.64}$$

which also means that we can define the power spectrum

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta_{\ell,\ell'} \delta_{m,m'} \tag{10.65}$$

where the right hand side has Kronecker δs. (Compare the related quantities for power pectra in Sec. 10.1 above.) The m index acts analogously to the angle of the wavevector in the 3-d case, and the spectrum and correlation function are related by

$$C(\theta) = \sum_\ell \frac{2\ell + 1}{4\pi} C_\ell P_\ell(\cos\theta) \tag{10.66}$$

where the P_ℓ are Legendre polynomials. Because of linearity, we can in principle express the CMB power spectrum as a functional of the initial power spectrum of density perturbations:

$$C_\ell = \int dk T_\ell(k) P(k, t_i) \ , \tag{10.67}$$

where the details of the transfer function depend upon the cosmological parameters. Just as in the BAO case, the CMB power spectrum lets us determine the cosmological parameters.

Boltzmann Solvers

In practice, the full description of the CMB and matter power spectrum in the presence of dark matter, baryons and radiation (as well as neutrinos and possible isocurvature fluctuations as described earlier) requires the solution of the Boltzmann equation, originally discussed in Chapter 5, but now generalized to allow for spatially-varying perturbations to the distribution functions and the full treatment of general relativity. This cannot be done analytically, but instead some powerful numerical codes have been developed in the cosmology community. The two most well-known and often-used are CMBFAST (`http://cfa-www.harvard.edu/~mzaldarr/CMBFAST/cmbfast.html`) and CAMB (`http://camb.info`).

In Figure 10.5 we show the CMB power spectrum for several different sets of cosmological parameters as calculated by the CMBFAST program. In Figure 10.6 we show recent measurements by the WMAP satellite and other experiments, as compiled by the WMAP team (for more and more recent data, see `http://lambda.gsfc.gov`).

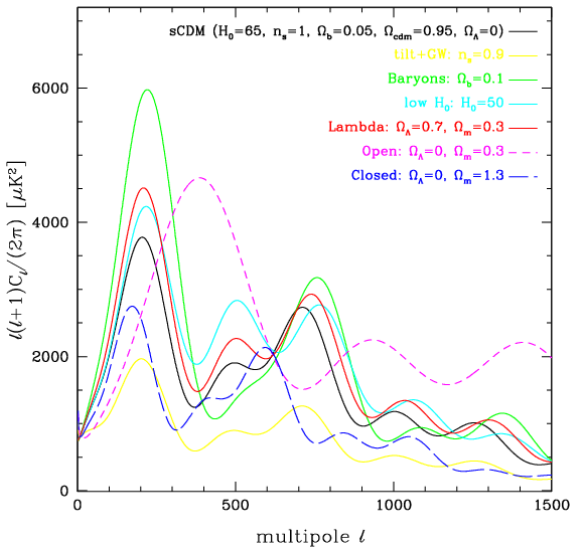

Figure 10.5: The CMB power spectrum for various sets of cosmological parameters.

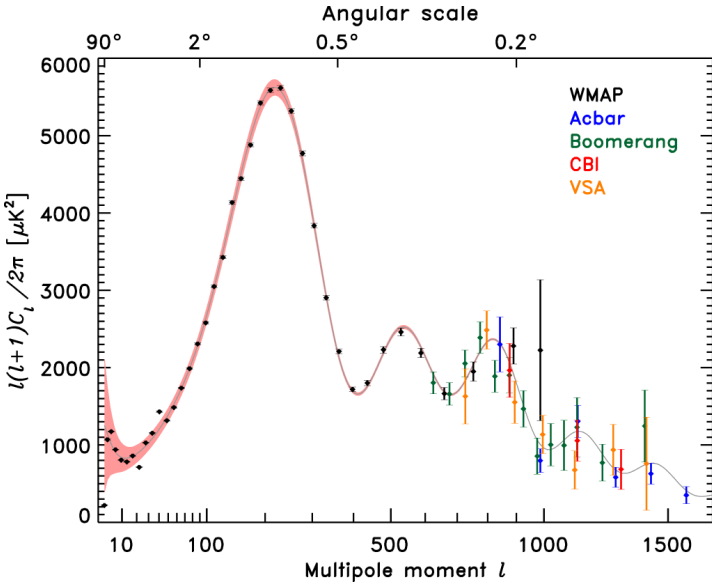

Figure 10.6: The measured CMB power spectrum, as compiled by the WMAP team.

Errata

The following is a list of errata and other changes made between the handouts and the version currently on the web. The version currently available has these problems corrected, except as noted!

Notes

1. In Eq. 2.59 there is a factor of c missing (which doesn't affect the argument at all).
2. In Chapters 3-4, there is some sloppiness in the signs of dr vs dt – this is a choice which depends upon whether we decide to have r increase away from the observer (backwards in time) or away from the emitter (forward in time). (*Not yet fixed in the main text.*)
3. Integral sign missing in Eq. 4.18
4. In Section 8.2.3 on relic particles, the third paragraph should say "$H_{\rm GUT} \sim T_{\rm GUT}^2/m_{\rm Pl}$ appropriate for an RD universe." and "This density would have been diluted by a factor of $(1+z_{\rm GUT})^{-3} = T_0^3/T_{\rm GUT}^3$" (several of the exponents were incorrect in these expressions).
5. Section 10.4.2, Eqs. 10.56–10.59 and surrounding text had the inequalities reversed: smaller k is larger scale, and vice-versa. Eq. 10.60 summarizing the results, was correct.

Problems

1. PS3, Q3. Derivatives $d_L'(z)$ incorrectly calculated, so the final results are incorrect for both d_A and d_L. (*Not yet fixed online.*)