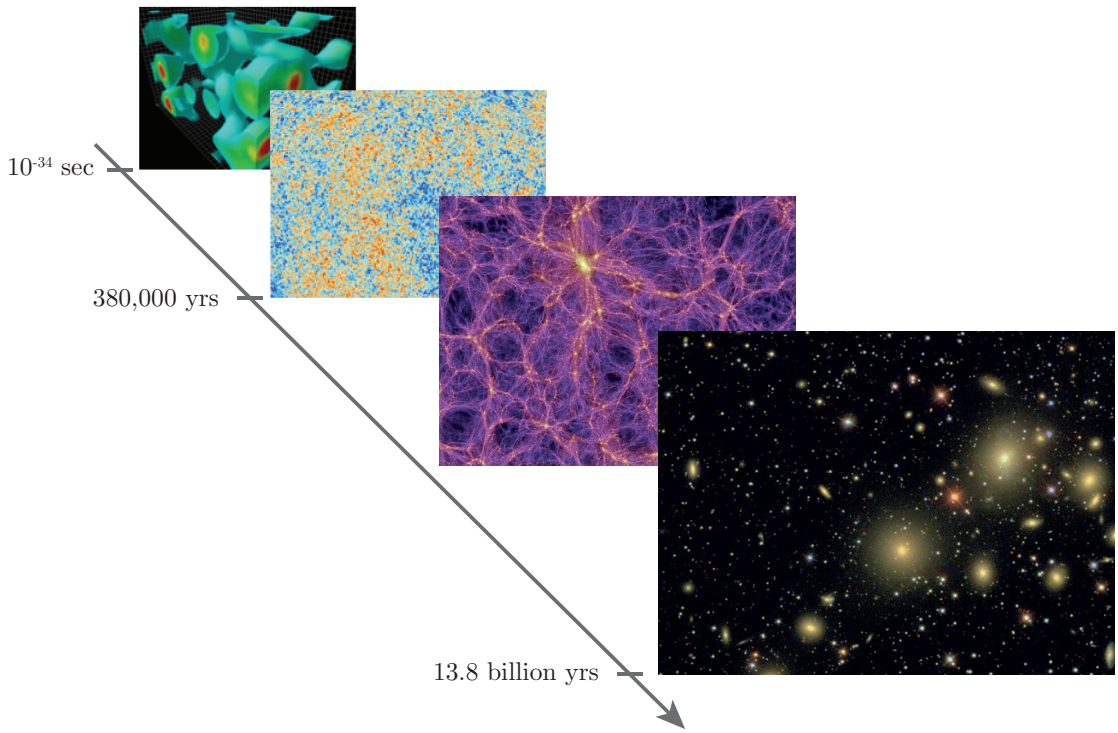
Cosmology

Part III Mathematical Tripos



Daniel Baumann

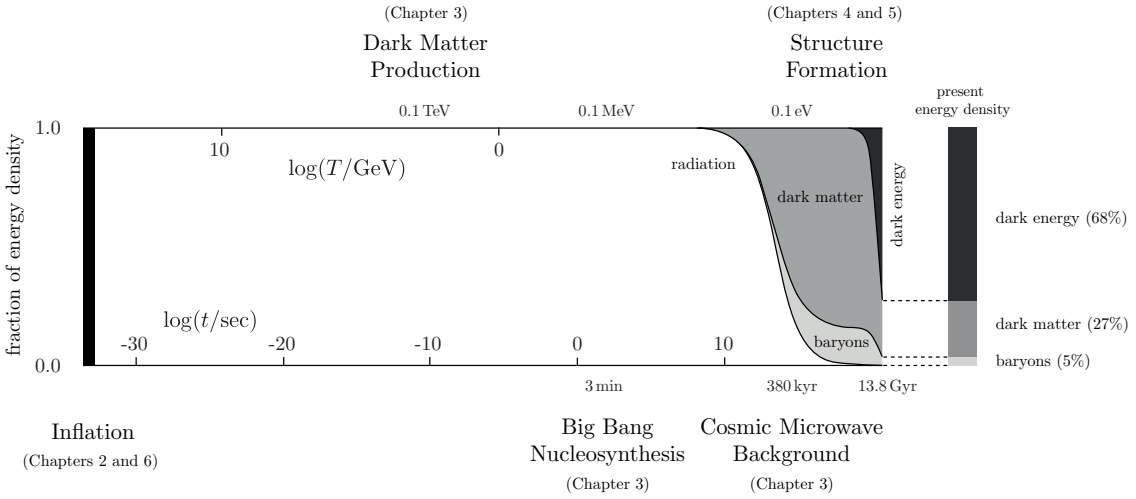
dbaumann@damtp.cam.ac.uk

Contents

Preface

This course is about 13.8 billion years of cosmic evolution:

At early times, the universe was hot and dense. Interactions between particles were frequent and energetic. Matter was in the form of free electrons and atomic nuclei with light bouncing between them. As the primordial plasma cooled, the light elements—hydrogen, helium and lithium—formed. At some point, the energy had dropped enough for the first stable atoms to exist. At that moment, photons started to stream freely. Today, billions of years later, we observe this afterglow of the Big Bang as microwave radiation. This radiation is found to be almost completely uniform, the same temperature (about 2.7 K) in all directions. Crucially, the cosmic microwave background contains small variations in temperature at a level of 1 part in 10 000. Parts of the sky are slightly hotter, parts slightly colder. These fluctuations reflect tiny variations in the primordial density of matter. Over time, and under the influence of gravity, these matter fluctuations grew. Dense regions were getting denser. Eventually, galaxies, stars and planets formed.



This picture of the universe—from fractions of a second after the Big Bang until today—is a scientific fact. However, the story isn't without surprises. The majority of the universe today consists of forms of matter and energy that are unlike anything we have ever seen in terrestrial experiments. Dark matter is required to explain the stability of galaxies and the rate of formation of large-scale structures. Dark energy is required to rationalise the striking fact that the expansion of the universe started to accelerate recently (meaning a few billion years ago). What dark matter and dark energy are is still a mystery. Finally, there is growing evidence that the primordial density perturbations originated from microscopic quantum fluctuations, stretched to cosmic sizes during a period of inflationary expansion. The physical origin of inflation is still a topic of active research.

Administrative comments.—Up-to-date versions of the lecture notes will be posted on the course website:

www.damtp.cam.ac.uk/user/db275/cosmology.pdf

Starred sections (*) are non-examinable.

> Boxed text contains technical details and derivations that may be omitted on a first reading.

Please do not hesitate to email me questions, comments or corrections:

dbaumann@damtp.cam.ac.uk

There will be four problem sets, which will appear in two-week intervals on the course website. Details regarding supervisions will be announced in the lectures.

Notation and conventions.—We will mostly use natural units, in which the speed of light and Planck's constant are set equal to one, $c = \hbar \equiv 1$. Length and time then have the same units. Our metric signature is $(+---)$, so that $\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}\boldsymbol{x}^2$ for Minkowski space. This is the same signature as used in the QFT course, but the opposite of the GR course. Spacetime four-vectors will be denoted by capital letters, e.g. X^μ and P^μ, where the Greek indices $\mu, \nu, \cdots$ run from 0 to 3. We will use the Einstein summation convention where repeated indices are summed over. Latin indices $i, j, k, \cdots$ will stand for spatial indices, e.g. x^i and p^i. Bold font will denote spatial three-vectors, e.g. $\boldsymbol{x}$ and $\boldsymbol{p}$.

Further reading.—I recommend the following textbooks:

- ▷ Dodelson, *Modern Cosmology*
  A very readable book at about the same level as these lectures. My Boltzmann-centric treatment of BBN and recombination was heavily inspired by Dodelson's Chapter 3.

- ▷ Peter and Uzan, *Primordial Cosmology*
  A recent book that contains a lot of useful reference material. Also good for *Advanced Cosmology*.

- ▷ Kolb and Turner, *The Early Universe*
  A remarkably timeless book. It is still one of the best treatments of the thermal history of the early universe.

- ▷ Weinberg, *Cosmology*
  Written by the hero of a whole generation of theoretical physicists, this is the text to consult if you are ever concerned about a lack of rigour. Unfortunately, Weinberg doesn't do plots.

Acknowledgements.—Thanks to Paolo Creminelli for comments on a previous version of these notes. Adam Solomon was a fantastic help in designing the problem sets and writing some of the solutions.

Part I

The Homogeneous Universe

1 Geometry and Dynamics

The further out we look into the universe, the simpler it seems to get (see fig. 1.1). Averaged over large scales, the clumpy distribution of galaxies becomes more and more *isotropic*—i.e. independent of direction. Despite what your mom might have told you, we shouldn't assume that we are the centre of the universe. (This assumption is sometimes called the *cosmological principle*). The universe should then appear isotropic to any (free-falling) observer throughout the universe. If the universe is isotropic around all points, then it is also *homogeneous*—i.e. independent of position. To a first approximation, we will therefore treat the universe as perfectly homogeneous and isotropic. As we will see, in §1.1, homogeneity and isotropy single out a unique form of the spacetime geometry. We discuss how particles and light propagate in this spacetime in §1.2. Finally, in §1.3, we derive the Einstein equations and relate the rate of expansion of the universe to its matter content.

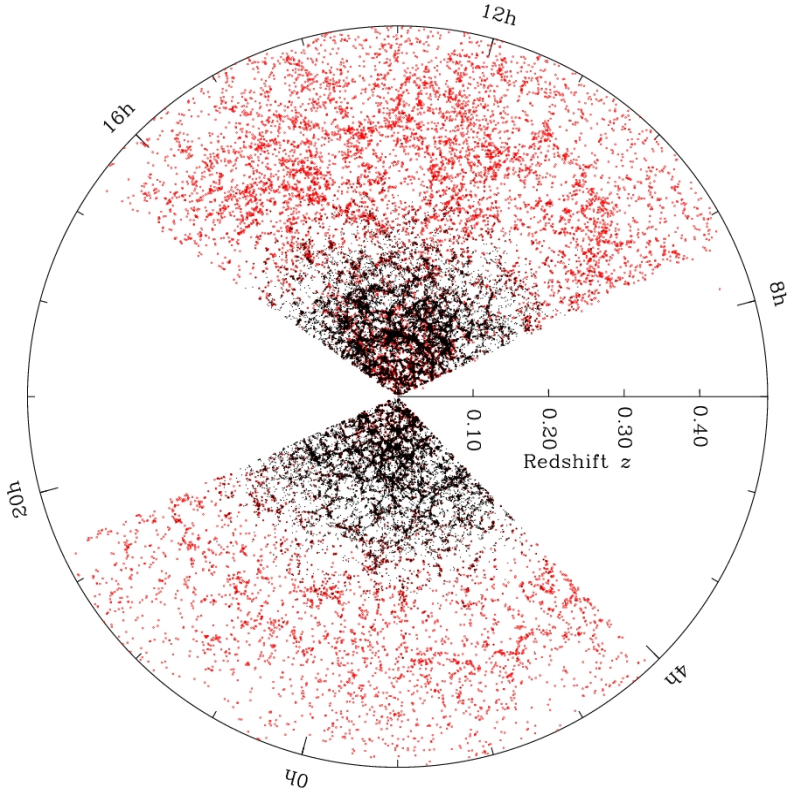


Figure 1.1: *The distribution of galaxies is clumpy on small scales, but becomes more uniform on large scales and early times.*

1.1 Geometry

1.1.1 Metric

The spacetime *metric* plays a fundamental role in relativity. It turns observer-dependent coordinates $X^\mu = (t, x^i)$ into the invariant line element[1]

$$\mathrm{d}s^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu}\mathrm{d}X^\mu \mathrm{d}X^\nu \equiv g_{\mu\nu}\mathrm{d}X^\mu \mathrm{d}X^\nu \ . \tag{1.1.1}$$

In special relativity, the Minkowski metric is the same everywhere in space and time,

$$g_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1) \ . \tag{1.1.2}$$

In general relativity, on the other hand, the metric will depend on where we are and when we are,

$$g_{\mu\nu}(t, \boldsymbol{x}) \ . \tag{1.1.3}$$

The spacetime dependence of the metric incorporates the effects of gravity. How the metric depends on the position in spacetime is determined by the distribution of matter and energy in the universe. For an arbitrary matter distribution, it can be next to impossible to find the metric from the Einstein equations. Fortunately, the large degree of symmetry of the homogeneous universe simplifies the problem.

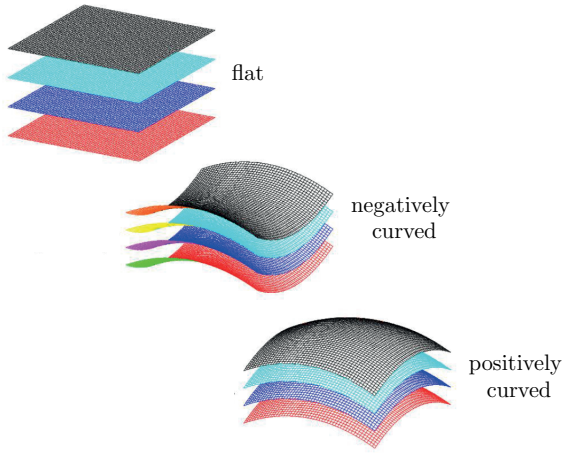


Figure 1.2: *The spacetime of the universe can be foliated into flat, positively curved or negatively curved spatial hypersurfaces.*

1.1.2 Symmetric Three-Spaces

Spatial homogeneity and isotropy mean that the universe can be represented by a time-ordered sequence of three-dimensional spatial slices Σ_t, each of which is homogeneous and isotropic (see fig. 1.2). We start with a classification of such maximally symmetric 3-spaces. First, we note that homogeneous and isotropic 3-spaces have constant 3-curvature.[2] There are only three options:

[1] Throughout the course, will use the Einstein summation convention where repeated indices are summed over. We will also use natural units with $c \equiv 1$, so that $\mathrm{d}X^0 = \mathrm{d}t$. Our metric signature will be mostly minus, $(+, -, -, -)$.

[2] We give a precise definition of Riemann curvature below.

zero curvature, positive curvature and negative curvature. Let us determine the metric for each case:

- *flat space*: the line element of three-dimensional Euclidean space E^3 is simply

$$\mathrm{d}\ell^2 = \mathrm{d}\boldsymbol{x}^2 = \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \ . \tag{1.1.4}$$

  This is clearly invariant under spatial translations ($x^i \mapsto x^i + a^i$, with $a^i = const.$) and rotations ($x^i \mapsto R^i{}_k\, x^k$, with $\delta_{ij}R^i{}_kR^j{}_l = \delta_{kl}$).

- *positively curved space*: a 3-space with constant positive curvature can be represented as a *3-sphere* S^3 embedded in four-dimensional Euclidean space E^4,

$$\mathrm{d}\ell^2 = \mathrm{d}\boldsymbol{x}^2 + \mathrm{d}u^2 \ , \qquad \boldsymbol{x}^2 + u^2 = a^2 \ , \tag{1.1.5}$$

  where a is the radius of the 3-sphere. Homogeneity and isotropy of the surface of the 3-sphere are inherited from the symmetry of the line element under four-dimensional rotations.

- *negatively curved space*: a 3-space with constant negative curvature can be represented as a *hyperboloid* H^3 embedded in four-dimensional Lorentzian space $\mathbb{R}^{1,3}$,

$$\mathrm{d}\ell^2 = \mathrm{d}\boldsymbol{x}^2 - \mathrm{d}u^2 \ , \qquad \boldsymbol{x}^2 - u^2 = -a^2 \ , \tag{1.1.6}$$

  where a^2 is an arbitrary constant. Homogeneity and isotropy of the induced geometry on the hyperboloid are inherited from the symmetry of the line element under four-dimensional pseudo-rotations (i.e. Lorentz transformations, with u playing the role of time).

In the last two cases, it is convenient to rescale the coordinates, $\boldsymbol{x} \to a\boldsymbol{x}$ and $u \to au$. The line elements of the spherical and hyperbolic cases then are

$$\mathrm{d}\ell^2 = a^2\left[\mathrm{d}\boldsymbol{x}^2 \pm \mathrm{d}u^2\right] \ , \qquad \boldsymbol{x}^2 \pm u^2 = \pm 1 \ . \tag{1.1.7}$$

Notice that the coordinates $\boldsymbol{x}$ and u are now dimensionless, while the parameter a carries the dimension of length. The differential of the embedding condition, $\boldsymbol{x}^2 \pm u^2 = \pm 1$, gives $u\,\mathrm{d}u = \mp\boldsymbol{x}\cdot\mathrm{d}\boldsymbol{x}$, so

$$\mathrm{d}\ell^2 = a^2\left[\mathrm{d}\boldsymbol{x}^2 \pm \frac{(\boldsymbol{x}\cdot\mathrm{d}\boldsymbol{x})^2}{1 \mp \boldsymbol{x}^2}\right] \ . \tag{1.1.8}$$

We can unify (1.1.8) with the Euclidean line element (1.1.4) by writing

$$\mathrm{d}\ell^2 = a^2\left[\mathrm{d}\boldsymbol{x}^2 + k\,\frac{(\boldsymbol{x}\cdot\mathrm{d}\boldsymbol{x})^2}{1 - k\boldsymbol{x}^2}\right] \equiv a^2\gamma_{ij}\,\mathrm{d}x^i\mathrm{d}x^j \ , \tag{1.1.9}$$

with

$$\gamma_{ij} \equiv \delta_{ij} + k\,\frac{x_ix_j}{1 - k(x_kx^k)} \ , \qquad \text{for} \qquad k \equiv \begin{cases} 0 & \text{Euclidean} \\ +1 & \text{spherical} \\ -1 & \text{hyperbolic} \end{cases} \ . \tag{1.1.10}$$

Note that we must take $a^2 > 0$ in order to have $\mathrm{d}\ell^2$ positive at $\boldsymbol{x} = 0$, and hence everywhere.[3] The form of the spatial metric γ_{ij} depends on the choice of coordinates:

[3] Notice that despite appearance $\boldsymbol{x} = 0$ is not a special point.

- It is convenient to use spherical polar coordinates, (r, θ, ϕ), because it makes the symmetries of the space manifest. Using

$$\mathrm{d}\boldsymbol{x}^2 = \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2) \ , \tag{1.1.11}$$

$$\boldsymbol{x}\cdot\mathrm{d}\boldsymbol{x} = r\,\mathrm{d}r \ , \tag{1.1.12}$$

  the metric in (1.1.9) becomes diagonal

$$\mathrm{d}\ell^2 = a^2\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right] \ , \tag{1.1.13}$$

  where $\mathrm{d}\Omega^2 \equiv \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2$.

- The complicated γ_{rr} component of (1.1.13) can sometimes be inconvenient. In that case, we may redefine the radial coordinate, $\mathrm{d}\chi \equiv \mathrm{d}r/\sqrt{1-kr^2}$, such that

$$\mathrm{d}\ell^2 = a^2\Big[\mathrm{d}\chi^2 + S_k^2(\chi)\,\mathrm{d}\Omega^2\Big] \ , \tag{1.1.14}$$

  where

$$S_k(\chi) \equiv \begin{cases} \sinh\chi & k=-1 \\ \chi & k=0 \\ \sin\chi & k=+1 \end{cases} \ . \tag{1.1.15}$$

1.1.3 Robertson-Walker Metric

To get the *Robertson-Walker metric*[4] for an expanding universe, we simply include $\mathrm{d}\ell^2 = a^2\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j$ into the spacetime line element and let the parameter a be an arbitrary function of time[5]

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j \ . \tag{1.1.16}$$

Notice that the symmetries of the universe have reduced the ten independent components of the spacetime metric to a single function of time, the *scale factor* $a(t)$, and a constant, the curvature parameter k. The coordinates $x^i \equiv \{x^1, x^2, x^3\}$ are called *comoving coordinates*. Fig. 1.3 illustrates the relation between comoving coordinates and *physical coordinates*, $x^i_{\rm phys} = a(t)\,x^i$. The physical velocity of an object is

$$v^i_{\rm phys} \equiv \frac{dx^i_{\rm phys}}{dt} = a(t)\frac{dx^i}{dt} + \frac{da}{dt}x^i \equiv v^i_{\rm pec} + Hx^i_{\rm phys} \ . \tag{1.1.17}$$

We see that this has two contributions: the so-called *peculiar velocity*, $v^i_{\rm pec} \equiv a(t)\dot{x}^i$, and the *Hubble flow*, $Hx^i_{\rm phys}$, where we have defined the *Hubble parameter* as[6]

$$H \equiv \frac{\dot{a}}{a} \ . \tag{1.1.18}$$

The peculiar velocity of an object is the velocity measured by a comoving observer (i.e. an observer who follows the Hubble flow).

[4] Sometimes this is called the Friedmann-Robertson-Walker (FRW) metric.

[5] Skeptics might worry about uniqueness. Why didn't we include a g_{0i} component? Because it would break isotropy. Why don't we allow for a non-trivial g_{00} component? Because it can always be absorbed into a redefinition of the time coordinate, $\mathrm{d}t' \equiv \sqrt{g_{00}}\,\mathrm{d}t$.

[6] Here, and in the following, an overdot denotes a time derivative, i.e. $\dot{a} \equiv da/dt$.

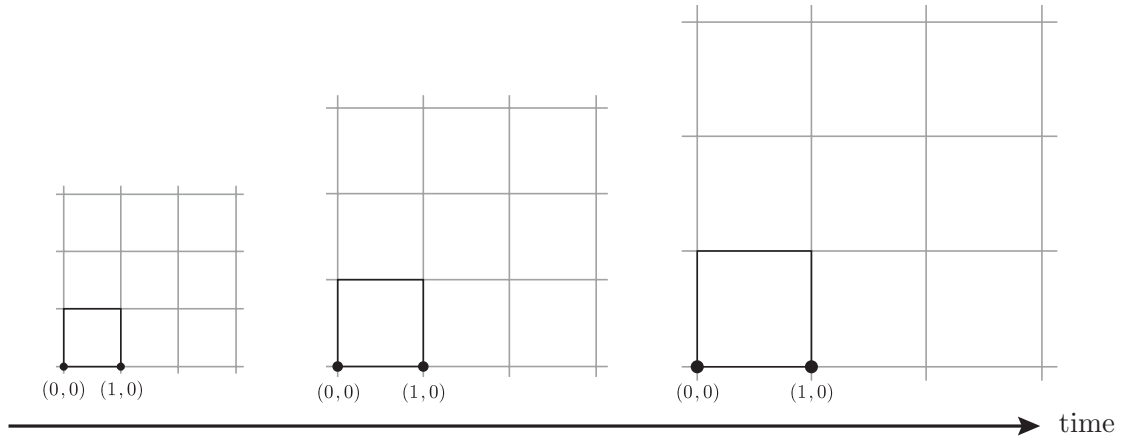


Figure 1.3: *Expansion of the universe. The comoving distance between points on an imaginary coordinate grid remains constant as the universe expands. The physical distance is proportional to the comoving distance times the scale factor $a(t)$ and hence gets larger as time evolves.*

- Using (1.1.13), the FRW metric in polar coordinates reads

$$\boxed{\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right]} \,. \tag{1.1.19}$$

  This result is worth memorizing — after all, it is the metric of our universe! Notice that the line element (1.1.19) has a rescaling symmetry

$$a \to \lambda a \,, \quad r \to r/\lambda \,, \quad k \to \lambda^2 k \,. \tag{1.1.20}$$

  This means that the geometry of the spacetime stays the same if we simultaneously rescale a, r and k as in (1.1.20). We can use this freedom to set the scale factor to unity today:[7] $a_0 \equiv a(t_0) \equiv 1$. In this case, $a(t)$ becomes dimensionless, and r and $k^{-1/2}$ inherit the dimension of length.

- Using (1.1.14), we can write the FRW metric as

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\Big[\mathrm{d}\chi^2 + S_k^2(\chi)\mathrm{d}\Omega^2\Big] \,. \tag{1.1.21}$$

  This form of the metric is particularly convenient for studying the propagation of light. For the same purpose, it is also useful to introduce *conformal time*,

$$\mathrm{d}\tau = \frac{\mathrm{d}t}{a(t)} \,, \tag{1.1.22}$$

  so that (1.1.21) becomes

$$\mathrm{d}s^2 = a^2(\tau)\Big[\mathrm{d}\tau^2 - \big(\mathrm{d}\chi^2 + S_k^2(\chi)\mathrm{d}\Omega^2\big)\Big] \,. \tag{1.1.23}$$

  We see that the metric has factorized into a static Minkowski metric multiplied by a time-dependent conformal factor $a(\tau)$. Since light travels along null geodesics, $\mathrm{d}s^2 = 0$, the propagation of light in FRW is the same as in Minkowski space if we first transform to conformal time. Along the path, the change in conformal time equals the change in comoving distance,

$$\Delta\tau = \Delta\chi \,. \tag{1.1.24}$$

  We will return to this in Chapter 2.

[7] Quantities that are evaluated at the present time t_0 will have a subscript '0'.

1.2 Kinematics

1.2.1 Geodesics

How do particles evolve in the FRW spacetime? In the absence of additional non-gravitational forces, freely-falling particles in a curved spacetime move along geodesics. I will briefly remind you of some basic facts about geodesic motion in general relativity[8] and then apply it to the FRW spacetime (1.1.16).

Geodesic Equation*

Consider a particle of mass m. In a curved spacetime it traces out a path $X^\mu(s)$. The *four-velocity* of the particle is defined by

$$U^\mu \equiv \frac{dX^\mu}{ds} \ . \tag{1.2.25}$$

A *geodesic* is a curve which extremises the proper time $\Delta s/c$ between two points in spacetime. In the box below, I show that this extremal path satisfies the *geodesic equation*[9]

$$\boxed{\frac{dU^\mu}{ds} + \Gamma^\mu_{\alpha\beta} U^\alpha U^\beta = 0} \ , \tag{1.2.26}$$

where $\Gamma^\mu_{\alpha\beta}$ are the *Christoffel symbols*,

$$\boxed{\Gamma^\mu_{\alpha\beta} \equiv \frac{1}{2} g^{\mu\lambda}(\partial_\alpha g_{\beta\lambda} + \partial_\beta g_{\alpha\lambda} - \partial_\lambda g_{\alpha\beta})} \ . \tag{1.2.27}$$

Here, I have introduced the notation $\partial_\mu \equiv \partial/\partial X^\mu$. Moreover, you should recall that the inverse metric is defined through $g^{\mu\lambda} g_{\lambda\nu} = \delta^\mu_\nu$.

*Derivation of the geodesic equation.**—Consider the motion of a massive particle between to points in spacetime A and B (see fig. 1.4). The relativistic action of the particle is

$$S = -m \int_A^B \mathrm{d}s \ . \tag{1.2.28}$$

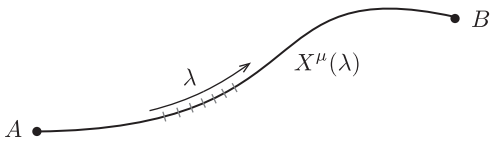


Figure 1.4: *Parameterisation of an arbitrary path in spacetime, $X^\mu(\lambda)$.*

We label each point on the curve by a parameter λ that increases monotonically from an initial value $\lambda(A) \equiv 0$ to a final value $\lambda(B) \equiv 1$. The action is a functional of the path $X^\mu(\lambda)$,

$$S[X^\mu(\lambda)] \ = \ -m \int_0^1 \left(g_{\mu\nu}(X) \dot{X}^\mu \dot{X}^\nu \right)^{1/2} \mathrm{d}\lambda \ \equiv \ \int_0^1 L[X^\mu, \dot{X}^\mu] \, \mathrm{d}\lambda \ , \tag{1.2.29}$$

[8] If all of this is new to you, you should arrange a crash-course with me and/or read Sean Carroll's *No-Nonsense Introduction to General Relativity*.

[9] If you want to learn about the beautiful geometrical story behind geodesic motion I recommend Harvey Reall's *Part III General Relativity* lectures. Here, I simply ask you to accept the geodesic equation as the $F = ma$ of general relativity (for $F = 0$). From now on, we will use (1.2.26) as our starting point.

where $\dot{X}^\mu \equiv dX^\mu/d\lambda$. The motion of the particle corresponds to the extremum of this action. The integrand in (1.2.29) is the Lagrangian L and it satisfies the Euler-Lagrange equation

$$\frac{d}{d\lambda}\left(\frac{\partial L}{\partial \dot{X}^\mu}\right) - \frac{\partial L}{\partial X^\mu} = 0 \ . \tag{1.2.30}$$

The derivatives in (1.2.30) are

$$\frac{\partial L}{\partial \dot{X}^\mu} = -\frac{1}{L} g_{\mu\nu}\dot{X}^\nu \quad , \quad \frac{\partial L}{\partial X^\mu} = -\frac{1}{2L}\partial_\mu g_{\nu\rho}\dot{X}^\nu \dot{X}^\rho \ . \tag{1.2.31}$$

Before continuing, it is convenient to switch from the general parameterisation λ to the parameterisation using proper time s. (We could not have used s from the beginning since the value of s at B is different for different curves. The range of integration would then have been different for different curves.) Notice that

$$\left(\frac{ds}{d\lambda}\right)^2 = g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu = L^2 \ , \tag{1.2.32}$$

and hence $ds/d\lambda = L$. In the above equations, we can therefore replace $d/d\lambda$ with Ld/ds. The Euler-Lagrange equation then becomes

$$\frac{d}{ds}\left(g_{\mu\nu}\frac{dX^\nu}{ds}\right) - \frac{1}{2}\partial_\mu g_{\nu\rho}\frac{dX^\nu}{ds}\frac{dX^\rho}{ds} = 0 \ . \tag{1.2.33}$$

Expanding the first term, we get

$$g_{\mu\nu}\frac{d^2X^\nu}{ds^2} + \partial_\rho g_{\mu\nu}\frac{dX^\rho}{ds}\frac{dX^\nu}{ds} - \frac{1}{2}\partial_\mu g_{\nu\rho}\frac{dX^\nu}{ds}\frac{dX^\rho}{ds} = 0 \ . \tag{1.2.34}$$

In the second term, we can replace $\partial_\rho g_{\mu\nu}$ with $\frac{1}{2}(\partial_\rho g_{\mu\nu} + \partial_\nu g_{\mu\rho})$ because it is contracted with an object that is symmetric in ν and ρ. Contracting (1.2.34) with the inverse metric and relabelling indices, we find

$$\frac{d^2X^\mu}{ds^2} + \Gamma^\mu_{\alpha\beta}\frac{dX^\alpha}{ds}\frac{dX^\beta}{ds} = 0 \ . \tag{1.2.35}$$

Substituting (1.2.25) gives the desired result (1.2.26).

The derivative term in (1.2.26) can be manipulated by using the chain rule

$$\frac{d}{ds}U^\mu(X^\alpha(s)) = \frac{dX^\alpha}{ds}\frac{\partial U^\mu}{\partial X^\alpha} = U^\alpha\frac{\partial U^\mu}{\partial X^\alpha} \ , \tag{1.2.36}$$

so that we get

$$U^\alpha\left(\frac{\partial U^\mu}{\partial X^\alpha} + \Gamma^\mu_{\alpha\beta}U^\beta\right) = 0 \ . \tag{1.2.37}$$

The term in brackets is the *covariant derivative* of U^μ, i.e. $\nabla_\alpha U^\mu \equiv \partial_\alpha U^\mu + \Gamma^\mu_{\alpha\beta}U^\beta$. This allows us to write the geodesic equation in the following slick way: $U^\alpha\nabla_\alpha U^\mu = 0$. In the GR course you will derive this form of the geodesic equation directly by thinking about *parallel transport*.

Using the definition of the *four-momentum* of the particle,

$$P^\mu = mU^\mu \ , \tag{1.2.38}$$

we may also write (1.2.37) as

$$\boxed{P^\alpha\frac{\partial P^\mu}{\partial X^\alpha} = -\Gamma^\mu_{\alpha\beta}P^\alpha P^\beta} \ . \tag{1.2.39}$$

For massless particles, the action (1.2.29) vanishes identically and our derivation of the geodesic equation breaks down. We don't have time to go through the more subtle derivation of the geodesic equation for massless particles. Luckily, we don't have to because the result is exactly the same as (1.2.39).[10] We only need to interpret P^μ as the four-momentum of a massless particle.

Accepting that the geodesic equation (1.2.39) applies to both massive and massless particles, we will move on. I will now show you how to apply the geodesic equation to particles in the FRW universe.

Geodesic Motion in FRW

To evaluate the r.h.s. of (1.2.39) we need to compute the Christoffel symbols for the FRW metric (1.1.16),

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j \ . \tag{1.2.40}$$

All Christoffel symbols with two time indices vanish, i.e. $\Gamma^\mu_{00} = \Gamma^0_{0\beta} = 0$. The only non-zero components are

$$\boxed{\Gamma^0_{ij} = a\dot{a}\gamma_{ij} \ , \qquad \Gamma^i_{0j} = \frac{\dot{a}}{a}\delta^i_j \ , \qquad \Gamma^i_{jk} = \frac{1}{2}\gamma^{il}(\partial_j\gamma_{kl} + \partial_k\gamma_{jl} - \partial_l\gamma_{jk})} \ , \tag{1.2.41}$$

or are related to these by symmetry (note that $\Gamma^\mu_{\alpha\beta} = \Gamma^\mu_{\beta\alpha}$). I will derive Γ^0_{ij} as an example and leave Γ^i_{0j} as an exercise.

Example.—The Christoffel symbol with upper index equal to zero is

$$\Gamma^0_{\alpha\beta} = \frac{1}{2}g^{0\lambda}(\partial_\alpha g_{\beta\lambda} + \partial_\beta g_{\alpha\lambda} - \partial_\lambda g_{\alpha\beta}) \ . \tag{1.2.42}$$

The factor $g^{0\lambda}$ vanishes unless $\lambda = 0$ in which case it is equal to 1. Therefore,

$$\Gamma^0_{\alpha\beta} = \frac{1}{2}(\partial_\alpha g_{\beta 0} + \partial_\beta g_{\alpha 0} - \partial_0 g_{\alpha\beta}) \ . \tag{1.2.43}$$

The first two terms reduce to derivatives of g_{00} (since $g_{i0} = 0$). The FRW metric has constant g_{00}, so these terms vanish and we are left with

$$\Gamma^0_{\alpha\beta} = -\frac{1}{2}\partial_0 g_{\alpha\beta} \ . \tag{1.2.44}$$

The derivative is non-zero only if α and β are spatial indices, $g_{ij} = -a^2\gamma_{ij}$ (don't miss the sign!). In that case, we find

$$\Gamma^0_{ij} = a\dot{a}\gamma_{ij} \ . \tag{1.2.45}$$

The homogeneity of the FRW background implies $\partial_i P^\mu = 0$, so that the geodesic equation (1.2.39) reduces to

$$\begin{aligned} P^0\frac{dP^\mu}{dt} &= -\Gamma^\mu_{\alpha\beta}P^\alpha P^\beta \ , \\ &= -\left(2\Gamma^\mu_{0j}P^0 + \Gamma^\mu_{ij}P^i\right)P^j \ , \end{aligned} \tag{1.2.46}$$

[10] One way to think about massless particles is as the zero-mass *limit* of massive particles. A more rigorous derivation of null geodesics from an action principle can be found in Paul Townsend's *Part III Black Holes* lectures [arXiv:gr-qc/9707012].

where I have used (1.2.41) in the second line.

- The first thing to notice from (1.2.46) is that massive particles at rest in the comoving frame, $P^j = 0$, will stay at rest because the r.h.s. then vanishes,

$$P^j = 0 \quad \Rightarrow \quad \frac{dP^i}{dt} = 0 \ . \tag{1.2.47}$$

- Next, we consider the $\mu = 0$ component of (1.2.46), but don't require the particles to be at rest. The first term on the r.h.s. vanishes because $\Gamma^0_{0j} = 0$. Using (1.2.41), we then find

$$E\,\frac{dE}{dt} = -\Gamma^0_{ij} P^i P^j = -\frac{\dot{a}}{a}\,p^2 \ , \tag{1.2.48}$$

where we have written $P^0 \equiv E$ and defined the amplitude of the *physical* three-momentum as

$$p^2 \equiv -g_{ij} P^i P^j = a^2 \gamma_{ij} P^i P^j \ . \tag{1.2.49}$$

Notice the appearance of the scale factor in (1.2.49) from the contraction with the spatial part of the FRW metric, $g_{ij} = -a^2\gamma_{ij}$. The components of the four-momentum satisfy the constraint $g_{\mu\nu} P^\mu P^\nu = m^2$, or $E^2 - p^2 = m^2$, where the r.h.s. vanishes for massless particles. It follows that $E\,\mathrm{d}E = p\,\mathrm{d}p$, so that (1.2.48) can be written as

$$\frac{\dot{p}}{p} = -\frac{\dot{a}}{a} \quad \Rightarrow \quad p \propto \frac{1}{a} \ . \tag{1.2.50}$$

We see that the physical three-momentum of any particle (both massive and massless) decays with the expansion of the universe.

  - For massless particles, eq. (1.2.50) implies

$$p = E \propto \frac{1}{a} \quad \text{(massless particles)} \ , \tag{1.2.51}$$

    i.e. the energy of massless particles decays with the expansion.

  - For massive particles, eq. (1.2.50) implies

$$p = \frac{m v}{\sqrt{1 - v^2}} \propto \frac{1}{a} \quad \text{(massive particles)} \ , \tag{1.2.52}$$

    where $v^i = dx^i/dt$ is the *comoving* peculiar velocity of the particles (i.e. the velocity relative to the comoving frame) and $v^2 \equiv a^2 \gamma_{ij} v^i v^j$ is the magnitude of the *physical* peculiar velocity, cf. eq. (1.1.17). To get the first equality in (1.2.52), I have used

$$P^i = m U^i = m\,\frac{dX^i}{ds} = m\,\frac{dt}{ds} v^i = \frac{m v^i}{\sqrt{1 - a^2 \gamma_{ij} v^i v^j}} = \frac{m v^i}{\sqrt{1 - v^2}} \ . \tag{1.2.53}$$

    Eq. (1.2.52) shows that freely-falling particles left on their own will converge onto the Hubble flow.

1.2.2 Redshift

Everything we know about the universe is inferred from the light we receive from distant objects. The light emitted by a distant galaxy can be viewed either quantum mechanically as freely-propagating photons, or classically as propagating electromagnetic waves. To interpret the observations correctly, we need to take into account that the wavelength of the light gets stretched (or, equivalently, the photons lose energy) by the expansion of the universe. We now quantify this effect.

Redshifting of photons.—In the quantum mechanical description, the wavelength of light is inversely proportional to the photon momentum, $\lambda = h/p$. Since according to (1.2.51) the momentum of a photon evolves as $a(t)^{-1}$, the wavelength scales as $a(t)$. Light emitted at time t_1 with wavelength λ_1 will be observed at t_0 with wavelength

$$\lambda_0 = \frac{a(t_0)}{a(t_1)}\,\lambda_1 \ . \tag{1.2.54}$$

Since $a(t_0) > a(t_1)$, the wavelength of the light increases, $\lambda_0 > \lambda_1$.

Redshifting of classical waves.—We can derive the same result by treating light as classical electromagnetic waves. Consider a galaxy at a fixed comoving distance d. At a time τ_1, the galaxy emits a signal of short conformal duration $\Delta\tau$ (see fig. 1.5). According to (1.1.24), the light arrives at our telescopes at time $\tau_0 = \tau_1 + d$. The conformal duration of the signal measured by the detector is the same as at the source, but the physical time intervals are different at the points of emission and detection,

$$\Delta t_1 = a(\tau_1)\Delta\tau \quad \text{and} \quad \Delta t_0 = a(\tau_0)\Delta\tau \ . \tag{1.2.55}$$

If Δt is the period of the light wave, the light is emitted with wavelength $\lambda_1 = \Delta t_1$ (in units where $c = 1$), but is observed with wavelength $\lambda_0 = \Delta t_0$, so that

$$\frac{\lambda_0}{\lambda_1} = \frac{a(\tau_0)}{a(\tau_1)} \ . \tag{1.2.56}$$

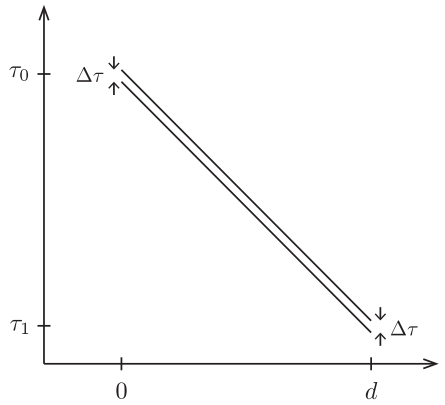


Figure 1.5: *In conformal time, the period of a light wave ($\Delta\tau$) is equal at emission (τ_1) and at observation (τ_0). However, measured in physical time ($\Delta t = a(\tau)\Delta\tau$) the period is longer when it reaches us, $\Delta t_0 > \Delta t_1$. We say that the light has redshifted since its wavelength is now longer, $\lambda_0 > \lambda_1$.*

It is conventional to define the *redshift* parameter as the fractional shift in wavelength of a photon emitted by a distant galaxy at time t_1 and observed on Earth today,

$$z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1} \ . \tag{1.2.57}$$

We then find

$$1 + z = \frac{a(t_0)}{a(t_1)} \ . \tag{1.2.58}$$

It is also common to define $a(t_0) \equiv 1$, so that

$$\boxed{1 + z = \frac{1}{a(t_1)}} \ . \tag{1.2.59}$$

Hubble's law.—For nearby sources, we may expand $a(t_1)$ in a power series,

$$a(t_1) = a(t_0)\big[1 + (t_1 - t_0)H_0 + \cdots\big] \ , \tag{1.2.60}$$

where H_0 is the *Hubble constant*

$$H_0 \equiv \frac{\dot{a}(t_0)}{a(t_0)} \ . \tag{1.2.61}$$

Eq. (1.2.58) then gives $z = H_0(t_0 - t_1) + \cdots$. For close objects, $t_0 - t_1$ is simply the physical distance d (in units with $c = 1$). We therefore find that the redshift increases linearly with distance

$$z \simeq H_0 d \ . \tag{1.2.62}$$

The slope in a redshift-distance diagram (cf. fig. 1.8) therefore measures the current expansion rate of the universe, H_0. These measurements used to come with very large uncertainties. Since H_0 normalizes everything else (see below), it became conventional to define[11]

$$H_0 \equiv 100\,h\,\mathrm{km\,s^{-1}Mpc^{-1}} \ , \tag{1.2.63}$$

where the parameter h is used to keep track of how uncertainties in H_0 propagate into other cosmological parameters. Today, measurements of H_0 have become much more precise,[12]

$$h \approx 0.67 \pm 0.01 \ . \tag{1.2.64}$$

1.2.3 Distances*

For distant objects, we have to be more careful about what we mean by "distance":

- *Metric distance.*—We first define a distance that isn't really observable, but that will be useful in defining observable distances. Consider the FRW metric in the form (1.1.21),

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\Big[\mathrm{d}\chi^2 + S_k^2(\chi)\mathrm{d}\Omega^2\Big] \ , \tag{1.2.65}$$

  where[13]

$$S_k(\chi) \equiv \begin{cases} R_0 \sinh(\chi/R_0) & k = -1 \\ \chi & k = 0 \\ R_0 \sin(\chi/R_0) & k = +1 \end{cases} \ . \tag{1.2.66}$$

  The distance multiplying the solid angle element $\mathrm{d}\Omega^2$ is the *metric distance*,

$$d_m = S_k(\chi) \ . \tag{1.2.67}$$

[11] A parsec (pc) is 3.26 light-years. Blame astronomers for the funny units in (6.3.29).

[12] Planck 2013 Results – *Cosmological Parameters* [arXiv:1303.5076].

[13] Notice that the definition of $S_k(\chi)$ contains a length scale R_0 after we chose to make the scale factor dimensionless, $a(t_0) \equiv 1$. This is achieved by using the rescaling symmetry $a \to \lambda a$, $\chi \to \chi/\lambda$, and $S_k^2 \to S_k^2/\lambda$.

In a flat universe ($k = 0$), the metric distance is simply equal to the *comoving distance* χ. The comoving distance between us and a galaxy at redshift z can be written as

$$\chi(z) = \int_{t_1}^{t_0} \frac{\mathrm{d}t}{a(t)} = \int_0^z \frac{\mathrm{d}z}{H(z)} \ , \tag{1.2.68}$$

where the redshift evolution of the Hubble parameter, $H(z)$, depends on the matter content of the universe (see §1.3). We emphasize that the comoving distance and the metric distance are not observables.

- *Luminosity distance.*—Type IA supernovae are called 'standard candles' because they are believed to be objects of known absolute luminosity L (= energy emitted per second). The observed flux F (= energy per second per receiving area) from a supernova explosion can then be used to infer its (luminosity) distance. Consider a source at a fixed comoving distance χ. In a static Euclidean space, the relation between absolute luminosity and observed flux is

$$F = \frac{L}{4\pi\chi^2} \ . \tag{1.2.69}$$

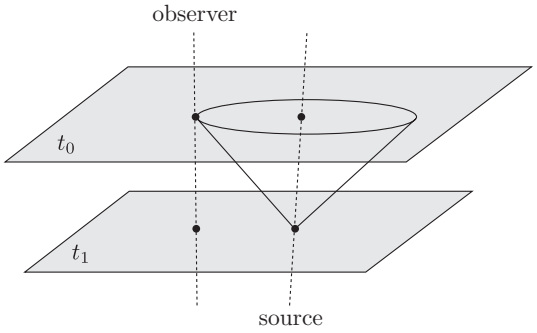


Figure 1.6: *Geometry associated with the definition of luminosity distance.*

In an FRW spacetime, this result is modified for three reasons:

1. At the time t_0 that the light reaches the Earth, the proper area of a sphere drawn around the supernova and passing through the Earth is $4\pi d_m^2$. The fraction of the light received in a telescope of aperture A is therefore $A/4\pi d_m^2$.
2. The rate of arrival of photons is lower than the rate at which they are emitted by the redshift factor $1/(1+z)$.
3. The energy E_0 of the photons when they are received is less than the energy E_1 with which they were emitted by the same redshift factor $1/(1+z)$.

Hence, the correct formula for the observed flux of a source with luminosity L at coordinate distance χ and redshift z is

$$F = \frac{L}{4\pi d_m^2 (1+z)^2} \equiv \frac{L}{4\pi d_L^2} \ , \tag{1.2.70}$$

where we have defined the *luminosity distance*, d_L, so that the relation between luminosity, flux and luminosity distance is the same as in (1.2.69). Hence, we find

$$d_L = d_m(1+z) \ . \tag{1.2.71}$$

- *Angular diameter distance.*—Sometimes we can make use of 'standard rulers', i.e. objects of known physical size D. (This is the case, for example, for the fluctuations in the CMB.) Let us assume again that the object is at a comoving distance χ and the photons which we observe today were emitted at time t_1. A naive astronomer could decide to measure the distance d_A to the object by measuring its angular size $\delta\theta$ and using the Euclidean formula for its distance,[14]

$$d_A = \frac{D}{\delta\theta} \ . \tag{1.2.72}$$

This quantity is called the *angular diameter distance*. The FRW metric (1.1.23) implies

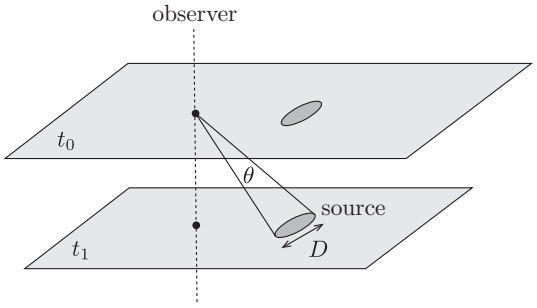


Figure 1.7: *Geometry associated with the definition of angular diameter distance.*

the following relation between the physical (transverse) size of the object and its angular size on the sky

$$D = a(t_1)S_k(\chi)\delta\theta = \frac{d_m}{1+z}\delta\theta \ . \tag{1.2.73}$$

Hence, we get

$$d_A = \frac{d_m}{1+z} \ . \tag{1.2.74}$$

The angular diameter distance measures the distance between us and the object when the light was *emitted*. We see that angular diameter and luminosity distances aren't independent, but related by

$$d_A = \frac{d_L}{(1+z)^2} \ . \tag{1.2.75}$$

Fig. 1.8 shows the redshift dependence of the three distance measures d_m, d_L, and d_A. Notice that all three distances are larger in a universe with dark energy (in the form of a cosmological constant Λ) than in one without. This fact was employed in the discovery of dark energy (see fig. 1.9 in §1.3.3).

1.3 Dynamics

The dynamics of the universe is determined by the Einstein equation

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \ . \tag{1.3.76}$$

This relates the Einstein tensor $G_{\mu\nu}$ (a measure of the "spacetime curvature" of the FRW universe) to the stress-energy tensor $T_{\mu\nu}$ (a measure of the "matter content" of the universe). We

[14] This formula assumes $\delta\theta \ll 1$ (in radians) which is true for all cosmological objects.

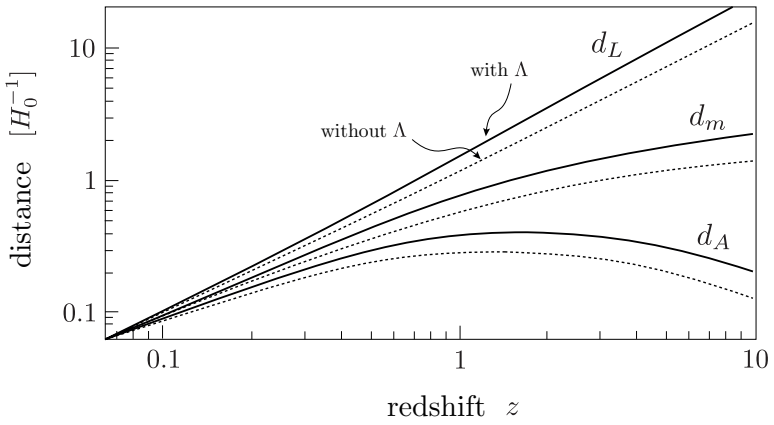


Figure 1.8: *Distance measures in a flat universe, with matter only (dotted lines) and with 70% dark energy (solid lines). In a dark energy dominated universe, distances out to a fixed redshift are larger than in a matter-dominated universe.*

will first discuss possible forms of cosmological stress-energy tensors $T_{\mu\nu}$ (§1.3.1), then compute the Einstein tensor $G_{\mu\nu}$ for the FRW background (§1.3.2), and finally put them together to solve for the evolution of the scale factor $a(t)$ as a function of the matter content (§1.3.3).

1.3.1 Matter Sources

We first show that the requirements of isotropy and homogeneity force the coarse-grained stress-energy tensor to be that of a *perfect fluid*,

$$T_{\mu\nu} = (\rho + P)\, U_\mu U_\nu - P\, g_{\mu\nu} \ , \tag{1.3.77}$$

where ρ and P are the *energy density* and the *pressure* of the fluid and U^μ is its *four-velocity* (relative to the observer).

Number Density

In fact, before we get to the stress-energy tensor, we study a simpler object: the number current four-vector N^μ. The $\mu = 0$ component, N^0, measures the number density of particles, where for us a "particle" may be an entire galaxy. The $\mu = i$ component, N^i, is the flux of the particles in the direction x^i. Isotropy requires that the mean value of any 3-vector, such as N^i, must vanish, and homogeneity requires that the mean value of any 3-scalar[15], such as N^0, is a function only of time. Hence, the current of galaxies, as measured by a comoving observer, has the following components

$$N^0 = n(t) \ , \qquad N^i = 0 \ , \tag{1.3.78}$$

where $n(t)$ is the number of galaxies per proper volume as measured by a comoving observer. A general observer (i.e. an observer in motion relative to the mean rest frame of the particles), would measure the following number current four-vector

$$N^\mu = n U^\mu \ , \tag{1.3.79}$$

where $U^\mu \equiv dX^\mu/ds$ is the relative four-velocity between the particles and the observer. Of course, we recover the previous result (1.3.78) for a comoving observer, $U^\mu = (1, 0, 0, 0)$. For

[15] A 3-scalar is a quantity that is invariant under purely spatial coordinate transformations.

$U^\mu = \gamma(1, v^i)$, eq. (1.3.79) gives the correctly boosted results. For instance, you may recall that the boosted number density is γn. (The number density increases because one of the dimensions of the volume is Lorentz contracted.)

The number of particles has to be conserved. In Minkowski space, this implies that the evolution of the number density satisfies the continuity equation

$$\dot{N}^0 = -\partial_i N^i \ , \tag{1.3.80}$$

or, in relativistic notation,

$$\partial_\mu N^\mu = 0 \ . \tag{1.3.81}$$

Eq. (1.3.81) is generalised to curved spacetimes by replacing the partial derivative ∂_μ with a *covariant derivative* ∇_μ,[16]

$$\nabla_\mu N^\mu = 0 \ . \tag{1.3.82}$$

Eq. (1.3.82) reduces to (1.3.81) in the local intertial frame.

Covariant derivative.—The covariant derivative is an important object in differential geometry and it is of fundamental importance in general relativity. The geometrical meaning of ∇_μ will be discussed in detail in the GR course. In this course, we will have to be satisfied with treating it as an operator that acts in a specific way on scalars, vectors and tensors:

- There is no difference between the covariant derivative and the partial derivative if it acts on a scalar
$$\nabla_\mu f = \partial_\mu f \ . \tag{1.3.83}$$
- Acting on a contravariant vector, V^ν, the covariant derivative is a partial derivative plus a correction that is linear in the vector:
$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\lambda} V^\lambda \ . \tag{1.3.84}$$
Look carefully at the index structure of the second term. A similar definition applies to the covariant derivative of covariant vectors, ω_ν,
$$\nabla_\mu \omega_\nu = \partial_\mu \omega_\nu - \Gamma^\lambda_{\mu\nu} \omega_\lambda \ . \tag{1.3.85}$$
Notice the change of the sign of the second term and the placement of the dummy index.
- For tensors with many indices, you just repeat (1.3.84) and (1.3.85) for each index. For each upper index you introduce a term with a single $+\Gamma$, and for each lower index a term with a single $-\Gamma$:
$$\begin{aligned} \nabla_\sigma T^{\mu_1\mu_2\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} \ = \ & \partial_\sigma T^{\mu_1\mu_2\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} \\ & + \ \Gamma^{\mu_1}{}_{\sigma\lambda} T^{\lambda\mu_2\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} + \Gamma^{\mu_2}{}_{\sigma\lambda} T^{\mu_1\lambda\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} + \cdots \\ & - \ \Gamma^{\lambda}{}_{\sigma\nu_1} T^{\mu_1\mu_2\cdots\mu_k}{}_{\lambda\nu_2\cdots\nu_l} - \Gamma^{\lambda}{}_{\sigma\nu_2} T^{\mu_1\mu_2\cdots\mu_k}{}_{\nu_1\lambda\cdots\nu_l} - \cdots \ . \end{aligned} \tag{1.3.86}$$
This is the general expression for the covariant derivative. Luckily, we will only be dealing with relatively simple tensors, so this monsterous expression will usually reduce to something managable.

[16] If this is the first time you have seen a covariant derivative, this will be a bit intimidating. Find me to talk about your fears.

Explicitly, eq. (1.3.82) can be written

$$\nabla_\mu N^\mu = \partial_\mu N^\mu + \Gamma^\mu_{\mu\lambda} N^\lambda = 0 \ . \tag{1.3.87}$$

Using (1.3.78), this becomes

$$\frac{dn}{dt} + \Gamma^i_{i0} n = 0 \ , \tag{1.3.88}$$

and substituting (1.2.41), we find

$$\frac{\dot{n}}{n} = -3\frac{\dot{a}}{a} \quad \Rightarrow \quad n(t) \propto a^{-3} \ . \tag{1.3.89}$$

As expected, the number density decreases in proportion to the increase of the proper volume.

Energy-Momentum Tensor

We will now use a similar logic to determine what form of the stress-energy tensor $T_{\mu\nu}$ is consistent with the requirements of homogeneity and isotropy. First, we decompose $T_{\mu\nu}$ into a 3-scalar, T_{00}, 3-vectors, T_{i0} and T_{0j}, and a 3-tensor, T_{ij}. As before, isotropy requires the mean values of 3-vectors to vanish, i.e. $T_{i0} = T_{0j} = 0$. Moreover, isotropy around a point $\boldsymbol{x} = 0$ requires the mean value of any 3-tensor, such as T_{ij}, at that point to be proportional to δ_{ij} and hence to g_{ij}, which equals $-a^2\delta_{ij}$ at $\boldsymbol{x} = 0$,

$$T_{ij}(\boldsymbol{x} = 0) \propto \delta_{ij} \propto g_{ij}(\boldsymbol{x} = 0) \ . \tag{1.3.90}$$

Homogeneity requires the proportionality coefficient to be only a function of time. Since this is a proportionality between two 3-tensors, T_{ij} and g_{ij}, it must remain unaffected by an arbitrary transformation of the spatial coordinates, including those transformations that preserve the form of g_{ij} while taking the origin into any other point. Hence, homogeneity and isotropy require the components of the stress-energy tensor everywhere to take the form

$$T_{00} = \rho(t) \ , \qquad \pi_i \equiv T_{i0} = 0 \ , \qquad T_{ij} = -P(t)\, g_{ij}(t, \boldsymbol{x}) \ . \tag{1.3.91}$$

It looks even nicer with mixed upper and lower indices

$$T^\mu{}_\nu = g^{\mu\lambda} T_{\lambda\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -P & 0 & 0 \\ 0 & 0 & -P & 0 \\ 0 & 0 & 0 & -P \end{pmatrix} . \tag{1.3.92}$$

This is the stress-energy tensor of a *perfect fluid* as seen by a comoving observer. More generally, the stress-energy tensor can be written in the following, explicitly covariant, form

$$T^\mu{}_\nu = (\rho + P)\, U^\mu U_\nu - P\, \delta^\mu_\nu \ , \tag{1.3.93}$$

where $U^\mu \equiv dX^\mu/ds$ is the relative four-velocity between the fluid and the observer, while ρ and P are the energy density and pressure in the *rest-frame* of the fluid. Of course, we recover the previous result (1.3.92) for a comoving observer, $U^\mu = (1, 0, 0, 0)$.

How do the density and pressure evolve with time? In Minkowski space, energy and momentum are conserved. The energy density therefore satisfies the continuity equation $\dot{\rho} = -\partial_i \pi^i$, i.e. the rate of change of the density equals the divergence of the energy flux. Similarly, the

evolution of the momentum density satisfies the Euler equation, $\dot{\pi}_i = \partial_i P$. These conservation laws can be combined into a four-component conservation equation for the stress-energy tensor

$$\partial_\mu T^\mu{}_\nu = 0 \ . \tag{1.3.94}$$

In general relativity, this is promoted to the covariant conservation equation

$$\nabla_\mu T^\mu{}_\nu = \partial_\mu T^\mu{}_\nu + \Gamma^\mu_{\mu\lambda} T^\lambda{}_\nu - \Gamma^\lambda_{\mu\nu} T^\mu{}_\lambda = 0 \ . \tag{1.3.95}$$

Eq. (1.3.95) reduces to (1.3.94) in the local intertial frame. This corresponds to four separate equations (one for each ν). The evolution of the energy density is determined by the $\nu = 0$ equation

$$\partial_\mu T^\mu{}_0 + \Gamma^\mu_{\mu\lambda} T^\lambda{}_0 - \Gamma^\lambda_{\mu 0} T^\mu{}_\lambda = 0 \ . \tag{1.3.96}$$

Since $T^i{}_0$ vanishes by isotropy, this reduces to

$$\frac{d\rho}{dt} + \Gamma^\mu_{\mu 0}\rho - \Gamma^\lambda_{\mu 0} T^\mu{}_\lambda = 0 \ . \tag{1.3.97}$$

From eq. (1.2.41) we see that $\Gamma^\lambda_{\mu 0}$ vanishes unless λ and μ are spatial indices equal to each other, in which case it is $\dot{a}/a$. The continuity equation (1.3.97) therefore reads

$$\boxed{\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0} \ . \tag{1.3.98}$$

Exercise.—Show that (1.3.98) can be written as, $\mathrm{d}U = -P\mathrm{d}V$, where $U = \rho V$ and $V \propto a^3$.

Cosmic Inventory

The universe is filled with a mixture of different matter components. It is useful to classify the different sources by their contribution to the pressure:

- **Matter**

  We will use the term "matter" to refer to all forms of matter for which the pressure is much smaller than the energy density, $|P| \ll \rho$. As we will show in Chapter 3, this is the case for a gas of non-relativistic particles (where the energy density is dominated by the mass). Setting $P = 0$ in (1.3.98) gives

  $$\rho \propto a^{-3} \ . \tag{1.3.99}$$

  This dilution of the energy density simply reflects the expansion of the volume $V \propto a^3$.

  - **Dark matter.** Most of the matter in the universe is in the form of invisible dark matter. This is usually thought to be a new heavy particle species, but what it really is, we don't know.
  - **Baryons.** Cosmologists refer to ordinary matter (nuclei and electrons) as baryons.[17]

[17] Of course, this is technically incorrect (electrons are *leptons*), but nuclei are so much heavier than electrons that most of the mass is in the baryons. If this terminology upsets you, you should ask your astronomer friends what they mean by "metals".

- **Radiation**

  We will use the term "radiation" to denote anything for which the pressure is about a third of the energy density, $P = \frac{1}{3}\rho$. This is the case for a gas of relativistic particles, for which the energy density is dominated by the kinetic energy (i.e. the momentum is much bigger than the mass). In this case, eq. (1.3.98) implies

  $$\rho \propto a^{-4} \ . \tag{1.3.100}$$

  The dilution now includes the redshifting of the energy, $E \propto a^{-1}$.

  - **Photons.** The early universe was dominated by photons. Being massless, they are always relativistic. Today, we detect those photons in the form of the cosmic microwave background.
  - **Neutrinos.** For most of the history of the universe, neutrinos behaved like radiation. Only recently have their small masses become relevant and they started to behave like matter.
  - **Gravitons.** The early universe may have produced a background of gravitons (i.e. gravitational waves, see §6.5.2). Experimental efforts are underway to detect them.

- **Dark energy**

  We have recently learned that matter and radiation aren't enough to describe the evolution of the universe. Instead, the universe today seems to be dominated by a mysterious *negative* pressure component, $P = -\rho$. This is unlike anything we have ever encountered in the lab. In particular, from eq. (1.3.98), we find that the energy *density* is constant,

  $$\rho \propto a^{0} \ . \tag{1.3.101}$$

  Since the energy density doesn't dilute, energy has to be created as the universe expands.[18]

  - **Vacuum energy.** In quantum field theory, this effect is actually predicted! The ground state energy of the vacuum corresponds to the following stress-energy tensor

    $$T^{\rm vac}_{\mu\nu} = \rho_{\rm vac}\, g_{\mu\nu} \ . \tag{1.3.102}$$

    Comparison with eq. (1.3.93), show that this indeed implies $P_{\rm vac} = -\rho_{\rm vac}$. Unfortunately, the predicted size of $\rho_{\rm vac}$ is completely off,

    $$\frac{\rho_{\rm vac}}{\rho_{\rm obs}} \sim 10^{120} \ . \tag{1.3.103}$$

  - **Something else?** The failure of quantum field theory to explain the size of the observed dark energy has lead theorists to consider more exotic possibilities (such as time-varying dark energy and modifications of general relativity). In my opinion, none of these ideas works very well.

[18] In a gravitational system this doesn't have to violate the conservation of energy. It is the conservation equation (1.3.98) that counts.

Cosmological constant.—The left-hand side of the Einstein equation (1.3.76) isn't uniquely defined. We can add the term $-\Lambda g_{\mu\nu}$, for some constant Λ, without changing the conservation of the stress tensor, $\nabla^\mu T_{\mu\nu} = 0$ (recall, or check, that $\nabla^\mu g_{\mu\nu} = 0$). In other words, we could have written the Einstein equation as

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \ . \tag{1.3.104}$$

Einstein, in fact, did add such a term and called it the *cosmological constant*. However, it has become modern practice to move this term to the r.h.s. and treat it as a contribution to the stress-energy tensor of the form

$$T^{(\Lambda)}_{\mu\nu} = \frac{\Lambda}{8\pi G}\, g_{\mu\nu} \equiv \rho_\Lambda\, g_{\mu\nu} \ . \tag{1.3.105}$$

This is of the same form as the stress-energy tensor from vacuum energy, eq. (1.3.102).

Summary

Most cosmological fluids can be parameterised in terms of a constant equation of state: $w = P/\rho$. This includes cold dark matter ($w = 0$), radiation ($w = 1/3$) and vacuum energy ($w = -1$). In that case, the solutions to (1.3.98) scale as

$$\rho \propto a^{-3(1+w)} \ , \tag{1.3.106}$$

and hence

$$\rho \propto \begin{cases} a^{-3} & \text{matter} \\ a^{-4} & \text{radiation} \\ a^{0} & \text{vacuum} \end{cases} \ . \tag{1.3.107}$$

1.3.2 Spacetime Curvature

We want to relate these matter sources to the evolution of the scale factor in the FRW metric (1.1.14). To do this we have to compute the Einstein tensor on the l.h.s. of the Einstein equation (1.3.76),

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} \ . \tag{1.3.108}$$

We will need the Ricci tensor

$$R_{\mu\nu} \equiv \partial_\lambda \Gamma^\lambda_{\mu\nu} - \partial_\nu \Gamma^\lambda_{\mu\lambda} + \Gamma^\lambda_{\lambda\rho}\Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\rho} \ , \tag{1.3.109}$$

and the Ricci scalar

$$R = R^\mu{}_\mu = g^{\mu\nu} R_{\mu\nu} \ . \tag{1.3.110}$$

Again, there is a lot of beautiful geometry behind these definitions. We will simply keep plugging-and-playing: given the Christoffel symbols (1.2.41) nothing stops us from computing (1.3.109).

We don't need to calculate $R_{i0} = R_{0i}$, because it is a 3-vector, and therefore must vanish due to the isotropy of the Robertson-Walker metric. (Try it, if you don't believe it!) The non-vanishing components of the Ricci tensor are

$$R_{00} = -3\frac{\ddot{a}}{a} \ , \tag{1.3.111}$$

$$R_{ij} = -\left[\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{k}{a^2}\right] g_{ij} \ . \tag{1.3.112}$$

Notice that we had to find $R_{ij} \propto g_{ij}$ to be consistent with homogeneity and isotropy.

Derivation of R_{00}.—Setting $\mu = \nu = 0$ in (1.3.109), we have

$$R_{00} = \partial_\lambda \Gamma^\lambda_{00} - \partial_0 \Gamma^\lambda_{0\lambda} + \Gamma^\lambda_{\lambda\rho}\Gamma^\rho_{00} - \Gamma^\rho_{0\lambda}\Gamma^\lambda_{0\rho} \ , \tag{1.3.113}$$

Since Christoffels with two time-components vanish, this reduces to

$$R_{00} = -\partial_0 \Gamma^i_{0i} - \Gamma^i_{0j}\Gamma^j_{0i} \ . \tag{1.3.114}$$

where in the second line we have used that Christoffels with two time-components vanish. Using $\Gamma^i_{0j} = (\dot{a}/a)\delta^i_j$, we find

$$R_{00} = -\frac{d}{dt}\left(3\frac{\dot{a}}{a}\right) - 3\left(\frac{\dot{a}}{a}\right)^2 \ = \ -3\frac{\ddot{a}}{a} \ . \tag{1.3.115}$$

*Derivation of R_{ij}.**—Evaluating (1.3.112) is a bit tedious. A useful trick is to compute $R_{ij}(\boldsymbol{x} = 0) \propto \delta_{ij} \propto g_{ij}(\boldsymbol{x} = 0)$ using (1.1.9) and then transform the resulting relation between 3-tensors to general $\boldsymbol{x}$.

We first read off the spatial metric from (1.1.9),

$$\gamma_{ij} = \delta_{ij} + \frac{k x_i x_j}{1 - k(x_k x^k)} \ . \tag{1.3.116}$$

The key point is to think ahead and anticipate that we will set $\boldsymbol{x} = 0$ at the end. This allows us to drop many terms. You may be tempted to use $\gamma_{ij}(\boldsymbol{x} = 0) = \delta_{ij}$ straight away. However, the Christoffel symbols contain a derivative of the metric and the Riemann tensor has another derivative, so there will be terms in the final answer with two derivatives acting on the metric. These terms get a contribution from the second term in (1.3.116). However, we can ignore the denominator in the second term of γ_{ij} and use

$$\gamma_{ij} = \delta_{ij} + k x_i x_j \ . \tag{1.3.117}$$

The difference in the final answer vanishes at $\boldsymbol{x} = 0$ [do you see why?]. The derivative of (1.3.117) is

$$\partial_l \gamma_{ij} = k\left(\delta_{li} x_j + \delta_{lj} x_i\right) \ . \tag{1.3.118}$$

With this, we can evaluate

$$\Gamma^i_{jk} = \frac{1}{2}\gamma^{il}\left(\partial_j \gamma_{kl} + \partial_k \gamma_{jl} - \partial_l \gamma_{jk}\right) \ . \tag{1.3.119}$$

The inverse metric is $\gamma^{ij} = \delta^{ij} - k x^i x^j$, but the second term won't contribute when we set $\boldsymbol{x} = 0$ in the end [do you see why?], so we are free to use $\gamma^{ij} = \delta^{ij}$. Using (1.3.118) in (1.3.119), we then get

$$\Gamma^i_{jk} = k x^i \delta_{jk} \ . \tag{1.3.120}$$

This vanishes at $\boldsymbol{x} = 0$, but its derivative does not

$$\Gamma^i_{jk}(\boldsymbol{x} = 0) = 0 \quad , \quad \partial_l \Gamma^i_{jk}(\boldsymbol{x} = 0) = k\,\delta^i_l \delta_{jk} \ . \tag{1.3.121}$$

We are finally ready to evaluate the Ricci tensor R_{ij} at $\boldsymbol{x} = 0$

$$R_{ij}(\boldsymbol{x} = 0) \equiv \underbrace{\partial_\lambda \Gamma^\lambda_{ij} - \partial_j \Gamma^\lambda_{i\lambda}}_{(A)} + \underbrace{\Gamma^\lambda_{\lambda\rho}\Gamma^\rho_{ij} - \Gamma^\rho_{i\lambda}\Gamma^\lambda_{j\rho}}_{(B)} \ . \tag{1.3.122}$$

Let us first look at the two terms labelled (B). Dropping terms that are zero at $\boldsymbol{x} = 0$, I find

$$\begin{aligned}(B) &= \Gamma^l_{l0}\Gamma^0_{ij} - \Gamma^0_{il}\Gamma^l_{j0} - \Gamma^l_{i0}\Gamma^0_{jl} \\ &= 3\frac{\dot{a}}{a}a\dot{a}\delta_{ij} - a\dot{a}\delta_{ij}\frac{\dot{a}}{a}\delta^l_j - \frac{\dot{a}}{a}\delta^l_j a\dot{a}\delta_{jl} \\ &= \dot{a}^2\delta_{ij}\ . \end{aligned} \tag{1.3.123}$$

The two terms labelled (A) in (1.3.122) can be evaluated by using (1.3.121),

$$\begin{aligned}(A) &= \partial_0\Gamma^0_{ij} + \partial_l\Gamma^l_{ij} - \partial_j\Gamma^l_{il} \\ &= \partial_0(a\dot{a})\delta_{ij} + k\delta^l_l\delta_{ij} - k\delta^l_j\delta_{il} \\ &= \left(a\ddot{a} + \dot{a}^2 + 2k\right)\delta_{ij}\ . \end{aligned} \tag{1.3.124}$$

Hence, I get

$$\begin{aligned}R_{ij}(\boldsymbol{x}=0) &= (A) + (B) \\ &= \left[a\ddot{a} + 2\dot{a}^2 + 2k\right]\delta_{ij} \\ &= -\left[\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{k}{a^2}\right]g_{ij}(\boldsymbol{x}=0)\ . \end{aligned} \tag{1.3.125}$$

As a relation between tensors this holds for general $\boldsymbol{x}$, so we get the promised result (1.3.112). To be absolutely clear, I will never ask you to reproduce a nasty computation like this.

The Ricci scalar is

$$R = -6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right]\ . \tag{1.3.126}$$

Exercise.—Verify eq. (1.3.126).

The non-zero components of the Einstein tensor $G^\mu{}_\nu \equiv g^{\mu\lambda}G_{\lambda\nu}$ are

$$G^0{}_0 = 3\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right]\ , \tag{1.3.127}$$

$$G^i{}_j = \left[2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right]\delta^i_j\ . \tag{1.3.128}$$

Exercise.—Verify eqs. (1.3.127) and (1.3.128).

1.3.3 Friedmann Equations

We combine eqs. (1.3.127) and (1.3.128) with stress-tensor (1.3.92), to get the *Friedmann equations*,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}\ , \tag{1.3.129}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P)\ . \tag{1.3.130}$$

Here, ρ and P should be understood as the sum of all contributions to the energy density and pressure in the universe. We write ρ_r for the contribution from radiation (with ρ_γ for photons and ρ_ν for neutrinos), ρ_m for the contribution by matter (with ρ_c for cold dark matter and ρ_b for baryons) and ρ_Λ for the vacuum energy contribution. The first Friedmann equation is often written in terms of the Hubble parameter, $H \equiv \dot{a}/a$,

$$\boxed{H^2 = \frac{8\pi G}{3}\,\rho - \frac{k}{a^2}}\,. \tag{1.3.131}$$

Let us use subscripts '0' to denote quantities evaluated today, at $t = t_0$. A flat universe ($k = 0$) corresponds to the following *critical density* today

$$\begin{aligned}\rho_{\rm crit,0} = \frac{3H_0^2}{8\pi G} &= 1.9\times 10^{-29}\,h^2\,{\rm grams\,cm^{-3}}\\ &= 2.8\times 10^{11}\,h^2\,M_\odot\,{\rm Mpc^{-3}}\\ &= 1.1\times 10^{-5}\,h^2\,{\rm protons\,cm^{-3}}\,.\end{aligned} \tag{1.3.132}$$

We use the critical density to define dimensionless density parameters

$$\Omega_{I,0} \equiv \frac{\rho_{I,0}}{\rho_{\rm crit,0}}\,. \tag{1.3.133}$$

The Friedmann equation (1.3.131) can then be written as

$$H^2(a) = H_0^2\left[\Omega_{r,0}\left(\frac{a_0}{a}\right)^4 + \Omega_{m,0}\left(\frac{a_0}{a}\right)^3 + \Omega_{k,0}\left(\frac{a_0}{a}\right)^2 + \Omega_{\Lambda,0}\right]\,, \tag{1.3.134}$$

where we have defined a "curvature" density parameter, $\Omega_{k,0} \equiv -k/(a_0H_0)^2$. It should be noted that in the literature, the subscript '0' is normally dropped, so that e.g. Ω_m usually denotes the matter density *today* in terms of the critical density *today*. From now on we will follow this convention and drop the '0' subscripts on the density parameters. We will also use the conventional normalization for the scale factor, $a_0 \equiv 1$. Eq. (1.3.134) then becomes

$$\frac{H^2}{H_0^2} \;=\; \Omega_r\,a^{-4} + \Omega_m\,a^{-3} + \Omega_k\,a^{-2} + \Omega_\Lambda\,. \tag{1.3.135}$$

ΛCDM

Observations (see figs. 1.9 and 1.10) show that the universe is filled with radiation ('r'), matter ('m') and dark energy ('Λ'):

$$|\Omega_k| \,\le\, 0.01\,,\quad \Omega_r = 9.4\times 10^{-5}\,,\quad \Omega_m = 0.32\,,\quad \Omega_\Lambda = 0.68\,.$$

The equation of state of dark energy seems to be that of a cosmological constant, $w_\Lambda \approx -1$. The matter splits into 5% ordinary matter (baryons, 'b') and 27% (cold) dark matter (CDM, 'c'):

$$\Omega_b = 0.05\,,\quad \Omega_c = 0.27\,.$$

We see that even today curvature makes up less than 1% of the cosmic energy budget. At earlier times, the effects of curvature are then completely negligible (recall that matter and radiation scale as a^{-3} and a^{-4}, respectively, while the curvature contribution only increases as a^{-2}). For the rest of these lectures, I will therefore set $\Omega_k \equiv 0$. In Chapter 2, we will show that inflation indeed predicts that the effects of curvature should be minuscule in the early universe (see also Problem Set 2).

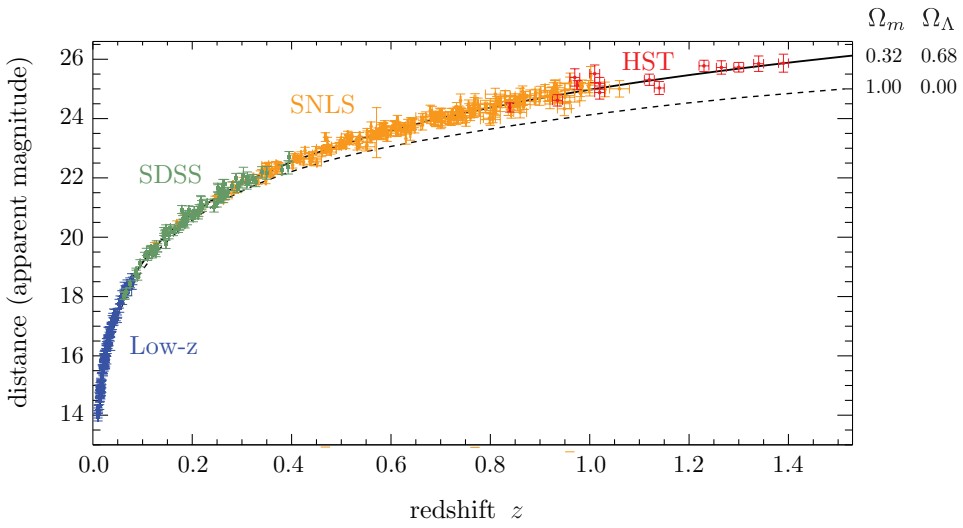


Figure 1.9: *Type IA supernovae and the discovery dark energy. If we assume a flat universe, then the supernovae clearly appear fainter (or more distant) than predicted in a matter-only universe ($\Omega_m = 1.0$). (SDSS = Sloan Digital Sky Survey; SNLS = SuperNova Legacy Survey; HST = Hubble Space Telescope.)*

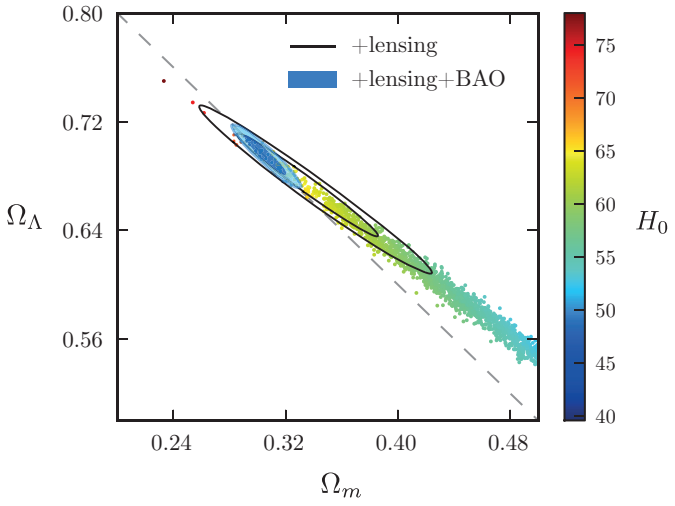


Figure 1.10: *A combination CMB and LSS observations indicate that the spatial geometry of the universe is flat. The energy density of the universe is dominated by a cosmological constant. Notice that the CMB data alone cannot exclude a matter-only universe with large spatial curvature. The evidence for dark energy requires additional input.*

Single-Component Universe

The different scalings of radiation (a^{-4}), matter (a^{-3}) and vacuum energy (a^0) imply that for most of its history the universe was dominated by a single component (first radiation, then matter, then vacuum energy; see fig. 1.11). Parameterising this component by its equation of state w_I captures all cases of interest. For a flat, single-component universe, the Friedmann equation (1.3.135) reduces to

$$\frac{\dot{a}}{a} = H_0 \sqrt{\Omega_I}\, a^{-\frac{3}{2}(1+w_I)} \ . \tag{1.3.136}$$

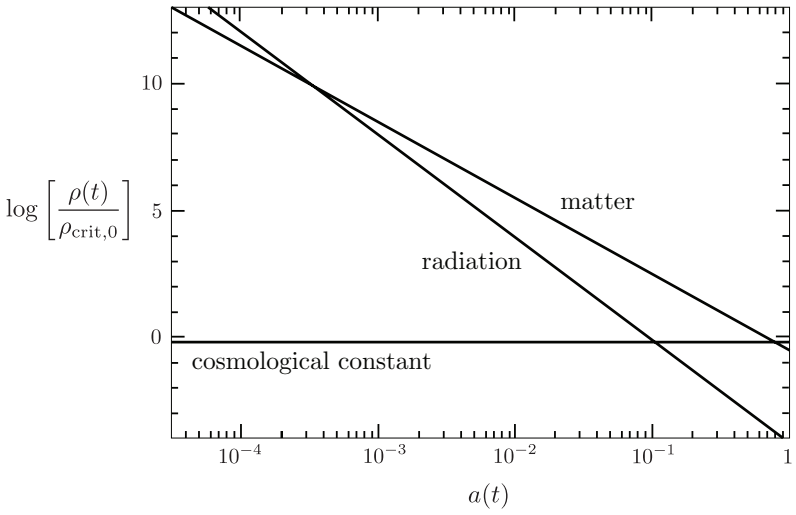


Figure 1.11: *Evolution of the energy densities in the universe.*

Integrating this equation, we obtain the time dependence of the scale factor

$$a(t) \propto \begin{cases} t^{2/3(1+w_I)} & w_I \neq -1 \quad \begin{matrix} t^{2/3} & \text{MD} \\ t^{1/2} & \text{RD} \end{matrix} \\ e^{Ht} & w_I = -1 \qquad \Lambda\text{D} \end{cases} \tag{1.3.137}$$

or, in conformal time,

$$a(\tau) \propto \begin{cases} \tau^{2/(1+3w_I)} & w_I \neq -1 \quad \begin{matrix} \tau^2 & \text{MD} \\ \tau & \text{RD} \end{matrix} \\ (-\tau)^{-1} & w_I = -1 \qquad \Lambda\text{D} \end{cases} \tag{1.3.138}$$

Exercise.—Derive eq. (1.3.138) from eq. (1.3.137).

Table 1.1 summarises the solutions for a flat universe during radiation domination (RD), matter domination (MD) and dark energy domination (ΛD).

	w	$\rho(a)$	$a(t)$	$a(\tau)$
RD	$\frac{1}{3}$	a^{-4}	$t^{1/2}$	τ
MD	0	a^{-3}	$t^{2/3}$	τ^2
ΛD	-1	a^0	e^{Ht}	$-\tau^{-1}$

Table 1.1: *FRW solutions for a flat single-component universe.*

Two-Component Universe*

Matter and radiation were equally important at $a_{\rm eq} \equiv \Omega_r/\Omega_m \approx 3\times 10^{-4}$, which was shortly before the cosmic microwave background was released (in §3.3.3, we will show that this happened at $a_{\rm rec} \approx 9\times 10^{-4}$). It will be useful to have an exact solution describing the transition era. Let us therefore consider a flat universe filled with a mixture of matter and radiation. To solve for the evolution of the scale factor, it proves convenient to move to conformal time. The Friedmann equations (1.3.129) and (1.3.130) then are

$$(a')^2 = \frac{8\pi G}{3}\rho a^4 \ , \tag{1.3.139}$$

$$a'' = \frac{4\pi G}{3}(\rho - 3P)a^3 \ , \tag{1.3.140}$$

where primes denote derivatives with respect to conformal time and

$$\rho \equiv \rho_m + \rho_r = \frac{\rho_{\rm eq}}{2}\left[\left(\frac{a_{\rm eq}}{a}\right)^3 + \left(\frac{a_{\rm eq}}{a}\right)^4\right] \ . \tag{1.3.141}$$

Exercise.—Derive eqs. (1.3.139) and (1.3.140). You will first need to convince yourself that $\dot a = a'/a$ and $\ddot a = a''/a^2 - (a')^2/a^3$.

Notice that radiation doesn't contribute as a source term in eq. (1.3.140), $\rho_r - 3P_r = 0$. Moreover, since $\rho_m a^3 = const. = \frac{1}{2}\rho_{\rm eq}a_{\rm eq}^3$, we can write eq. (1.3.140) as

$$a'' = \frac{2\pi G}{3}\rho_{\rm eq}a_{\rm eq}^3 \ . \tag{1.3.142}$$

This equation has the following solution

$$a(\tau) = \frac{\pi G}{3}\rho_{\rm eq}a_{\rm eq}^3\tau^2 + C\tau + D \ . \tag{1.3.143}$$

Imposing $a(\tau = 0) \equiv 0$, fixes one integration constant, $D = 0$. We find the second integration constant by substituting (1.3.143) and (1.3.141) into (1.3.139),

$$C = \left(\frac{4\pi G}{3}\rho_{\rm eq}a_{\rm eq}^4\right)^{1/2} \ . \tag{1.3.144}$$

Eq. (1.3.143) can then be written as

$$a(\tau) = a_{\rm eq}\left[\left(\frac{\tau}{\tau_\star}\right)^2 + 2\left(\frac{\tau}{\tau_\star}\right)\right] \ , \tag{1.3.145}$$

where

$$\tau_\star \equiv \left(\frac{\pi G}{3}\rho_{\rm eq}a_{\rm eq}^2\right)^{-1/2} = \frac{\tau_{\rm eq}}{\sqrt{2}-1} \ . \tag{1.3.146}$$

For $\tau \ll \tau_{\rm eq}$, we recover the radiation-dominated limit, $a \propto \tau$, while for $\tau \gg \tau_{\rm eq}$, we agree with the matter-dominated limit, $a \propto \tau^2$.

2 Inflation

The FRW cosmology described in the previous chapter is incomplete. It doesn't explain why the universe is homogeneous and isotropic on large scales. In fact, the standard cosmology predicts that the early universe was made of many causally disconnected regions of space. The fact that these apparently disjoint patches of space have very nearly the same densities and temperatures is called the *horizon problem*. In this chapter, I will explain how inflation—an early period of accelerated expansion—drives the primordial universe towards homogeneity and isotropy, even if it starts in a more generic initial state.

Throughout this chapter, we will trade Newton's constant for the (reduced) Planck mass,

$$M_{\rm pl} \equiv \sqrt{\frac{\hbar c}{8\pi G}} = 2.4 \times 10^{18}\,{\rm GeV}\ ,$$

so that the Friedmann equation (1.3.131) is written as $H^2 = \rho/(3M_{\rm pl}^2)$.

2.1 The Horizon Problem

2.1.1 Light and Horizons

The size of a causal patch of space it determined by how far light can travel in a certain amount of time. As we mentioned in §1.1.3, in an expanding spacetime the propagation of light (photons) is best studied using conformal time. Since the spacetime is isotropic, we can always define the coordinate system so that the light travels purely in the radial direction (i.e. $\theta = \phi = const.$). The evolution is then determined by a two-dimensional line element[1]

$$ {\rm d}s^2 = a^2(\tau)\left[{\rm d}\tau^2 - {\rm d}\chi^2\right]\ . \tag{2.1.1}$$

Since photons travel along null geodesics, ${\rm d}s^2 = 0$, their path is defined by

$$\Delta\chi(\tau) = \pm\Delta\tau\ , \tag{2.1.2}$$

where the plus sign corresponds to outgoing photons and the minus sign to incoming photons. This shows the main benefit of working with conformal time: light rays correspond to straight lines at 45° angles in the χ-τ coordinates. If instead we had used physical time t, then the light cones for curved spacetimes would be curved. With these preliminaries, we now define two different types of cosmological horizons. One which limits the distances at which past events can be observed and one which limits the distances at which it will ever be possible to observe future events.

[1] For the radial coordinate χ we have used the parameterisation of (1.1.23), so that (2.1.1) is conformal to two-dimensional Minkowski space and the curvature k of the three-dimensional spatial slices is absorbed into the definition of the coordinate χ. Had we used the regular polar coordinate r, the two-dimensional line element would have retained a dependence on k. For flat slices, χ and r are of course the same.

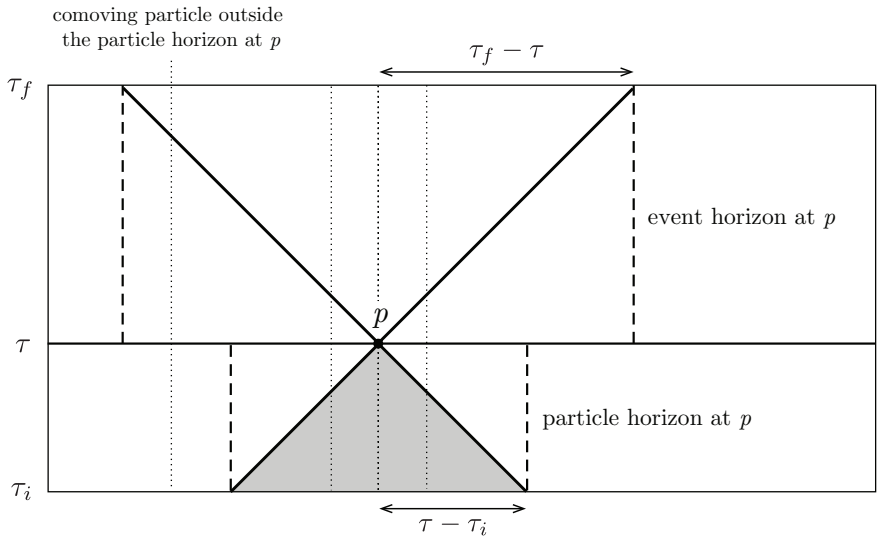


Figure 2.1: *Spacetime diagram illustrating the concept of horizons. Dotted lines show the worldlines of comoving objects. The event horizon is the maximal distance to which we can send signal. The particle horizon is the maximal distance from which we can receive signals.*

- *Particle horizon.*—Eq. (2.1.2) tells us that the maximal comoving distance that light can travel between two times τ_1 and $\tau_2 > \tau_1$ is simply $\Delta\tau = \tau_2 - \tau_1$ (recall that $c \equiv 1$). Hence, if the Big Bang 'started' with the singularity at $t_i \equiv 0$,[2] then the greatest comoving distance from which an observer at time t will be able to receive signals travelling at the speed of light is given by

$$\chi_{\text{ph}}(\tau) = \tau - \tau_i = \int_{t_i}^{t} \frac{\mathrm{d}t}{a(t)} \,. \tag{2.1.3}$$

  This is called the (comoving) particle horizon. The size of the particle horizon at time τ may be visualised by the intersection of the past light cone of an observer p with the spacelike surface $\tau = \tau_i$ (see fig. 2.1). Causal influences have to come from within this region. Only comoving particles whose worldlines intersect the past light cone of p can send a signal to an observer at p. The boundary of the region containing such worldlines is the particle horizon at p. Notice that every observer has his of her own particle horizon.

- *Event horizon.*—Just as there are past events that we cannot see now, there may be future events that we will never be able to see (and distant regions that we will never be able to influence). In comoving coordinates, the greatest distance from which an observer at time t_f will receive signals emitted at any time later than t is given by

$$\chi_{\text{eh}}(\tau) = \tau_f - \tau = \int_{t}^{t_f} \frac{\mathrm{d}t}{a(t)} \,. \tag{2.1.4}$$

  This is called the (comoving) event horizon. It is similar to the event horizon of black holes. Here, τ_f denotes the 'final moment of (conformal) time'. Notice that this may be finite even if physical time is infinite, $t_f = +\infty$. Whether this is the case or not depends on the form of $a(t)$. In particular, τ_f is finite for our universe, if dark energy is really a cosmological constant.

[2]Notice that the Big Bang singularity is a *moment in time*, but **not** *a point space*. Indeed, in figs. 2.1 and 2.2 we describe the singularity by an extended (possibly infinite) spacelike hypersurface.

2.1.2 Growing Hubble Sphere

It is the particle horizon that is relevant for the horizon problem of the standard Big Bang cosmology. Eq. (2.1.3) can be written in the following illuminating way

$$\chi_{\rm ph}(\tau) = \int_{t_i}^{t} \frac{\mathrm{d}t}{a} = \int_{a_i}^{a} \frac{\mathrm{d}a}{a\dot{a}} = \int_{\ln a_i}^{\ln a} (aH)^{-1}\, \mathrm{d}\ln a \ , \tag{2.1.5}$$

where $a_i \equiv 0$ corresponds to the Big Bang singularity. The causal structure of the spacetime can hence be related to the evolution of the *comoving Hubble radius* $(aH)^{-1}$. For a universe dominated by a fluid with constant equation of state $w \equiv P/\rho$, we get

$$(aH)^{-1} = H_0^{-1}\, a^{\frac{1}{2}(1+3w)} \ . \tag{2.1.6}$$

Note the dependence of the exponent on the combination $(1+3w)$. All familiar matter sources satisfy the strong energy condition (SEC), $1+3w > 0$, so it used to be a standard assumption that the comoving Hubble radius increases as the universe expands. In this case, the integral in (2.1.5) is dominated by the upper limit and receives vanishing contributions from early times. We see this explicitly in the example of a perfect fluid. Using (2.1.6) in (2.1.5), we find

$$\chi_{\rm ph}(a) = \frac{2H_0^{-1}}{(1+3w)} \left[a^{\frac{1}{2}(1+3w)} - a_i^{\frac{1}{2}(1+3w)} \right] \equiv \tau - \tau_i \ . \tag{2.1.7}$$

The fact that the comoving horizon receives its largest contribution from late times can be made manifest by defining

$$\tau_i \equiv \frac{2H_0^{-1}}{(1+3w)}\, a_i^{\frac{1}{2}(1+3w)} \quad \xrightarrow{a_i \to 0\,,\, w > -\frac{1}{3}} \quad 0 \ . \tag{2.1.8}$$

The comoving horizon is finite,

$$\chi_{\rm ph}(t) = \frac{2H_0^{-1}}{(1+3w)}\, a(t)^{\frac{1}{2}(1+3w)} = \frac{2}{(1+3w)}\, (aH)^{-1} \ . \tag{2.1.9}$$

We see that in the standard cosmology $\chi_{\rm ph} \sim (aH)^{-1}$. This has lead to the confusing practice of referring to both the particle horizon and the Hubble radius as the "horizon" (see §2.2.2).

2.1.3 Why is the CMB so uniform?

About 380 000 years after the Big Bang, the universe had cooled enough to allow the formation of hydrogen atoms and the decoupling of photons from the primordial plasma (see §3.3.3). We observe this event in the form of the cosmic microwave background (CMB), an afterglow of the hot Big Bang. Remarkably, this radiation is almost perfectly isotropic, with anisotropies in the CMB temperature being smaller than one part in ten thousand.

A moment's thought will convince you that the finiteness of the conformal time elapsed between $t_i = 0$ and the time of the formation of the CMB, t_{rec}, implies a serious problem: it means that most spots in the CMB have non-overlapping past light cones and hence never were in causal contact. This is illustrated by the spacetime diagram in fig. 2.2. Consider two opposite directions on the sky. The CMB photons that we receive from these directions were emitted at the points labelled p and q in fig. 2.2. We see that the photons were emitted sufficiently close to the Big Bang singularity that the past light cones of p and q don't overlap. This implies that no point lies inside the particle horizons of both p and q. So the puzzle is: how do the photons

coming from p and q "know" that they should be at almost exactly the same temperature? The same question applies to any two points in the CMB that are separated by more than 1 degree in the sky. The homogeneity of the CMB spans scales that are much larger than the particle horizon at the time when the CMB was formed. In fact, in the standard cosmology the CMB is made of about 10^4 disconnected patches of space. If there wasn't enough time for these regions to communicate, why do they look so similar? This is the *horizon problem*.

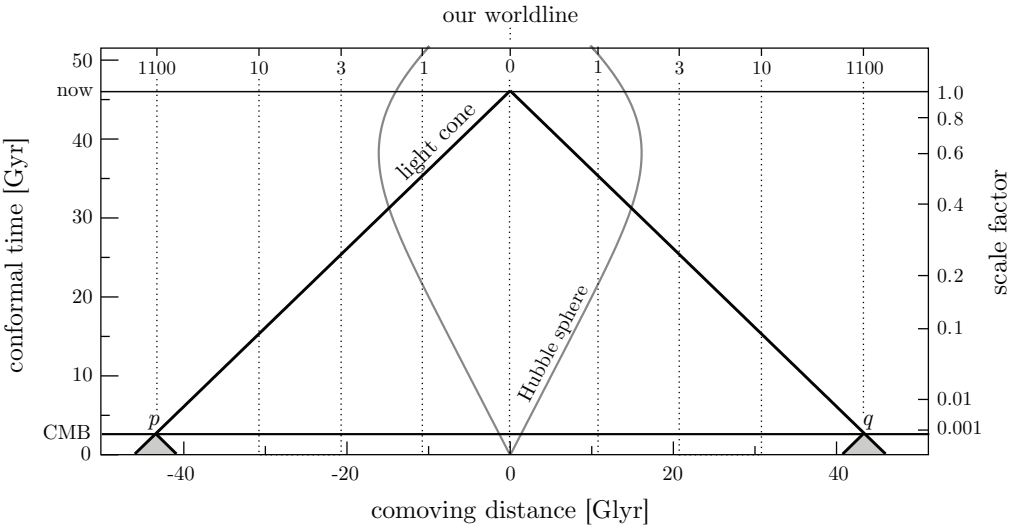


Figure 2.2: *The horizon problem in the conventional Big Bang model. All events that we currently observe are on our past light cone. The intersection of our past light cone with the spacelike slice labelled CMB corresponds to two opposite points in the observed CMB. Their past light cones don't overlap before they hit the singularity, $a = 0$, so the points appear never to have been in causal contact. The same applies to any two points in the CMB that are separated by more than 1 degree on the sky.*

2.2 A Shrinking Hubble Sphere

Our description of the horizon problem has highlighted the fundamental role played by the growing Hubble sphere of the standard Big Bang cosmology. A simple solution to the horizon problem therefore suggests itself: let us conjecture a phase of *decreasing Hubble radius* in the early universe,

$$\frac{d}{dt}(aH)^{-1} \;<\; 0\;. \tag{2.2.10}$$

If this lasts long enough, the horizon problem can be avoided. Physically, the shrinking Hubble sphere requires a SEC-violating fluid, $1 + 3w < 0$.

2.2.1 Solution of the Horizon Problem

For a shrinking Hubble sphere, the integral in (2.1.5) is dominated by the lower limit. The Big Bang singularity is now pushed to *negative conformal time*,

$$\tau_i = \frac{2H_0^{-1}}{(1+3w)}\, a_i^{\frac{1}{2}(1+3w)} \quad \xrightarrow{a_i\to 0\,,\, w<-\frac{1}{3}} \quad -\infty\;. \tag{2.2.11}$$

This implies that there was "much more conformal time between the singularity and decoupling than we had thought"! Fig. 2.3 shows the new spacetime diagram. The past light cones of

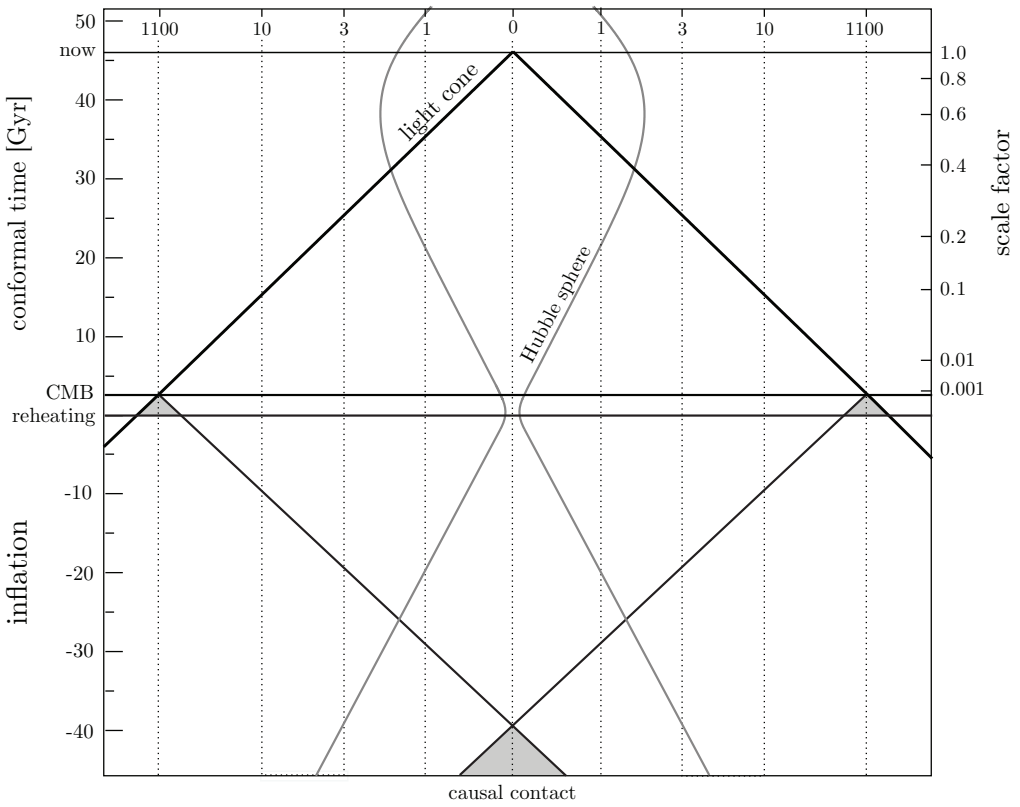


Figure 2.3: *Inflationary solution to the horizon problem. The comoving Hubble sphere shrinks during inflation and expands during the conventional Big Bang evolution (at least until dark energy takes over at $a \approx 0.5$). Conformal time during inflation is negative. The spacelike singularity of the standard Big Bang is replaced by the reheating surface, i.e. rather than marking the beginning of time it now corresponds simply to the transition from inflation to the standard Big Bang evolution. All points in the CMB have overlapping past light cones and therefore originated from a causally connected region of space.*

widely separated points in the CMB now had enough time to intersect before the time τ_i. The uniformity of the CMB is not a mystery anymore. In inflationary cosmology, $\tau = 0$ isn't the initial singularity, but instead becomes only a transition point between inflation and the standard Big Bang evolution. There is time both before and after $\tau = 0$.

2.2.2 Hubble Radius vs. Particle Horizon

A quick word of warning about bad (but unfortunately standard) language in the inflationary literature: Both the particle horizon χ_{ph} and the Hubble radius $(aH)^{-1}$ are often referred to simply as the "horizon". In the standard FRW evolution (with ordinary matter) the two are roughly the same—cf. eq. (2.1.9)—so giving them the same name isn't an issue. However, the whole point of inflation is to make the particle horizon much larger than the Hubble radius.

The Hubble radius $(aH)^{-1}$ is the (comoving) distance over which particles can travel in the course of one expansion time.[3] It is therefore another way of measuring whether particles are causally connected with each other: comparing the comoving separation λ of two particles with $(aH)^{-1}$ determines whether the particles can communicate with each other *at a given moment* (i.e. within the next Hubble time). This makes it clear that χ_{ph} and $(aH)^{-1}$ are conceptually very different:

[3]The expansion time, $t_H \equiv H^{-1} = dt/d\ln a$, is roughly the time in which the scale factor doubles.

- if $\lambda > \chi_{\rm ph}$, then the particles could *never* have communicated.
- if $\lambda > (aH)^{-1}$, then the particles cannot talk to each other *now*.

Inflation is a mechanism to achieve $\chi_{\rm ph} \gg (aH)^{-1}$. This means that particles can't communicate now (or when the CMB was created), but were in causal contact early on. In particular, the shrinking Hubble sphere means that particles which were initially in causal contact with another—i.e. separated by a distance $\lambda < (a_I H_I)^{-1}$—can no longer communicate after a sufficiently long period of inflation: $\lambda > (aH)^{-1}$; see fig. 2.4. However, at any moment before horizon exit (careful: I really mean exit of the Hubble radius!) the particles could still talk to each other and establish similar conditions. Everything within the Hubble sphere at the beginning of inflation, $(a_I H_I)^{-1}$, was causally connected.

Since the Hubble radius is easier to calculate than the particle horizon it is common to use the Hubble radius as a means of judging the horizon problem. If the entire observable universe was within the comoving Hubble radius at the beginning of inflation—i.e. $(a_I H_I)^{-1}$ was larger than the comoving radius of the observable universe $(a_0 H_0)^{-1}$—then there is no horizon problem. Notice that this is more conservative than using the particle horizon since $\chi_{\rm ph}(t)$ is always bigger than $(aH)^{-1}(t)$. Moreover, using $(a_I H_I)^{-1}$ as a measure of the horizon problem means that we don't have to assume anything about earlier times $t < t_I$.

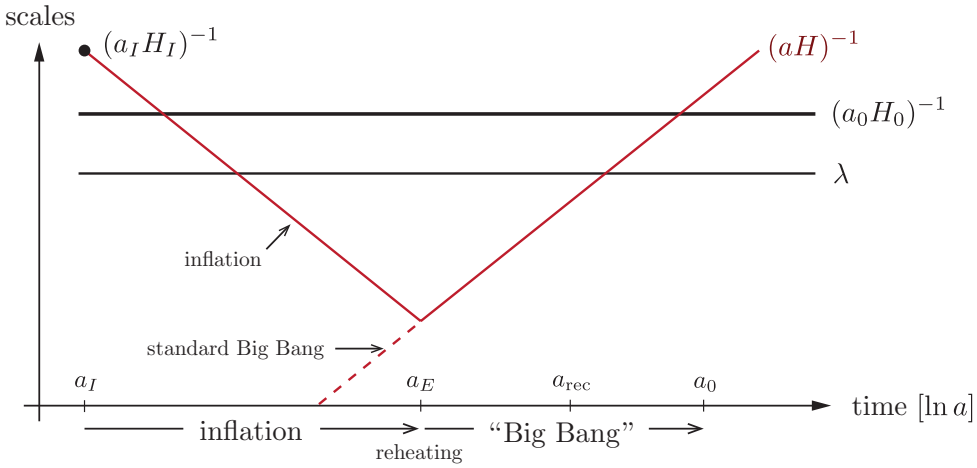


Figure 2.4: *Scales of cosmological interest were larger than the Hubble radius until $a \approx 10^{-5}$ (where today is at $a(t_0) \equiv 1$). However, at very early times, before inflation operated, all scales of interest were smaller than the Hubble radius and therefore susceptible to microphysical processing. Similarly, at very late times, the scales of cosmological interest are back within the Hubble radius. Notice the symmetry of the inflationary solution. Scales just entering the horizon today, 60 e-folds after the end of inflation, left the horizon 60 e-folds before the end of inflation.*

Duration of inflation.—How much inflation do we need to solve the horizon problem? At the very least, we require that the observable universe today fits in the comoving Hubble radius at the beginning of inflation,

$$(a_0 H_0)^{-1} < (a_I H_I)^{-1} \ . \tag{2.2.12}$$

Let us assume that the universe was radiation dominated since the end of inflation and ignore the relatively recent matter- and dark energy-dominated epochs. Remembering that $H \propto a^{-2}$ during radiation domination, we have

$$\frac{a_0 H_0}{a_E H_E} \sim \frac{a_0}{a_E}\left(\frac{a_E}{a_0}\right)^2 = \frac{a_E}{a_0} \sim \frac{T_0}{T_E} \sim 10^{-28} \ , \tag{2.2.13}$$

where in the numerical estimate we used $T_E \sim 10^{15}$ GeV and $T_0 = 10^{-3}$ eV (~ 2.7 K). Eq. (2.2.12) can then be written as

$$(a_I H_I)^{-1} > (a_0 H_0)^{-1} \sim 10^{28} (a_E H_E)^{-1} \ . \tag{2.2.14}$$

For inflation to solve the horizon problem, $(aH)^{-1}$ should therefore shrink by a factor of 10^{28}. The most common way to arrange this it to have $H \sim const.$ during inflation (see below). This implies $H_I \approx H_E$, so eq. (2.2.14) becomes

$$\frac{a_E}{a_I} > 10^{28} \quad \Rightarrow \quad \ln\left(\frac{a_E}{a_I}\right) > 64 \ . \tag{2.2.15}$$

This is the famous statement that the solution of the horizon problem requires about 60 e-folds of inflation.

2.2.3 Conditions for Inflation

I like the shrinking Hubble sphere as the fundamental definition of inflation since it relates most directly to the horizon problem and is also key for the inflationary mechanism of generating fluctuations (see Chapter 6). However, before we move on to discuss what physics can lead to a shrinking Hubble sphere, let me show you that this definition of inflation is equivalent to other popular ways of describing inflation.

- *Accelerated expansion.*—From the relation

$$\frac{d}{dt}(aH)^{-1} = \frac{d}{dt}(\dot{a})^{-1} = -\frac{\ddot{a}}{(\dot{a})^2} \ , \tag{2.2.16}$$

  we see that a shrinking comoving Hubble radius implies accelerated expansion

$$\ddot{a} > 0 \ . \tag{2.2.17}$$

  This explains why inflation is often defined as a period of acceleration.

- *Slowly-varying Hubble parameter.*—Alternatively, we may write

$$\frac{d}{dt}(aH)^{-1} = -\frac{\dot{a}H + a\dot{H}}{(aH)^2} = -\frac{1}{a}(1-\varepsilon) \ , \qquad \text{where} \quad \varepsilon \equiv -\frac{\dot{H}}{H^2} \ . \tag{2.2.18}$$

  The shrinking Hubble sphere therefore also corresponds to

$$\varepsilon = -\frac{\dot{H}}{H^2} < 1 \ . \tag{2.2.19}$$

- *Quasi-de Sitter expansion.*—For perfect inflation, $\varepsilon = 0$, the spacetime becomes de Sitter space

$$\mathrm{d}s^2 = \mathrm{d}t^2 - e^{2Ht}\mathrm{d}\boldsymbol{x}^2 \ , \tag{2.2.20}$$

  where $H = \partial_t \ln a = const.$ Inflation has to end, so it shouldn't correspond to perfect de Sitter space. However, for small, but finite $\varepsilon \neq 0$, the line element (2.2.20) is still a good approximation to the inflationary background. This is why we will often refer to inflation as a quasi-de Sitter period.

- *Negative pressure.*—What forms of stress-energy source accelerated expansion? Let us consider a perfect fluid with pressure P and density ρ. The Friedmann equation, $H^2 = \rho/(3M_{\rm pl}^2)$, and the continuity equation, $\dot{\rho} = -3H(\rho + P)$, together imply

$$\dot{H} + H^2 = -\frac{1}{6M_{\rm pl}^2}(\rho + 3P) = -\frac{H^2}{2}\left(1 + \frac{3P}{\rho}\right) \ . \tag{2.2.21}$$

  We rearrange this to find that

$$\varepsilon = -\frac{\dot{H}}{H^2} = \frac{3}{2}\left(1 + \frac{P}{\rho}\right) < 1 \quad \Leftrightarrow \quad w \equiv \frac{P}{\rho} < -\frac{1}{3} \ , \tag{2.2.22}$$

  i.e. inflation requires negative pressure or a violation of the strong energy condition. How this can arise in a physical theory will be explained in the next section. We will see that there is nothing sacred about the strong energy condition and that it can easily be violated.

- *Constant density.*—Combining the continuitiy equation, $\dot{\rho} = -3H(\rho+P)$, with eq. (2.2.21), we find

$$\left|\frac{d\ln\rho}{d\ln a}\right| = 2\varepsilon < 1 \ . \tag{2.2.23}$$

  For small ε, the energy density is therefore nearly constant. Conventional matter sources all dilute with expansion, so we need to look for something more unusual.

2.3 The Physics of Inflation

We have shown that a given FRW spacetime with time-dependent Hubble parameter $H(t)$ corresponds to cosmic acceleration if and only if

$$\varepsilon \equiv -\frac{\dot{H}}{H^2} = -\frac{d\ln H}{dN} < 1 \ . \tag{2.3.24}$$

Here, we have defined $\mathrm{d}N \equiv \mathrm{d}\ln a = H\mathrm{d}t$, which measures the number of e-folds N of inflationary expansion. Eq. (2.3.24) implies that the fractional change of the Hubble parameter per e-fold is small. Moreover, in order to solve the horizon problem, we want inflation to last for a sufficiently long time (usually at least $N \sim 40$ to 60 e-folds). To achieve this requires ε to remain small for a sufficiently large number of Hubble times. This condition is measured by a second parameter

$$\eta \equiv \frac{d\ln\varepsilon}{dN} = \frac{\dot{\varepsilon}}{H\varepsilon} \ . \tag{2.3.25}$$

For $|\eta| < 1$, the fractional change of ε per Hubble time is small and inflation persists. In this section, we discuss what microscopic physics can lead to the conditions $\varepsilon < 1$ and $|\eta| < 1$.

2.3.1 Scalar Field Dynamics

As a simple toy model for inflation we consider a scalar field, the *inflaton* $\phi(t, \boldsymbol{x})$. As indicated by the notation, the value of the field can depend on time t and the position in space $\boldsymbol{x}$. Associated with each field value is a potential energy density $V(\phi)$ (see fig. 2.5). If the field is dynamical (i.e. changes with time) then it also carries kinetic energy density. If the stress-energy associated with the scalar field dominates the universe, it sources the evolution of the FRW background. We want to determine under which conditions this can lead to accelerated expansion.

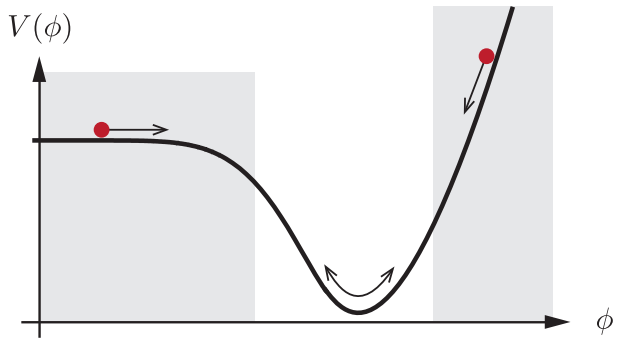


Figure 2.5: *Example of a slow-roll potential. Inflation occurs in the shaded parts of the potential.*

The stress-energy tensor of the scalar field is [4]

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi - V(\phi)\right) . \tag{2.3.26}$$

Consistency with the symmetries of the FRW spacetime requires that the background value of the inflaton only depends on time, $\phi = \phi(t)$. From the time-time component $T^0{}_0 = \rho_\phi$, we infer that

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) . \tag{2.3.27}$$

We see that the total energy density, ρ_ϕ, is simply the sum of the kinetic energy density, $\frac{1}{2}\dot{\phi}^2$, and the potential energy density, $V(\phi)$. From the space-space component $T^i{}_j = -P_\phi\,\delta^i_j$, we find that the pressure is the *difference* of kinetic and potential energy densities,

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) . \tag{2.3.28}$$

We see that a field configuration leads to inflation, $P_\phi < -\frac{1}{3}\rho_\phi$, if the potential energy dominates over the kinetic energy.

Next, we look in more detail at the evolution of the inflaton $\phi(t)$ and the FRW scale factor $a(t)$. Substituting ρ_ϕ from (2.3.27) into the *Friedmann equation*, $H^2 = \rho_\phi/(3M_{\rm pl}^2)$, we get

$$\boxed{H^2 = \frac{1}{3M_{\rm pl}^2}\left[\frac{1}{2}\dot{\phi}^2 + V\right]} . \tag{F}$$

Taking a time derivative, we find

$$2H\dot{H} = \frac{1}{3M_{\rm pl}^2}\left[\dot{\phi}\ddot{\phi} + V'\dot{\phi}\right] , \tag{2.3.29}$$

where $V' \equiv dV/d\phi$. Substituting ρ_ϕ and P_ϕ into the second Friedmann equation (2.2.21), $\dot{H} = -(\rho_\phi + P_\phi)/(2M_{\rm pl}^2)$, we get

$$\dot{H} = -\frac{1}{2}\frac{\dot{\phi}^2}{M_{\rm pl}^2} . \tag{2.3.30}$$

Notice that $\dot{H}$ is sourced by the kinetic energy density. Combining (2.3.30) with (2.3.29) leads to the *Klein-Gordon equation*

$$\boxed{\ddot{\phi} + 3H\dot{\phi} + V' = 0} . \tag{KG}$$

[4] You can derive this stress-energy tensor either from Noether's theorem or from the action of a scalar field. You will see those derivations in the QFT course: David Tong, *Part III Quantum Field Theory*.

This is the evolution equation for the scalar field. Notice that the potential acts like a *force*, V', while the expansion of the universe adds *friction*, $H\dot{\phi}$.

2.3.2 Slow-Roll Inflation

Substituting eq. (2.3.30) into the definition of ε, eq. (2.3.24), we find

$$\varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M_{\rm pl}^2 H^2} \ . \tag{2.3.31}$$

Inflation ($\varepsilon < 1$) therefore occurs if the kinetic energy, $\frac{1}{2}\dot{\phi}^2$, only makes a small contribution to the total energy, $\rho_\phi = 3M_{\rm pl}^2 H^2$. This situation is called *slow-roll inflation*.

In order for this condition to persist, the acceleration of the scalar field has to be small. To assess this, it is useful to define the dimensionless acceleration per Hubble time

$$\delta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}} \ . \tag{2.3.32}$$

Taking the time-derivative of (2.3.31),

$$\dot{\varepsilon} = \frac{\dot{\phi}\ddot{\phi}}{M_{\rm pl}^2 H^2} - \frac{\dot{\phi}^2 \dot{H}}{M_{\rm pl}^2 H^3} \ , \tag{2.3.33}$$

and comparing to (2.3.25), we find

$$\eta = \frac{\dot{\varepsilon}}{H\varepsilon} = 2\frac{\ddot{\phi}}{H\dot{\phi}} - 2\frac{\dot{H}}{H^2} = 2(\varepsilon - \delta) \ . \tag{2.3.34}$$

Hence, $\{\varepsilon, |\delta|\} \ll 1$ implies $\{\varepsilon, |\eta|\} \ll 1$.

Slow-roll approximation.—So far, no approximations have been made. We simply noted that in a regime where $\{\varepsilon, |\delta|\} \ll 1$, inflation occurs and persists. We now use these conditions to simplify the equations of motion. This is called the *slow-roll approximation*. The condition $\varepsilon \ll 1$ implies $\frac{1}{2}\dot{\phi}^2 \ll V$ and hence leads to the following simplification of the Friedmann equation (F),

$$\boxed{H^2 \approx \frac{V}{3M_{\rm pl}^2}} \ . \tag{F$_{\rm SR}$}$$

In the slow-roll approximation, the Hubble expansion is determined completely by the potential energy. The condition $|\delta| \ll 1$ simplifies the Klein-Gordon equation (KG) to

$$\boxed{3H\dot{\phi} \approx -V'} \ . \tag{KG$_{\rm SR}$}$$

This provides a simple relationship between the gradient of the potential and the speed of the inflaton. Substituting (F$_{\rm SR}$) and (KG$_{\rm SR}$) into (2.3.31) gives

$$\varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M_{\rm pl}^2 H^2} \approx \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2 \ . \tag{2.3.35}$$

Furthermore, taking the time-derivative of (KG$_{\rm SR}$),

$$3\dot{H}\dot{\phi} + 3H\ddot{\phi} = -V''\dot{\phi} \ , \tag{2.3.36}$$

leads to

$$\delta + \varepsilon = -\frac{\ddot{\phi}}{H\dot{\phi}} - \frac{\dot{H}}{H^2} \approx M_{\rm pl}^2 \frac{V''}{V} \ . \tag{2.3.37}$$

Hence, a convenient way to assess whether a given potential $V(\phi)$ can lead to slow-roll inflation is to compute the *potential slow-roll parameters*[5]

$$\boxed{\epsilon_{\rm v} \equiv \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2 \quad , \quad |\eta_{\rm v}| \equiv M_{\rm pl}^2 \frac{|V''|}{V}} \ . \tag{2.3.38}$$

Successful slow-roll inflation occurs when these parameters are small, $\{\epsilon_{\rm v}, |\eta_{\rm v}|\} \ll 1$.

Amount of inflation.—The total number of 'e-folds' of accelerated expansion are

$$N_{\rm tot} \equiv \int_{a_I}^{a_E} {\rm d}\ln a = \int_{t_I}^{t_E} H(t)\,{\rm d}t \ , \tag{2.3.39}$$

where t_I and t_E are defined as the times when $\varepsilon(t_I) = \varepsilon(t_E) \equiv 1$. In the slow-roll regime, we can use

$$H{\rm d}t = \frac{H}{\dot{\phi}}\,{\rm d}\phi = \frac{1}{\sqrt{2\varepsilon}}\frac{|{\rm d}\phi|}{M_{\rm pl}} \approx \frac{1}{\sqrt{2\epsilon_{\rm v}}}\frac{|{\rm d}\phi|}{M_{\rm pl}} \tag{2.3.40}$$

to write (2.3.39) as an integral in the field space of the inflaton[6]

$$N_{\rm tot} = \int_{\phi_I}^{\phi_E} \frac{1}{\sqrt{2\epsilon_{\rm v}}}\frac{|{\rm d}\phi|}{M_{\rm pl}} \ , \tag{2.3.41}$$

where ϕ_I and ϕ_E are defined as the boundaries of the interval where $\epsilon_{\rm v} < 1$. The largest scales observed in the CMB are produced about 60 e-folds before the end of inflation

$$N_{\rm cmb} = \int_{\phi_{\rm cmb}}^{\phi_E} \frac{1}{\sqrt{2\epsilon_{\rm v}}}\frac{|{\rm d}\phi|}{M_{\rm pl}} \approx 60 \ . \tag{2.3.42}$$

A successful solution to the horizon problem requires $N_{\rm tot} > N_{\rm cmb}$.

Case study: $m^2\phi^2$ inflation.—As an example, let us give the slow-roll analysis of arguably the simplest model of inflation: single-field inflation driven by a mass term

$$V(\phi) = \frac{1}{2}m^2\phi^2 \ . \tag{2.3.43}$$

The slow-roll parameters are

$$\epsilon_{\rm v}(\phi) = \eta_{\rm v}(\phi) = 2\left(\frac{M_{\rm pl}}{\phi}\right)^2 \ . \tag{2.3.44}$$

To satisfy the slow-roll conditions $\epsilon_{\rm v}, |\eta_{\rm v}| < 1$, we therefore need to consider super-Planckian values for the inflaton

$$\phi > \sqrt{2}M_{\rm pl} \equiv \phi_E \ . \tag{2.3.45}$$

[5] In contrast, the parameters ε and η are often called the *Hubble slow-roll parameters*. During slow-roll, the parameters are related as follows: $\epsilon_{\rm v} \approx \varepsilon$ and $\eta_{\rm v} \approx 2\varepsilon - \frac{1}{2}\eta$.

[6] The absolute value around the integration measure indicates that we pick the overall sign of the integral in such a way as to make $N_{\rm tot} > 0$.

The relation between the inflaton field value and the number of e-folds before the end of inflation is

$$N(\phi) = \frac{\phi^2}{4M_{\rm pl}^2} - \frac{1}{2} \ . \tag{2.3.46}$$

Fluctuations observed in the CMB are created at

$$\phi_{\rm cmb} = 2\sqrt{N_{\rm cmb}}\, M_{\rm pl} \sim 15 M_{\rm pl} \ . \tag{2.3.47}$$

2.3.3 Reheating*

During inflation most of the energy density in the universe is in the form of the inflaton potential $V(\phi)$. Inflation ends when the potential steepens and the inflaton field picks up kinetic energy. The energy in the inflaton sector then has to be transferred to the particles of the Standard Model. This process is called *reheating* and starts the *Hot Big Bang*. We will only have time for a very brief and mostly qualitative description of the absolute basics of the reheating phenomenon.[7] This is non-examinable.

Scalar field oscillations.—After inflation, the inflaton field ϕ begins to oscillate at the bottom of the potential $V(\phi)$, see fig. 2.5. Assume that the potential can be approximated as $V(\phi) = \frac{1}{2}m^2\phi^2$ near the minimum of $V(\phi)$, where the amplitude of ϕ is small. The inflaton is still homogeneous, $\phi(t)$, so its equation of motion is

$$\ddot{\phi} + 3H\dot{\phi} = -m^2\phi \ . \tag{2.3.48}$$

The expansion time scale soon becomes much longer than the oscillation period, $H^{-1} \gg m^{-1}$. We can then neglect the friction term, and the field undergoes oscillations with frequency m. We can write the energy continuity equation as

$$\dot{\rho}_\phi + 3H\rho_\phi = -3HP_\phi = -\frac{3}{2}H(m^2\phi^2 - \dot{\phi}^2) \ . \tag{2.3.49}$$

The r.h.s. averages to zero over one oscillation period. The oscillating field therefore behaves like pressureless matter, with $\rho_\phi \propto a^{-3}$. The fall in the energy density is reflected in a decrease of the oscillation amplitude.

Inflaton decay.—To avoid that the universe ends up empty, the inflaton has to couple to Standard Model fields. The energy stored in the inflaton field will then be transferred into ordinary particles. If the decay is slow (which is the case if the inflaton can only decay into fermions) the inflaton energy density follows the equation

$$\dot{\rho}_\phi + 3H\rho_\phi = -\Gamma_\phi\rho_\phi \ , \tag{2.3.50}$$

where Γ_ϕ parameterizes the inflaton decay rate. If the inflaton can decay into bosons, the decay may be very rapid, involving a mechanism called *parametric resonance* (sourced by Bose condensation effects). This kind of rapid decay is called *preheating*, since the bosons thus created are far from thermal equilibrium.

Thermalisation.—The particles produced by the decay of the inflaton will interact, create other particles through particle reactions, and the resulting particle soup will eventually reach thermal

[7] For more details see Baumann, *The Physics of Inflation*, DAMTP Lecture Notes.

equilibrium with some temperature $T_{\rm rh}$. This reheating temperature is determined by the energy density $\rho_{\rm rh}$ at the end of the reheating epoch. Necessarily, $\rho_{\rm rh} < \rho_{\phi,E}$ (where $\rho_{\phi,E}$ is the inflaton energy density at the end of inflation). If reheating takes a long time, we may have $\rho_{\rm rh} \ll \rho_{\phi,E}$. The evolution of the gas of particles into a thermal state can be quite involved. Usually it is just assumed that it happens eventually, since the particles are able to interact. However, it is possible that some particles (such as gravitinos) never reach thermal equilibrium, since their interactions are so weak. In any case, as long as the momenta of the particles are much higher than their masses, the energy density of the universe behaves like radiation regardless of the momentum space distribution. After thermalisation of at least the baryons, photons and neutrinos is complete, the standard Hot Big Bang era begins.

3 Thermal History

In this chapter, we will describe the first three minutes[1] in the history of the universe, starting from the hot and dense state following inflation. At early times, the thermodynamical properties of the universe were determined by local equilibrium. However, it are the departures from thermal equilibrium that make life interesting. As we will see, non-equilibrium dynamics allows massive particles to acquire cosmological abundances and therefore explains why there is something rather than nothing. Deviations from equilibrium are also crucial for understanding the origin of the cosmic microwave background and the formation of the light chemical elements.

We will start, in §3.1, with a schematic description of the basic principles that shape the thermal history of the universe. This provides an overview of the story that will be fleshed out in much more detail in the rest of the chapter: in §3.2, will present equilibrium thermodynamics in an expanding universe, while in 3.3, we will introduce the Boltzmann equation and apply it to several examples of non-equilibrium physics. We will use units in which Boltzmann's constant is set equal to unity, $k_{\rm B} \equiv 1$, so that temperature has units of energy.

3.1 The Hot Big Bang

The key to understanding the thermal history of the universe is the comparison between the *rate of interactions* Γ and the *rate of expansion* H. When $\Gamma \gg H$, then the time scale of particle interactions is much smaller than the characteristic expansion time scale:

$$t_c \equiv \frac{1}{\Gamma} \ll t_H \equiv \frac{1}{H} \, . \tag{3.1.1}$$

Local thermal equilibrium is then reached before the effect of the expansion becomes relevant. As the universe cools, the rate of interactions may decrease faster than the expansion rate. At $t_c \sim t_H$, the particles decouple from the thermal bath. Different particle species may have different interaction rates and so may decouple at different times.

3.1.1 Local Thermal Equilibrium

Let us first show that the condition (3.1.1) is satisfied for Standard Model processes at temperatures above a few hundred GeV. We write the rate of particle interactions as[2]

$$\Gamma \equiv n \sigma v \, , \tag{3.1.2}$$

where n is the number density of particles, σ is their interaction cross section, and v is the average velocity of the particles. For $T \gtrsim 100$ GeV, all known particles are ultra-relativistic,

[1] A wonderful popular account of this part of cosmology is Weinberg's book *The First Three Minutes*.

[2] For a process of the form $1 + 2 \leftrightarrow 3 + 4$, we would write the interaction rate of species 1 as $\Gamma_1 = n_2 \sigma v$, where n_2 is the density of the target species 2 and v is the average relative velocity of 1 and 2. The interaction rate of species 2 would be $\Gamma_2 = n_1 \sigma v$. We have used the expectation that at high energies $n_1 \sim n_2 \equiv n$.

and hence $v \sim 1$. Since particle masses can be ignored in this limit, the only dimensionful scale is the temperature T. Dimensional analysis then gives $n \sim T^3$. Interactions are mediated by gauge bosons, which are massless above the scale of electroweak symmetry breaking. The cross sections for the strong and electroweak interactions then have a similar dependence, which also can be estimated using dimensional analysis[3]

$$\sigma \sim \left| \quad \right|^2 \sim \frac{\alpha^2}{T^2} \,, \tag{3.1.3}$$

where $\alpha \equiv g_A^2/4\pi$ is the generalized structure constant associated with the gauge boson A. We find that

$$\Gamma = n\sigma v \sim T^3 \times \frac{\alpha^2}{T^2} = \alpha^2 T \,. \tag{3.1.4}$$

We wish to compare this to the Hubble rate $H \sim \sqrt{\rho}/M_{\rm pl}$. The same dimensional argument as before gives $\rho \sim T^4$ and hence

$$H \sim \frac{T^2}{M_{\rm pl}^2} \,. \tag{3.1.5}$$

The ratio of (3.1.4) and (3.1.5) is

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_{\rm pl}}{T} \sim \frac{10^{16}\,{\rm GeV}}{T} \,, \tag{3.1.6}$$

where we have used $\alpha \sim 0.01$ in the numerical estimate. Below $T \sim 10^{16}\,{\rm GeV}$, but above 100 GeV, the condition (3.1.1) is therefore satisfied.

When particles exchange energy and momentum efficiently they reach a state of maximum entropy. It is a standard result of statistical mechanics that the number of particles per unit volume in phase space—the *distribution function*—then takes the form[4]

$$f(E) = \frac{1}{e^{E/T} \pm 1} \,, \tag{3.1.7}$$

where the + sign is for fermions and the − sign for bosons. When the temperature drops below the mass of the particles, $T \ll m$, they become non-relativistic and their distribution function receives an exponential suppression, $f \to e^{-m/T}$. This means that relativistic particles ('radiation') dominate the density and pressure of the primordial plasma. The total energy density is therefore well approximated by summing over all relativistic particles, $\rho_r \propto \sum_i \int {\rm d}^3p\, f_i(p) E_i(p)$. The result can be written as (see below)

$$\rho_r = \frac{\pi^2}{30} g_\star(T) T^4 \,, \tag{3.1.8}$$

where $g_\star(T)$ is the number of relativistic degrees of freedom. Fig. 3.1 shows the evolution of $g_\star(T)$ assuming the particle content of the Standard Model. At early times, all particles are relativistic and $g_\star = 106.75$. The value of $g_\star$ decreases whenever the temperature of the universe drops below the mass of a particle species and it becomes non-relativistic. Today, only photons and (maybe) neutrinos are still relativistic and $g_\star = 3.38$.

[3] Shown in eq. (3.1.3) is the Feynman diagram associated with a $2 \to 2$ scattering process mediated by the exchange of a gauge boson. Each vertex contributes a factor of the gauge coupling $g_A \propto \sqrt{\alpha}$. The dependence of the cross section on α follows from squaring the dependence on α derived from the Feynman diagram, i.e. $\sigma \propto (\sqrt{\alpha} \times \sqrt{\alpha})^2 = \alpha^2$. For more details see the *Part III Standard Model* course.

[4] The precise formula will include the chemical potential – see below.

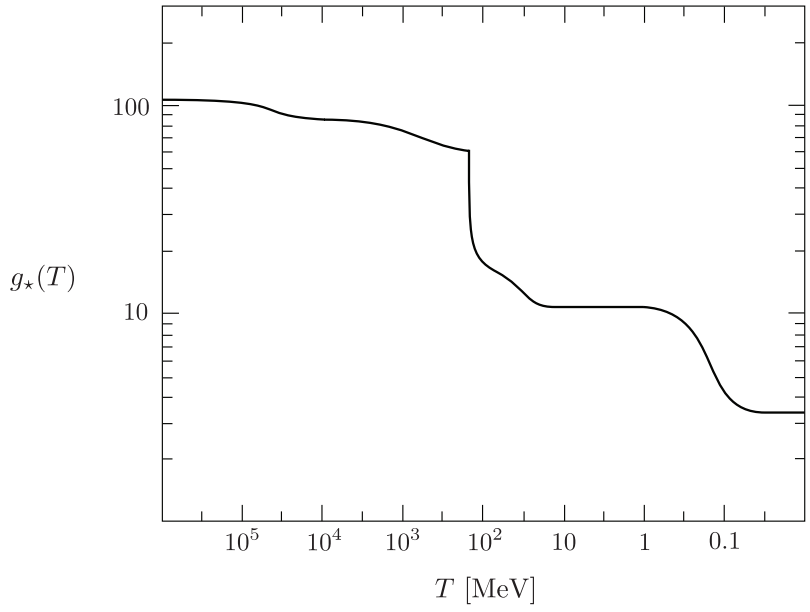


Figure 3.1: *Evolution of the number of relativistic degrees of freedom assuming the Standard Model.*

3.1.2 Decoupling and Freeze-Out

If equilibrium had persisted until today, the universe would be mostly photons. Any massive particle species would be exponentially suppressed.[5] To understand the world around us, it is therefore crucial to understand the deviations from equilibrium that led to the *freeze-out* of massive particles (see fig. 3.2).

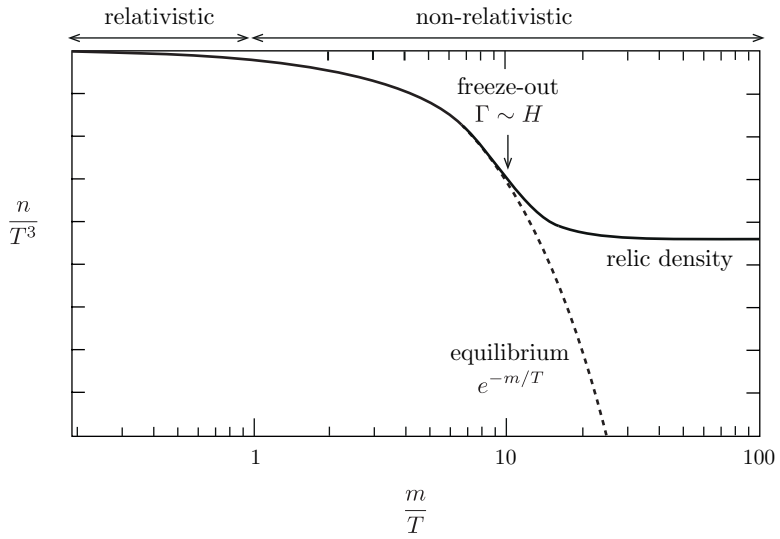


Figure 3.2: *A schematic illustration of particle freeze-out. At high temperatures, $T \gg m$, the particle abundance tracks its equilibrium value. At low temperatures, $T \ll m$, the particles freeze out and maintain a density that is much larger than the Boltzmann-suppressed equilibrium abundance.*

Below the scale of electroweak symmetry breaking, $T \lesssim 100$ GeV, the gauge bosons of the weak interactions, $W^\pm$ and Z, receive masses $M_W \sim M_Z$. The cross section associated with

[5] This isn't quite correct for baryons. Since baryon number is a symmetry of the Standard Model, the number density of baryons can remain significant even in equilibrium.

processes mediated by the weak force becomes

$$\sigma \sim \left| \quad \right|^2 \sim G_F^2 T^2 \ , \tag{3.1.9}$$

where we have introduced Fermi's constant,[6] $G_F \sim \alpha/M_W^2 \sim 1.17 \times 10^{-5}$ GeV^{-2}. Notice that the strength of the weak interactions now decreases as the temperature of the universe drops. We find that

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_{\rm pl} T^3}{M_W^4} \sim \left(\frac{T}{1 \text{ MeV}}\right)^3 \ , \tag{3.1.10}$$

which drops below unity at $T_{dec} \sim 1$ MeV. Particles that interact with the primordial plasma only through the weak interaction therefore *decouple* around 1 MeV. This decoupling of weak scale interactions has important consequences for the thermal history of the universe.

3.1.3 A Brief History of the Universe

Table 3.1 lists the key events in the thermal history of the universe:

- **Baryogenesis.*** Relativistic quantum field theory requires the existence of anti-particles (see *Part III Quantum Field Theory*). This poses a slight puzzle. Particles and anti-particles annihilate through processes such as $e^+ + e^- \to \gamma + \gamma$. If initially the universe was filled with equal amounts of matter and anti-matter then we expect these annihilations to lead to a universe dominated by radiation. However, we do observe an overabundance of matter (mostly baryons) over anti-matter in the universe today. Models of *baryogenesis* try to derive the observed baryon-to-photon ratio

  $$\eta \equiv \frac{n_b}{n_\gamma} \sim 10^{-9} \ , \tag{3.1.11}$$

  from some dynamical mechanism, i.e. without assuming a primordial matter-antimatter asymmetry as an initial condition. Although many ideas for baryogenesis exist, none is singled out by experimental tests. We will not have much to say about baryogenesis in this course.

- **Electroweak phase transition.** At 100 GeV particles receive their masses through the Higgs mechanism. Above we have seen how this leads to a drastic change in the strength of the weak interaction.

- **QCD phase transition.** While quarks are *asymptotically free* (i.e. weakly interacting) at high energies, below 150 MeV, the strong interactions between the quarks and the gluons become important. Quarks and gluons then form bound three-quark systems, called *baryons*, and quark-antiquark pairs, called *mesons*. These baryons and mesons are the relevant degrees of freedom below the scale of the QCD phase transition.

- **Dark matter freeze-out.** Since dark matter is very weakly interacting with ordinary matter we expect it to decouple relatively early on. In §3.3.2, we will study the example of WIMPs—weakly interacting massive particles that freeze out around 1 MeV. We will

[6] The $1/M_W^2$ comes from the low-momentum limit of the propagator of a massive gauge field.

Event	time t	redshift z	temperature T
Inflation	10^{-34} s (?)	–	–
Baryogenesis	?	?	?
EW phase transition	20 ps	10^{15}	100 GeV
QCD phase transition	20 μs	10^{12}	150 MeV
Dark matter freeze-out	?	?	?
Neutrino decoupling	1 s	6×10^9	1 MeV
Electron-positron annihilation	6 s	2×10^9	500 keV
Big Bang nucleosynthesis	3 min	4×10^8	100 keV
Matter-radiation equality	60 kyr	3400	0.75 eV
Recombination	260–380 kyr	1100–1400	0.26–0.33 eV
Photon decoupling	380 kyr	1000–1200	0.23–0.28 eV
Reionization	100–400 Myr	11–30	2.6–7.0 meV
Dark energy-matter equality	9 Gyr	0.4	0.33 meV
Present	13.8 Gyr	0	0.24 meV

Table 3.1: *Key events in the thermal history of the universe.*

show that choosing natural values for the mass of the dark matter particles and their interaction cross section with ordinary matter reproduces the observed relic dark matter density surprisingly well.

- **Neutrino decoupling.** Neutrinos only interact with the rest of the primordial plasma through the weak interaction. The estimate in (3.1.10) therefore applies and neutrinos decouple at 0.8 MeV.

- **Electron-positron annihilation.** Electrons and positrons annihilate shortly after neutrino decoupling. The energies of the electrons and positrons gets transferred to the photons, but not the neutrinos. In §3.2.4, we will explain that this is the reason why the photon temperature today is greater than the neutrino temperature.

- **Big Bang nucleosynthesis.** Around 3 minutes after the Big Bang, the light elements were formed. In §3.3.4, we will study this process of *Big Bang nucleosynthesis* (BBN).

- **Recombination.** Neutral hydrogen forms through the reaction $e^- + p^+ \to \mathrm{H} + \gamma$ when the temperature has become low enough that the reverse reaction is energetically disfavoured. We will study *recombination* in §3.3.3.

- **Photon decoupling.** Before recombination the strongest coupling between the photons and the rest of the plasma is through Thomson scattering, $e^- + \gamma \to e^- + \gamma$. The sharp drop in the free electron density after recombination means that this process becomes inefficient and the photons decouple. They have since streamed freely through the universe and are today observed as the *cosmic microwave background* (CMB).

In the rest of this chapter we will explore in detail where this knowledge about the thermal history of the universe comes from.

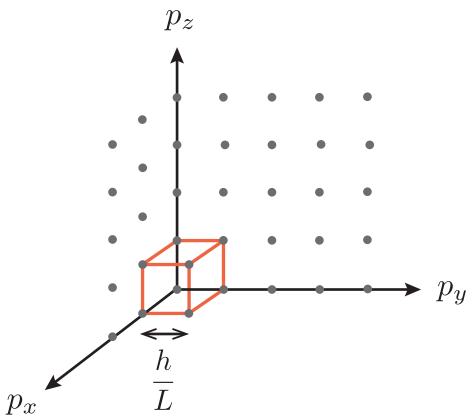
3.2 Equilibrium

3.2.1 Equilibrium Thermodynamics

We have good observational evidence (from the perfect blackbody spectrum of the CMB) that the early universe was in *local thermal equilibrium*.[7] Moreover, we have seen above that the Standard Model predicts thermal equilibrium above 100 GeV. To describe this state and the subsequent evolution of the universe, we need to recall some basic facts of equilibrium thermodynamics, suitably generalized to apply to an expanding universe.

Microscopic to Macroscopic

Statistical mechanics is the art of turning microscopic laws into an understanding of the macroscopic world. I will briefly review this approach for a gas of weakly interacting particles. It is convenient to describe the system in *phase space*, where the gas is described by the positions and momenta of all particles. In quantum mechanics, the momentum eigenstates of a particle in a volume $V = L^3$ have a discrete spectrum:



The density of states in momentum space $\{\boldsymbol{p}\}$ then is $L^3/h^3 = V/h^3$, and the state density in phase space $\{\boldsymbol{x}, \boldsymbol{p}\}$ is

$$\frac{1}{h^3} \,. \tag{3.2.12}$$

If the particle has g internal degrees of freedom (e.g. spin), then the density of states becomes

$$\frac{g}{h^3} = \frac{g}{(2\pi)^3} \,, \tag{3.2.13}$$

[7] Strictly speaking, the universe can never truly be in equilibrium since the FRW spacetime doesn't posses a time-like Killing vector. But this is physics not mathematics: if the expansion is slow enough, particles have enough time to settle close to local equilibrium. (And since the universe is homogeneous, the local values of thermodynamics quantities are also global values.)

where in the second equality we have used natural units with $\hbar = h/(2\pi) \equiv 1$. To obtain the number density of a gas of particles we need to know how the particles are distributed amongst the momentum eigenstates. This information is contained in the *(phase space) distribution function* $f(\boldsymbol{x}, \boldsymbol{p}, t)$. Because of homogeneity, the distribution function should, in fact, be independent of the position $\boldsymbol{x}$. Moreover, isotropy requires that the momentum dependence is only in terms of the magnitude of the momentum $p \equiv |\boldsymbol{p}|$. We will typically leave the time dependence implicit—it will manifest itself in terms of the temperature dependence of the distribution functions. The particle density in phase space is then the density of states times the distribution function

$$\frac{g}{(2\pi)^3} \times f(p) \ . \tag{3.2.14}$$

The *number density* of particles (in real space) is found by integrating (3.2.14) over momentum,

$$\boxed{n \ = \ \frac{g}{(2\pi)^3} \int \mathrm{d}^3p \, f(p)} \ . \tag{3.2.15}$$

To obtain the energy density of the gas of particles, we have to weight each momentum eigenstate by its energy. To a good approximation, the particles in the early universe were *weakly interacting*. This allows us to ignore the interaction energies between the particles and write the energy of a particle of mass m and momentum p simply as

$$E(p) = \sqrt{m^2 + p^2} \ . \tag{3.2.16}$$

Integrating the product of (3.2.16) and (3.2.14) over momentum then gives the *energy density*

$$\boxed{\rho \ = \ \frac{g}{(2\pi)^3} \int \mathrm{d}^3p \, f(p) \, E(p)} \ . \tag{3.2.17}$$

Similarly, we define the *pressure* as

$$\boxed{P \ = \ \frac{g}{(2\pi)^3} \int \mathrm{d}^3p \, f(p) \, \frac{p^2}{3E}} \ . \tag{3.2.18}$$

*Pressure.**—Let me remind you where the $p^2/3E$ factor in (3.2.18) comes from. Consider a small area element of size $\mathrm{d}A$, with unit normal vector $\hat{\boldsymbol{n}}$ (see fig. 3.3). All particles with velocity $|\boldsymbol{v}|$, striking this area element in the time interval between t and $t + \mathrm{d}t$, were located at $t = 0$ in a spherical shell of radius $R = |\boldsymbol{v}|t$ and width $|\boldsymbol{v}|\mathrm{d}t$. A solid angle $\mathrm{d}\Omega^2$ of this shell defines the volume $\mathrm{d}V = R^2|\boldsymbol{v}|\mathrm{d}t \, \mathrm{d}\Omega^2$ (see the grey shaded region in fig. 3.3). Multiplying the phase space density (3.2.14) by $\mathrm{d}V$ gives the number of particles in the volume (per unit volume in momentum space) with energy $E(|\boldsymbol{v}|)$,

$$\mathrm{d}N = \frac{g}{(2\pi)^3} f(E) \times R^2|\boldsymbol{v}|\mathrm{d}t \, \mathrm{d}\Omega \ . \tag{3.2.19}$$

Not all particles in $\mathrm{d}V$ reach the target, only those with velocities directed to the area element. Taking into account the isotropy of the velocity distribution, we find that the total number of particles striking the area element $\mathrm{d}A \, \hat{\boldsymbol{n}}$ with velocity $\boldsymbol{v} = |\boldsymbol{v}| \, \hat{\boldsymbol{v}}$ is

$$\mathrm{d}N_A = \frac{|\hat{\boldsymbol{v}} \cdot \hat{\boldsymbol{n}}| \, \mathrm{d}A}{4\pi R^2} \times \mathrm{d}N = \frac{g}{(2\pi)^3} f(E) \times \frac{|\boldsymbol{v} \cdot \hat{\boldsymbol{n}}|}{4\pi} \, \mathrm{d}A \, \mathrm{d}t \, \mathrm{d}\Omega \ , \tag{3.2.20}$$

where $\boldsymbol{v}\cdot\hat{\boldsymbol{n}} < 0$. If these particles are reflected elastically, each transfer momentum $2|\boldsymbol{p}\cdot\hat{\boldsymbol{n}}|$ to the target. Therefore, the contribution of particles with velocity $|\boldsymbol{v}|$ to the pressure is

$$\mathrm{d}P(|\boldsymbol{v}|) = \int \frac{2|\boldsymbol{p}\cdot\hat{\boldsymbol{n}}|}{\mathrm{d}A\,\mathrm{d}t}\,\mathrm{d}N_A = \frac{g}{(2\pi)^3} f(E) \times \frac{p^2}{2\pi E}\int \cos^2\theta\,\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi = \frac{g}{(2\pi)^3}\times f(E)\,\frac{p^2}{3E}\ , \quad (3.2.21)$$

where we have used $|\boldsymbol{v}| = |\boldsymbol{p}|/E$ and integrated over the hemisphere defined by $\hat{\boldsymbol{v}}\cdot\hat{\boldsymbol{n}} \equiv -\cos\theta < 0$ (i.e. integrating only over particles moving towards $\mathrm{d}A$—see fig. 3.3). Integrating over energy E (or momentum p), we obtain (3.2.18).

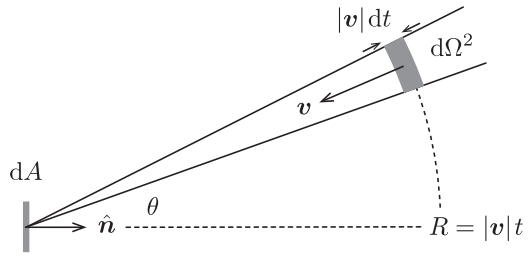


Figure 3.3: *Pressure in a weakly interacting gas of particles.*

Local Thermal Equilibrium

A system of particles is said to be in *kinetic equilibrium* if the particles exchange energy and momentum efficiently. This leads to a state of maximum entropy in which the distribution functions are given by the *Fermi-Dirac* and *Bose-Einstein* distributions[8]

$$\boxed{f(p) = \frac{1}{e^{(E(p)-\mu)/T}\pm 1}}\ , \quad (3.2.22)$$

where the + sign is for fermions and the − sign for bosons. At low temperatures, $T < E - \mu$, both distribution functions reduce to the *Maxwell-Boltzmann* distribution

$$f(p) \approx e^{-(E(p)-\mu)/T}\ . \quad (3.2.23)$$

The equilibrium distribution functions have two parameters: the *temperature* T and the *chemical potential* μ. The chemical potential may be temperature-dependent. As the universe expands, T and $\mu(T)$ change in such a way that the continuity equations for the energy density ρ and the particle number density n are satisfied. Each particle species i (with possibly distinct m_i, μ_i, T_i) has its own distribution function f_i and hence its own n_i, ρ_i, and P_i.

*Chemical potential.**—In thermodynamics, the chemical potential characterizes the response of a system to a change in particle number. Specifically, it is defined as the derivative of the entropy with respect to the number of particles, at fixed energy and fixed volume,

$$\mu = -T\left(\frac{\partial S}{\partial N}\right)_{U,V}\ . \quad (3.2.24)$$

[8] We use units where Boltzmann's constant is $k_{\mathrm{B}} \equiv 1$.

The change in entropy of a system therefore is

$$\mathrm{d}S = \frac{\mathrm{d}U + P\mathrm{d}V - \mu\mathrm{d}N}{T} \ , \tag{3.2.25}$$

where $\mu \mathrm{d}N$ is sometimes called the *chemical work*. A knowledge of the chemical potential of reacting particles can be used to indicate which way a reaction proceeds. The second law of thermodynamics means that particles flow to the side of the reaction with the lower total chemical potential. *Chemical equilibrium* is reached when the sum of the chemical potentials of the reacting particles is equal to the sum of the chemical potentials of the products. The rates of the forward and reverse reactions are then equal.

If a species i is in *chemical equilibrium*, then its chemical potential μ_i is related to the chemical potentials μ_j of the other species it interacts with. For example, if a species 1 interacts with species 2, 3 and 4 via the reaction $1 + 2 \leftrightarrow 3 + 4$, then chemical equilibrium implies

$$\mu_1 + \mu_2 = \mu_3 + \mu_4 \ . \tag{3.2.26}$$

Since the number of photons is not conserved (e.g. double Compton scattering $e^- + \gamma \leftrightarrow e^- + \gamma + \gamma$ happens in equilibrium at high temperatures), we know that

$$\mu_\gamma = 0 \ . \tag{3.2.27}$$

This implies that if the chemical potential of a particle X is μ_X, then the chemical potential of the corresponding anti-particle $\bar{X}$ is

$$\mu_{\bar{X}} = -\mu_X \ , \tag{3.2.28}$$

To see this, just consider particle-antiparticle annihilation, $X + \bar{X} \leftrightarrow \gamma + \gamma$.

Thermal equilibrium is achieved for species which are both in kinetic and chemical equilibrium. These species then share a common temperature $T_i = T$.[9]

3.2.2 Densities and Pressure

Let us now use the results from the previous section to relate the densities and pressure of a gas of weakly interacting particles to the temperature of the universe.

At early times, the chemical potentials of all particles are so small that they can be neglected.[10] Setting the chemical potential to zero, we get

$$n = \frac{g}{2\pi^2} \int_0^\infty \mathrm{d}p \, \frac{p^2}{\exp\left[\sqrt{p^2 + m^2}/T\right] \pm 1} \ , \tag{3.2.29}$$

$$\rho = \frac{g}{2\pi^2} \int_0^\infty \mathrm{d}p \, \frac{p^2 \sqrt{p^2 + m^2}}{\exp\left[\sqrt{p^2 + m^2}/T\right] \pm 1} \ . \tag{3.2.30}$$

[9] This temperature is often identified with the photon temperature T_γ — the "temperature of the universe".

[10] For electrons and protons this is a fact (see Problem Set 2), while for neutrinos it is likely true, but not proven.

Defining $x \equiv m/T$ and $\xi \equiv p/T$, this can be written as

$$n = \frac{g}{2\pi^2} T^3 \, I_\pm(x) \ , \qquad I_\pm(x) \equiv \int_0^\infty \mathrm{d}\xi \, \frac{\xi^2}{\exp\left[\sqrt{\xi^2 + x^2}\right] \pm 1} \ , \tag{3.2.31}$$

$$\rho = \frac{g}{2\pi^2} T^4 \, J_\pm(x) \ , \qquad J_\pm(x) \equiv \int_0^\infty \mathrm{d}\xi \, \frac{\xi^2 \sqrt{\xi^2 + x^2}}{\exp\left[\sqrt{\xi^2 + x^2}\right] \pm 1} \ . \tag{3.2.32}$$

In general, the functions $I_\pm(x)$ and $J_\pm(x)$ have to be evaluated numerically. However, in the (ultra)relativistic and non-relativistic limits, we can get analytical results.

The following standard integrals will be useful

$$\int_0^\infty \mathrm{d}\xi \, \frac{\xi^n}{e^\xi - 1} = \zeta(n+1)\,\Gamma(n+1) \ , \tag{3.2.33}$$

$$\int_0^\infty \mathrm{d}\xi \, \xi^n e^{-\xi^2} = \tfrac{1}{2}\,\Gamma\left(\tfrac{1}{2}(n+1)\right) \ , \tag{3.2.34}$$

where $\zeta(z)$ is the Riemann zeta-function.

Relativistic Limit

In the limit $x \to 0$ ($m \ll T$), the integral in (3.2.31) reduces to

$$I_\pm(0) = \int_0^\infty \mathrm{d}\xi \, \frac{\xi^2}{e^\xi \pm 1} \ . \tag{3.2.35}$$

For bosons, this takes the form of the integral (3.2.33) with $n = 2$,

$$I_-(0) = 2\zeta(3) \ , \tag{3.2.36}$$

where $\zeta(3) \approx 1.20205\cdots$. To find the corresponding result for fermions, we note that

$$\frac{1}{e^\xi + 1} = \frac{1}{e^\xi - 1} - \frac{2}{e^{2\xi} - 1} \ , \tag{3.2.37}$$

so that

$$I_+(0) = I_-(0) - 2 \times \left(\frac{1}{2}\right)^3 I_-(0) = \frac{3}{4}\, I_-(0) \ . \tag{3.2.38}$$

Hence, we get

$$\boxed{n = \frac{\zeta(3)}{\pi^2}\, g T^3 \begin{cases} 1 & \text{bosons} \\ \frac{3}{4} & \text{fermions} \end{cases}} \ . \tag{3.2.39}$$

A similar computation for the energy density gives

$$\boxed{\rho = \frac{\pi^2}{30}\, g T^4 \begin{cases} 1 & \text{bosons} \\ \frac{7}{8} & \text{fermions} \end{cases}} \ . \tag{3.2.40}$$

Relic photons.—Using that the temperature of the cosmic microwave background is $T_0 = 2.73$ K, show that

$$n_{\gamma,0} = \frac{2\zeta(3)}{\pi^2} T_0^3 \ \approx \ 410 \text{ photons cm}^{-3} \ , \tag{3.2.41}$$

$$\rho_{\gamma,0} = \frac{\pi^2}{15} T_0^4 \ \approx \ 4.6 \times 10^{-34} \text{g cm}^{-3} \qquad \Rightarrow \qquad \Omega_\gamma h^2 \approx 2.5 \times 10^{-5} \ . \tag{3.2.42}$$

Finally, from (3.2.18), it is easy to see that we recover the expected pressure-density relation for a relativistic gas (i.e. 'radiation')

$$P = \frac{1}{3}\rho \ . \tag{3.2.43}$$

*Exercise.**—For $\mu = 0$, the numbers of particles and anti-particles are equal. To find the "net particle number" let us restore finite μ in the relativistic limit. For fermions with $\mu \neq 0$ and $T \gg m$, show that

$$\begin{aligned} n - \bar{n} &= \frac{g}{2\pi^2}\int_0^\infty \mathrm{d}p\, p^2 \left(\frac{1}{e^{(p-\mu)/T}+1} - \frac{1}{e^{(p+\mu)/T}+1}\right) \\ &= \frac{1}{6\pi^2}\, gT^3 \left[\pi^2\left(\frac{\mu}{T}\right) + \left(\frac{\mu}{T}\right)^3\right] \ . \end{aligned} \tag{3.2.44}$$

Note that this result is exact and not a truncated series.

Non-Relativistic Limit

In the limit $x \gg 1$ ($m \gg T$), the integral (3.2.31) is the same for bosons and fermions

$$I_\pm(x) \approx \int_0^\infty \mathrm{d}\xi\, \frac{\xi^2}{e^{\sqrt{\xi^2+x^2}}} \ . \tag{3.2.45}$$

Most of the contribution to the integral comes from $\xi \ll x$. We can therefore Taylor expand the square root in the exponential to lowest order in ξ,

$$I_\pm(x) \approx \int_0^\infty \mathrm{d}\xi\, \frac{\xi^2}{e^{x+\xi^2/(2x)}} = e^{-x}\int_0^\infty \mathrm{d}\xi\, \xi^2 e^{-\xi^2/(2x)} = (2x)^{3/2}e^{-x}\int_0^\infty \mathrm{d}\xi\, \xi^2 e^{-\xi^2} \ . \tag{3.2.46}$$

The last integral is of the form of the integral (3.2.34) with $n = 2$. Using $\Gamma(\frac{3}{2}) = \sqrt{\pi}/2$, we get

$$I_\pm(x) = \sqrt{\frac{\pi}{2}}\, x^{3/2} e^{-x} \ , \tag{3.2.47}$$

which leads to

$$\boxed{n = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T}} \ . \tag{3.2.48}$$

As expected, massive particles are exponentially rare at low temperatures, $T \ll m$. At lowest order in the non-relativistic limit, we have $E(p) \approx m$ and the energy density is simply equal to the mass density

$$\rho \approx mn \ . \tag{3.2.49}$$

Exercise.—Using $E(p) = \sqrt{m^2+p^2} \approx m + p^2/2m$, show that

$$\rho = mn + \frac{3}{2}nT \ . \tag{3.2.50}$$

Finally, from (3.2.18), it is easy to show that a non-relativistic gas of particles acts like pressureless dust (i.e. 'matter')

$$P = nT \ \ll \ \rho = mn \ . \tag{3.2.51}$$

Exercise.—Derive (3.2.51). Notice that this is nothing but the ideal gas law, $PV = Nk_{\rm B}T$.

By comparing the relativistic limit ($T \gg m$) and the non-relativistic limit ($T \ll m$), we see that the number density, energy density, and pressure of a particle species fall exponentially (are "Boltzmann suppressed") as the temperature drops below the mass of the particle. We interpret this as the annihilation of particles and anti-particles. At higher energies these annihilations also occur, but they are balanced by particle-antiparticle pair production. At low temperatures, the thermal particle energies aren't sufficient for pair production.

Exercise.—Restoring finite μ in the non-relativistic limit, show that

$$n = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{-(m-\mu)/T} , \tag{3.2.52}$$

$$n - \bar{n} = 2g\left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T} \sinh\left(\frac{\mu}{T}\right) . \tag{3.2.53}$$

Effective Number of Relativistic Species

Let T be the temperature of the photon gas. The total radiation density is the sum over the energy densities of all relativistic species

$$\boxed{\rho_r = \sum_i \rho_i = \frac{\pi^2}{30} g_\star(T) T^4} , \tag{3.2.54}$$

where $g_\star(T)$ is the *effective number of relativistic degrees of freedom* at the temperature T. The sum over particle species may receive two types of contributions:

- Relativistic species in thermal equilibrium with the photons, $T_i = T \gg m_i$,

$$g_\star^{th}(T) = \sum_{i=b} g_i + \frac{7}{8} \sum_{i=f} g_i . \tag{3.2.55}$$

  When the temperature drops below the mass m_i of a particle species, it becomes non-relativistic and is removed from the sum in (3.2.55). Away from mass thresholds, the thermal contribution is independent of temperature.

- Relativistic species that are not in thermal equilibrium with the photons, $T_i \neq T \gg m_i$,

$$g_\star^{dec}(T) = \sum_{i=b} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{i=f} g_i \left(\frac{T_i}{T}\right)^4 . \tag{3.2.56}$$

  We have allowed for the decoupled species to have different temperatures T_i. This will be relevant for neutrinos after e^+e^- annihilation (see §3.2.4).

Fig. 3.4 shows the evolution of $g_\star(T)$ assuming the Standard Model particle content (see table 3.2). At $T \gtrsim 100$ GeV, all particles of the Standard Model are relativistic. Adding up

Table 3.2: *Particle content of the Standard Model.*

type		mass	spin	g
quarks	$t, \bar{t}$	173 GeV	$\frac{1}{2}$	$2 \cdot 2 \cdot 3 = 12$
	$b, \bar{b}$	4 GeV		
	$c, \bar{c}$	1 GeV		
	$s, \bar{s}$	100 MeV		
	$d, \bar{s}$	5 MeV		
	$u, \bar{u}$	2 MeV		
gluons	g_i	0	1	$8 \cdot 2 = 16$
leptons	$\tau^{\pm}$	1777 MeV	$\frac{1}{2}$	$2 \cdot 2 = 4$
	$\mu^{\pm}$	106 MeV		
	$e^{\pm}$	511 keV		
	$\nu_\tau, \bar{\nu}_\tau$	< 0.6 eV	$\frac{1}{2}$	$2 \cdot 1 = 2$
	$\nu_\mu, \bar{\nu}_\mu$	< 0.6 eV		
	$\nu_e, \bar{\nu}_e$	< 0.6 eV		
gauge bosons	W^+	80 GeV	1	3
	W^-	80 GeV		
	Z^0	91 GeV		
	γ	0		2
Higgs boson	H^0	125 GeV	0	1

their internal degrees of freedom we get:[11]

$$\begin{aligned} g_b &= 28 \qquad \text{photons}\,(2),\ W^{\pm} \text{ and } Z^0\,(3 \cdot 3),\ \text{gluons}\,(8 \cdot 2),\ \text{and Higgs}\,(1) \\ g_f &= 90 \qquad \text{quarks}\,(6 \cdot 12),\ \text{charged leptons}\,(3 \cdot 4),\ \text{and neutrinos}\,(3 \cdot 2) \end{aligned}$$

and hence

$$g_\star = g_b + \frac{7}{8} g_f \ = \ 106.75 \ . \tag{3.2.57}$$

As the temperature drops, various particle species become non-relativistic and annihilate. To estimate $g_\star$ at a temperature T we simply add up the contributions from all relativistic degrees of freedom (with $m \ll T$) and discard the rest.

Being the heaviest particles of the Standard Model, the top quarks annihilates first. At $T \sim \frac{1}{6} m_t \sim 30$ GeV,[12] the effective number of relativistic species is reduced to $g_\star = 106.75 -$

[11] Here, we have used that massless spin-1 particles (photons and gluons) have two polarizations, massive spin-1 particles ($W^{\pm}, Z$) have three polarizations and massive spin-$\frac{1}{2}$ particles ($e^{\pm}, \mu^{\pm}, \tau^{\pm}$ and quarks) have two spin states. We assumed that the neutrinos are purely left-handed (i.e. we only counted one helicity state). Also, remember that fermions have anti-particles.

[12] The transition from relativistic to non-relativistic behaviour isn't instantaneous. About 80% of the particle-antiparticle annihilations takes place in the interval $T = m \to \frac{1}{6} m$.

$\frac{7}{8} \times 12 = 96.25$. The Higgs boson and the gauge bosons $W^{\pm}$, Z^0 annihilate next. This happens roughly at the same time. At $T \sim 10$ GeV, we have $g_\star = 96.26 - (1 + 3 \cdot 3) = 86.25$. Next, the bottom quarks annihilate ($g_\star = 86.25 - \frac{7}{8} \times 12 = 75.75$), followed by the charm quarks and the tau leptons ($g_\star = 75.75 - \frac{7}{8} \times (12 + 4) = 61.75$). Before the strange quarks had time to annihilate, something else happens: matter undergoes the QCD phase transition. At $T \sim 150$ MeV, the quarks combine into baryons (protons, neutrons, ...) and mesons (pions, ...). There are many different species of baryons and mesons, but all except the pions ($\pi^{\pm}, \pi^0$) are non-relativistic below the temperature of the QCD phase transition. Thus, the only particle species left in large numbers are the pions, electrons, muons, neutrinos, and the photons. The three pions (spin-0) correspond to $g = 3 \cdot 1 = 3$ internal degrees of freedom. We therefore get $g_\star = 2 + 3 + \frac{7}{8} \times (4 + 4 + 6) = 17.25$. Next electrons and positrons annihilate. However, to understand this process we first need to talk about entropy.

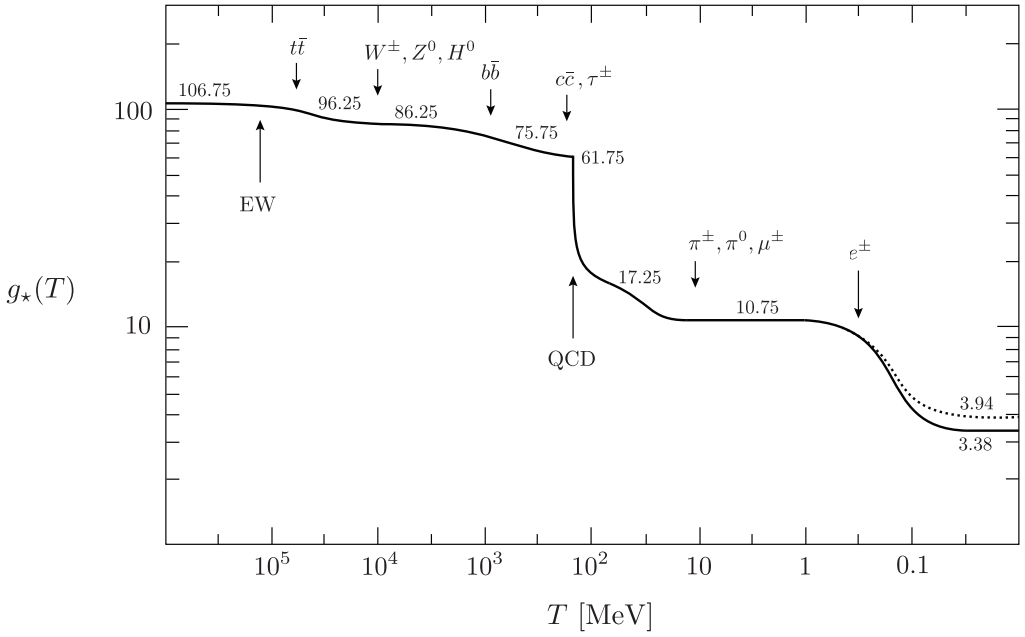


Figure 3.4: *Evolution of relativistic degrees of freedom $g_\star(T)$ assuming the Standard Model particle content. The dotted line stands for the number of effective degrees of freedom in entropy $g_{\star S}(T)$.*

3.2.3 Conservation of Entropy

To describe the evolution of the universe it is useful to track a conserved quantity. As we will see, in cosmology entropy is more informative than energy. According to the second law of thermodynamics, the total entropy of the universe only increases or stays constant. It is easy to show that the entropy is conserved in equilibrium (see below). Since there are far more photons than baryons in the universe, the entropy of the universe is dominated by the entropy of the photon bath (at least as long as the universe is sufficiently uniform). Any entropy production from non-equilibrium processes is therefore total insignificant relative to the total entropy. To a good approximation we can therefore treat the expansion of the universe as *adiabatic*, so that the total entropy stays constant even beyond equilibrium.

Exercise.—Show that the following holds for particles in equilibrium (which therefore have the corresponding distribution functions) and $\mu = 0$:

$$\boxed{\frac{\partial P}{\partial T} = \frac{\rho + P}{T}} \,. \tag{3.2.58}$$

Consider the second law of thermodynamics: $T\mathrm{d}S = \mathrm{d}U + P\mathrm{d}V$. Using $U = \rho V$, we get

$$\begin{aligned} \mathrm{d}S &= \frac{1}{T}\Big(\mathrm{d}\big[(\rho + P)V\big] - V\mathrm{d}P\Big) \\ &= \frac{1}{T}\,\mathrm{d}\big[(\rho + P)V\big] - \frac{V}{T^2}(\rho + P)\,\mathrm{d}T \\ &= \mathrm{d}\left[\frac{\rho + P}{T}\,V\right] \,, \end{aligned} \tag{3.2.59}$$

where we have used (3.2.58) in the second line. To show that entropy is conserved in equilibrium, we consider

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\rho + P}{T}\,V\right] \\ &= \frac{V}{T}\left[\frac{\mathrm{d}\rho}{\mathrm{d}t} + \frac{1}{V}\frac{\mathrm{d}V}{\mathrm{d}t}(\rho + P)\right] + \frac{V}{T}\left[\frac{\mathrm{d}P}{\mathrm{d}t} - \frac{\rho + P}{T}\frac{\mathrm{d}T}{\mathrm{d}t}\right] \,. \end{aligned} \tag{3.2.60}$$

The first term vanishes by the continuity equation, $\dot\rho + 3H(\rho + P) = 0$. (Recall that $V \propto a^3$.) The second term vanishes by (3.2.58). This established the conservation of entropy in equilibrium.

In the following, it will be convenient to work with the *entropy density*, $s \equiv S/V$. From (3.2.59), we learn that

$$s = \frac{\rho + P}{T} \,. \tag{3.2.61}$$

Using (3.2.40) and (3.2.51), the total entropy density for a collection of different particle species is

$$s = \sum_i \frac{\rho_i + P_i}{T_i} \equiv \frac{2\pi^2}{45} g_{\star S}(T) T^3 \,, \tag{3.2.62}$$

where we have defined the *effective number of degrees of freedom in entropy*,

$$g_{\star S}(T) = g_{\star S}^{th}(T) + g_{\star S}^{dec}(T) \,. \tag{3.2.63}$$

Note that for species in thermal equilibrium $g_{\star S}^{th}(T) = g_\star^{th}(T)$. However, given that $s_i \propto T_i^3$, for decoupled species we get

$$g_{\star S}^{dec}(T) \;\equiv\; \sum_{i=b} g_i \left(\frac{T_i}{T}\right)^3 \;+\; \frac{7}{8}\sum_{i=f} g_i \left(\frac{T_i}{T}\right)^3 \;\neq\; g_\star^{dec}(T) \,. \tag{3.2.64}$$

Hence, $g_{\star S}$ is equal to $g_\star$ only when *all* the relativistic species are in equilibrium at the same temperature. In the real universe, this is the case until $t \approx 1$ sec (cf. fig. 3.4).

The conservation of entropy has two important consequences:

- It implies that $s \propto a^{-3}$. The number of particles in a comoving volume is therefore proportional to the number density n_i divided by the entropy density

$$N_i \equiv \frac{n_i}{s} \ . \tag{3.2.65}$$

If particles are neither produced nor destroyed, then $n_i \propto a^{-3}$ and N_i is constant. This is case, for example, for the total baryon number after baryogenesis, $n_B/s \equiv (n_b - n_{\bar{b}})/s$.

- It implies, via eq. (3.2.62), that

$$g_{\star S}(T)\, T^3\, a^3 = const. \ , \qquad \text{or} \qquad T \propto g_{\star S}^{-1/3} a^{-1} \ . \tag{3.2.66}$$

Away from particle mass thresholds $g_{\star S}$ is approximately constant and $T \propto a^{-1}$, as expected. The factor of $g_{\star S}^{-1/3}$ accounts for the fact that whenever a particle species becomes non-relativistic and disappears, its entropy is transferred to the other relativistic species still present in the thermal plasma, causing T to decrease slightly less slowly than a^{-1}. We will see an example in the next section (cf. fig. 3.5).

Substituting $T \propto g_{\star S}^{-1/3} a^{-1}$ into the Friedmann equation

$$H = \frac{1}{a}\frac{da}{dt} \simeq \left(\frac{\rho_r}{3M_{\rm pl}^2}\right)^{1/2} \simeq \frac{\pi}{3}\left(\frac{g_\star}{10}\right)^{1/2} \frac{T^2}{M_{\rm pl}} \ , \tag{3.2.67}$$

we reproduce the usual result for a radiation dominated universe, $a \propto t^{1/2}$, except that there is a change in the scaling every time $g_{\star S}$ changes. For $T \propto t^{-1/2}$, we can integrate the Friedmann equation and get the temperature as a function of time

$$\frac{T}{1\,\text{MeV}} \simeq 1.5\, g_\star^{-1/4} \left(\frac{1\,\text{sec}}{t}\right)^{1/2} \ . \tag{3.2.68}$$

It is a useful rule of thumb that the temperature of the universe 1 second after the Big Bang was about 1 MeV, and evolved as $t^{-1/2}$ before that.

3.2.4 Neutrino Decoupling

Neutrinos are coupled to the thermal bath via weak interaction processes like

$$\begin{aligned} \nu_e + \bar{\nu}_e &\leftrightarrow e^+ + e^- \ , \\ e^- + \bar{\nu}_e &\leftrightarrow e^- + \bar{\nu}_e \ . \end{aligned} \tag{3.2.69}$$

The cross section for these interactions was estimated in (3.1.9), $\sigma \sim G_F^2 T^2$, and hence it was found that $\Gamma \sim G_F^2 T^5$. As the temperature decreases, the interaction rate drops much more rapidly that the Hubble rate $H \sim T^2/M_{\rm pl}$:

$$\frac{\Gamma}{H} \sim \left(\frac{T}{1\,\text{MeV}}\right)^3 \ . \tag{3.2.70}$$

We conclude that neutrinos decouple around 1 MeV. (A more accurate computation gives $T_{dec} \sim 0.8$ MeV.) After decoupling, the neutrinos move freely along geodesics and preserve to an excellent approximate the *relativistic* Fermi-Dirac distribution (even after they become non-relativistic at later times). In §1.2.1, we showed the physical momentum of a particle scales

as $p \propto a^{-1}$. It is therefore convenient to define the time-independent combination $q \equiv ap$, so that the neutrino number density is

$$n_\nu \propto a^{-3} \int \mathrm{d}^3 q \, \frac{1}{\exp(q/aT_\nu) + 1} \ . \tag{3.2.71}$$

After decoupling, particle number conservation requires $n_\nu \propto a^{-3}$. This is only consistent with (3.2.71) if the neutrino temperature evolves as $T_\nu \propto a^{-1}$. As long as the photon temperature[13] T_γ scales in the same way, we still have $T_\nu = T_\gamma$. However, particle annihilations will cause a deviation from $T_\gamma \propto a^{-1}$ in the photon temperature.

3.2.5 Electron-Positron Annihilation

Shortly after the neutrinos decouple, the temperature drops below the electron mass and electron-positron annihilation occurs

$$e^+ + e^- \ \leftrightarrow \ \gamma + \gamma \ . \tag{3.2.72}$$

The energy density and entropy of the electrons and positrons are transferred to the photons, but not to the decoupled neutrinos. The photons are thus "heated" (the photon temperature does not decrease as much) relative to the neutrinos (see fig. 3.5). To quantify this effect, we

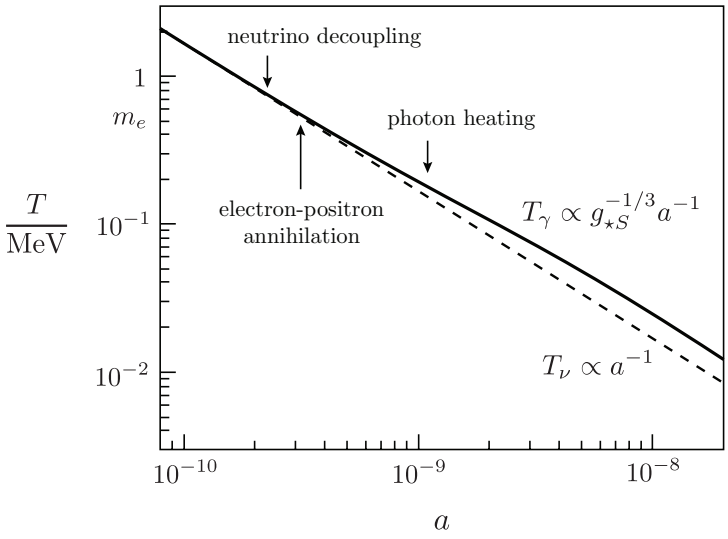


Figure 3.5: *Thermal history through electron-positron annihilation. Neutrinos are decoupled and their temperature redshifts simply as $T_\nu \propto a^{-1}$. The energy density of the electron-positron pairs is transferred to the photon gas whose temperature therefore redshifts more slowly, $T_\gamma \propto g_{\star S}^{-1/3} a^{-1}$.*

consider the change in the effective number of degrees of freedom in entropy. If we neglect neutrinos and other decoupled species,[14] we have

$$g_{\star S}^{th} = \begin{cases} 2 + \frac{7}{8} \times 4 \ = \ \frac{11}{2} & T \gtrsim m_e \\ 2 & T < m_e \end{cases} \ . \tag{3.2.73}$$

Since, in equilibrium, $g_{\star S}^{th}(aT_\gamma)^3$ remains constant, we find that aT_γ increases after electron-positron annihilation, $T < m_e$, by a factor $(11/4)^{1/3}$, while aT_ν remains the same. This means

[13] For the moment we will restore the subscript on the photon temperature to highlight the difference with the neutrino temperature.

[14] Obviously, entropy is separately conserved for the thermal bath and the decoupling species.

that the temperature of neutrinos is slightly lower than the photon temperature after e^+e^- annihilation,

$$T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma \ . \tag{3.2.74}$$

For $T \ll m_e$, the effective number of relativistic species (in energy density and entropy) therefore is

$$g_\star = 2 + \frac{7}{8} \times 2N_{\rm eff} \left(\frac{4}{11}\right)^{4/3} = 3.36 \ , \tag{3.2.75}$$

$$g_{\star S} = 2 + \frac{7}{8} \times 2N_{\rm eff} \left(\frac{4}{11}\right) = 3.94 \ , \tag{3.2.76}$$

where we have introduced the parameter $N_{\rm eff}$ as the *effective* number of neutrino species in the universe. If neutrinos decoupling was instantaneous then we have $N_{\rm eff} = 3$. However, neutrino decoupling was not quite complete when e^+e^- annihilation began, so some of the energy and entropy did leak to the neutrinos. Taking this into account[15] raises the effective number of neutrinos to $N_{\rm eff} = 3.046$.[16] Using this value in (3.2.75) and (3.2.76) explains the final values of $g_\star(T)$ and $g_{\star S}(T)$ in fig. 3.1.

3.2.6 Cosmic Neutrino Background

The relation (3.2.74) holds until the present. The cosmic neutrino background (CνB) therefore has a slightly lower temperature, $T_{\nu,0} = 1.95$ K $= 0.17$ meV, than the cosmic microwave background, $T_0 = 2.73$ K $= 0.24$ meV. The number density of neutrinos is

$$n_\nu = \frac{3}{4} N_{\rm eff} \times \frac{4}{11} n_\gamma \ . \tag{3.2.77}$$

Using (3.2.41), we see that this corresponds to 112 neutrinos cm^{-3} per flavour. The present energy density of neutrinos depends on whether the neutrinos are relativistic or non-relativistic today. It used to be believe that neutrinos were massless in which case we would have

$$\rho_\nu = \frac{7}{8} N_{\rm eff} \left(\frac{4}{11}\right)^{4/3} \rho_\gamma \quad \Rightarrow \quad \Omega_\nu h^2 \approx 1.7 \times 10^{-5} \quad (m_\nu = 0) \ . \tag{3.2.78}$$

Neutrino oscillation experiments have since shown that neutrinos do have mass. The minimum sum of the neutrino masses is $\sum m_{\nu,i} > 60$ meV. Massive neutrinos behave as radiation-like particles in the early universe[17], and as matter-like particles in the late universe (see fig. 3.6). On Problem Set 2, you will show that energy density of massive neutrinos, $\rho_\nu = \sum m_{\nu,i} n_{\nu,i}$, corresponds to

$$\Omega_\nu h^2 \approx \frac{\sum m_{\nu,i}}{94 \text{ eV}} \ . \tag{3.2.79}$$

By demanding that neutrinos don't over close the universe, i.e. $\Omega_\nu < 1$, one sets a cosmological upper bound on the sum of the neutrino masses, $\sum m_{\nu,i} < 15$ eV (using $h = 0.7$). Measurements

[15] To get the precise value of $N_{\rm eff}$ one also has to consider the fact that the neutrino spectrum after decoupling deviates slightly from the Fermi-Dirac distribution. This spectral distortion arises because the energy dependence of the weak interaction causes neutrinos in the high-energy tail to interact more strongly.

[16] The Planck constraint on $N_{\rm eff}$ is 3.36 ± 0.34. This still leaves room for discovering that $N_{\rm eff} \neq 3.046$, which is one of the avenues in which cosmology could discover new physics beyond the Standard Model.

[17] For $m_\nu < 0.2$ eV, neutrinos are relativistic at recombination.

of tritium β-decay, in fact, find that $\sum m_{\nu,i} < 6$ eV. Moreover, observations of the cosmic microwave background, galaxy clustering and type Ia supernovae together put an even stronger bound, $\sum m_{\nu,i} < 1$ eV. This implies that although neutrinos contribute at least 25 times the energy density of photons, they are still a subdominant component overall, $0.001 < \Omega_\nu < 0.02$.

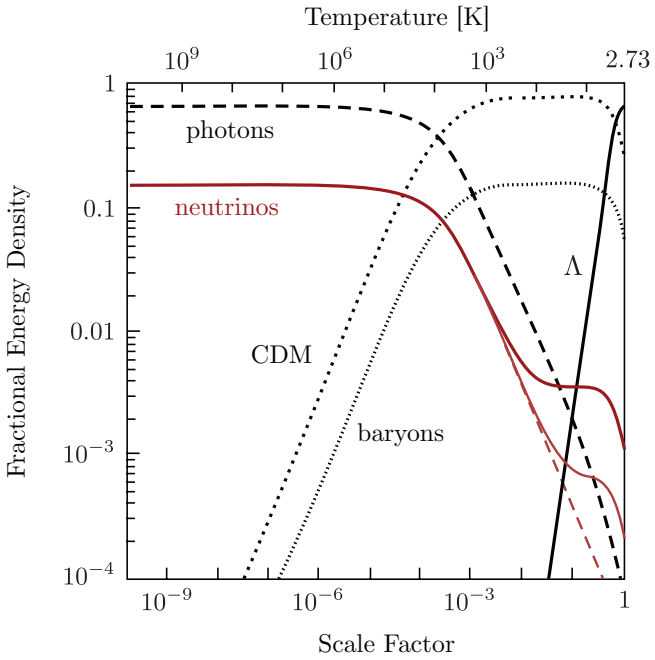


Figure 3.6: *Evolution of the fractional energy densities of photons, three neutrino species (one massless and two massive – 0.05 and 0.01 eV), cold dark matter (CDM), baryons, and a cosmological constant (Λ). Notice the change in the behaviour of the two massive neutrinos when they become non-relativistic particles.*

3.3 Beyond Equilibrium

The formal tool to describe the evolution beyond equilibrium is the Boltzmann equation. In this section, we first introduce the Boltzmann equation and then apply it to three important examples: (i) the production of dark matter; (ii) the formation of the light elements during Big Bang nucleosynthesis; and (iii) the recombination of electrons and protons into neutral hydrogen.

3.3.1 Boltzmann Equation

In the absence of interactions, the number density of a particle species i evolves as

$$\frac{dn_i}{dt} + 3\frac{\dot{a}}{a}n_i = 0 \ . \tag{3.3.80}$$

This is simply a reflection of the fact that the number of particles in a fixed physical volume ($V \propto a^3$) is conserved, so that the density dilutes with the expanding volume, $n_i \propto a^{-3}$, cf. eq. (1.3.89). To include the effects of interactions we add a collision term to the r.h.s. of (3.3.80),

$$\frac{1}{a^3}\frac{d(n_i a^3)}{dt} = C_i[\{n_j\}] \ . \tag{3.3.81}$$

This is the *Boltzmann equation*. The form of the collision term depends on the specific interactions under consideration. Interactions between three or more particles are very unlikely, so

we can limit ourselves to single-particle decays and two-particle scatterings / annihilations. For concreteness, let us consider the following process

$$1 + 2 \;\rightleftarrows\; 3 + 4 \;, \tag{3.3.82}$$

i.e. particle 1 can annihilate with particle 2 to produce particles 3 and 4, or the inverse process can produce 1 and 2. This reaction will capture all processes studied in this chapter. Suppose we are interested in tracking the number density n_1 of species 1. Obviously, the rate of change in the abundance of species 1 is given by the difference between the rates for producing and eliminating the species. The Boltzmann equation simply formalises this statement,

$$\frac{1}{a^3}\frac{d(n_1 a^3)}{dt} = -\alpha\, n_1 n_2 + \beta\, n_3 n_4 \;. \tag{3.3.83}$$

We understand the r.h.s. as follows: The first term, $-\alpha n_1 n_2$, describes the destruction of particles 1, while that second term, $+\beta\, n_3 n_4$. Notice that the first term is proportional to n_1 and n_2 and the second term is proportional to n_3 and n_4. The parameter $\alpha = \langle\sigma v\rangle$ is the *thermally averaged cross section.*[18] The second parameter β can be related to α by noting that the collision term has to vanish in (chemical) equilibrium

$$\beta = \left(\frac{n_1 n_2}{n_3 n_4}\right)_{\rm eq} \alpha \;, \tag{3.3.84}$$

where $n_i^{\rm eq}$ are the equilibrium number densities we calculated above. We therefore find

$$\boxed{\frac{1}{a^3}\frac{d(n_1 a^3)}{dt} = -\langle\sigma v\rangle\left[n_1 n_2 - \left(\frac{n_1 n_2}{n_3 n_4}\right)_{\rm eq} n_3 n_4\right]} \;. \tag{3.3.85}$$

It is instructive to write this in terms of the number of particles in a comoving volume, as defined in (3.2.65), $N_i \equiv n_i/s$. This gives

$$\frac{d\ln N_1}{d\ln a} = -\frac{\Gamma_1}{H}\left[1 - \left(\frac{N_1 N_2}{N_3 N_4}\right)_{\rm eq}\frac{N_3 N_4}{N_1 N_2}\right] \;, \tag{3.3.86}$$

where $\Gamma_1 \equiv n_2\langle\sigma v\rangle$. The r.h.s. of (3.3.86) contains a factor describing the *interaction efficiency*, Γ_1/H, and a factor characterizing the *deviation from equilibrium*, $[1 - \cdots]$.

For $\Gamma_1 \gg H$, the natural state of the system is chemical equilibrium. Imagine that we start with $N_1 \gg N_1^{\rm eq}$ (while $N_i \sim N_i^{\rm eq}$, $i = 2, 3, 4$). The r.h.s. of (3.3.86) then is negative, particles of type 1 are destroyed and N_1 is reduced towards the equilibrium value $N_1^{\rm eq}$. Similarly, if $N_1 \ll N_1^{\rm eq}$, the r.h.s. of (3.3.86) is positive and N_1 is driven towards $N_1^{\rm eq}$. The same conclusion applies if several species deviate from their equilibrium values. As long as the interaction rates are large, the system quickly relaxes to a steady state where the r.h.s. of (3.3.86) vanishes and the particles assume their equilibrium abundances.

When the reaction rate drops below the Hubble scale, $\Gamma_1 < H$, the r.h.s. of (3.3.86) gets suppressed and the comoving density of particles approaches a constant relic density, i.e. $N_1 = const$. This is illustrated in fig. 3.2. We will see similar types of evolution when we study the freeze-out of dark matter particles in the early universe (fig. 3.7), neutrons in BBN (fig. 3.9) and electrons in recombination (fig. 3.8).

[18] You will learn in the *QFT* and *Standard Model* courses how to compute *cross sections* σ for elementary processes. In this course, we will simply use dimensional analysis to estimate the few cross sections that we will need. The cross section may depend on the *relative velocity* v of particles 1 and 2. The angle brackets in $\alpha = \langle\sigma v\rangle$ denote an average over v.

3.3.2 Dark Matter Relics

We start with the slightly speculative topic of dark matter freeze-out. I call this speculative because it requires us to make some assumptions about the nature of the unknown dark matter particles. For concreteness, we will focus on the hypothesis that the dark matter is a weakly interacting massive particle (WIMP).

Freeze-Out

WIMPs were in close contact with the rest of the cosmic plasma at high temperatures, but then experienced freeze-out at a critical temperature T_f. The purpose of this section is to solve the Boltzmann equation for such a particle, determining the epoch of freeze-out and its relic abundance.

To get started we have to assume something about the WIMP interactions in the early universe. We will imagine that a heavy dark matter particle X and its antiparticle $\bar{X}$ can annihilate to produce two light (essentially massless) particles ℓ and $\bar{\ell}$,

$$X + \bar{X} \;\leftrightarrow\; \ell + \bar{\ell} \,. \tag{3.3.87}$$

Moreover, we assume that the light particles are tightly coupled to the cosmic plasma,[19] so that throughout they maintain their equilibrium densities, $n_\ell = n_\ell^{\rm eq}$. Finally, we assume that there is no initial asymmetry between X and $\bar{X}$, i.e. $n_X = n_{\bar{X}}$. The Boltzmann equation (3.3.85) for the evolution of the number of WIMPs in a comoving volume, $N_X \equiv n_X/s$, then is

$$\frac{dN_X}{dt} = -s\langle\sigma v\rangle\Big[N_X^2 - (N_X^{\rm eq})^2\Big] \,, \tag{3.3.88}$$

where $N_X^{\rm eq} \equiv n_X^{\rm eq}/s$. Since most of the interesting dynamics will take place when the temperature is of order the particle mass, $T \sim M_X$, it is convenient to define a new measure of time,

$$x \equiv \frac{M_X}{T} \,. \tag{3.3.89}$$

To write the Boltzmann equation in terms of x rather than t, we note that

$$\frac{dx}{dt} = \frac{d}{dt}\left(\frac{M_X}{T}\right) = -\frac{1}{T}\frac{{\rm d}T}{dt}\,x \simeq Hx \,, \tag{3.3.90}$$

where we have assumed that $T \propto a^{-1}$ (i.e. $g_{\star S} \approx const. \equiv g_{\star S}(M_X)$) for the times relevant to the freeze-out. We assume radiation domination so that $H = H(M_X)/x^2$. Eq. (3.3.88) then becomes the so-called *Riccati equation*,

$$\boxed{\frac{dN_X}{dx} = -\frac{\lambda}{x^2}\Big[N_X^2 - (N_X^{\rm eq})^2\Big]} \,, \tag{3.3.91}$$

where we have defined

$$\lambda \equiv \frac{2\pi^2}{45} g_{\star S} \frac{M_X^3 \langle\sigma v\rangle}{H(M_X)} \,. \tag{3.3.92}$$

We will treat λ as a constant (which in more fundamental theories of WIMPs is usually a good approximation). Unfortunately, even for constant λ, there are no analytic solutions to (3.3.91). Fig. 3.7 shows the result of a numerical solution for two different values of λ. As expected,

[19] This would be case case, for instance, if ℓ and $\bar{\ell}$ were electrically charged.

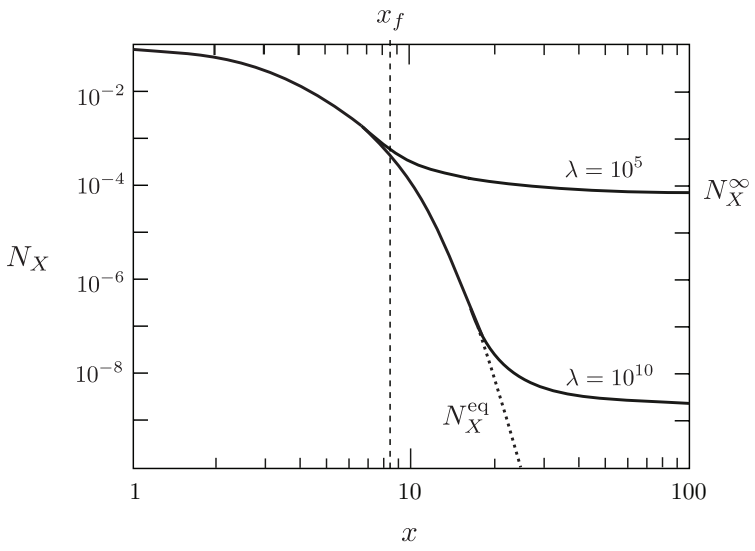


Figure 3.7: *Abundance of dark matter particles as the temperature drops below the mass.*

at very high temperatures, $x < 1$, we have $N_X \approx N_X^{\rm eq} \simeq 1$. However, at low temperatures, $x \gg 1$, the equilibrium abundance becomes exponentially suppressed, $N_X^{\rm eq} \sim e^{-x}$. Ultimately, X-particles will become so rare that they will not be able to find each other fast enough to maintain the equilibrium abundance. Numerically, we find that freeze-out happens at about $x_f \sim 10$. This is when the solution of the Boltzmann equation starts to deviate significantly from the equilibrium abundance.

The final relic abundance, $N_X^\infty \equiv N_X(x = \infty)$, determines the freeze-out density of dark matter. Let us estimate its magnitude as a function of λ. Well after freeze-out, N_X will be much larger than $N_X^{\rm eq}$ (see fig. 3.7). Thus at late times, we can drop $N_X^{\rm eq}$ from the Boltzmann equation,

$$\frac{dN_X}{dx} \simeq -\frac{\lambda N_X^2}{x^2} \qquad (x > x_f) \ . \tag{3.3.93}$$

Integrating from x_f, to $x = \infty$, we find

$$\frac{1}{N_X^\infty} - \frac{1}{N_X^f} = \frac{\lambda}{x_f} \ , \tag{3.3.94}$$

where $N_X^f \equiv N_X(x_f)$. Typically, $N_X^f \gg N_X^\infty$ (see fig. 3.7), so a simple analytic approximation is

$$\boxed{N_X^\infty \simeq \frac{x_f}{\lambda}} \ . \tag{3.3.95}$$

Of course, this still depends on the unknown freeze-out time (or temperature) x_f. As we see from fig. 3.7, a good order-of-magnitude estimate is $x_f \sim 10$. The value of x_f isn't terribly sensitive to the precise value of λ, namely $x_f(\lambda) \propto |\ln \lambda|$.

Exercise.—Estimate $x_f(\lambda)$ from $\Gamma(x_f) = H(x_f)$.

Eq. (3.3.95) predicts that the freeze-out abundance N_X^∞ decreases as the interaction rate λ increases. This makes sense intuitively: larger interactions maintain equilibrium longer, deeper into the Boltzmann-suppressed regime. Since the estimate in (3.3.95) works quite well, we will use it in the following.

WIMP Miracle*

It just remains to relate the freeze-out abundance of dark matter relics to the dark matter density today:

$$\begin{aligned}\Omega_X &\equiv \frac{\rho_{X,0}}{\rho_{\mathrm{crit},0}} \\ &= \frac{M_X n_{X,0}}{3M_{\mathrm{pl}}^2 H_0^2} = \frac{M_X N_{X,0} s_0}{3M_{\mathrm{pl}}^2 H_0^2} = M_X N_X^\infty \frac{s_0}{3M_{\mathrm{pl}}^2 H_0^2} \ . \end{aligned} \tag{3.3.96}$$

where we have used that the number of WIMPs is conserved after freeze-out, i.e. $N_{X,0} = N_X^\infty$. Substituting $N_X^\infty = x_f/\lambda$ and $s_0 \equiv s(T_0)$, we get

$$\Omega_X = \frac{H(M_X)}{M_X^2} \frac{x_f}{\langle\sigma v\rangle} \frac{g_{\star S}(T_0)}{g_{\star S}(M_X)} \frac{T_0^3}{3M_{\mathrm{pl}}^2 H_0^2} \ , \tag{3.3.97}$$

where we have used (3.3.92) and (3.2.62). Using (3.2.67) for $H(M_X)$, gives

$$\Omega_X = \frac{\pi}{9} \frac{x_f}{\langle\sigma v\rangle} \left(\frac{g_\star(M_X)}{10}\right)^{1/2} \frac{g_{\star S}(T_0)}{g_{\star S}(M_X)} \frac{T_0^3}{M_{\mathrm{pl}}^3 H_0^2} \ . \tag{3.3.98}$$

Finally, we substitute the measured values of T_0 and H_0 and use $g_{\star S}(T_0) = 3.91$ and $g_{\star S}(M_X) = g_\star(M_X)$:

$$\Omega_X h^2 \sim 0.1 \left(\frac{x_f}{10}\right) \left(\frac{10}{g_\star(M_X)}\right)^{1/2} \frac{10^{-8}\,\mathrm{GeV}^{-2}}{\langle\sigma v\rangle} \ . \tag{3.3.99}$$

This reproduces the observed dark matter density if

$$\sqrt{\langle\sigma v\rangle} \sim 10^{-4}\,\mathrm{GeV}^{-1} \sim 0.1\sqrt{G_F} \ .$$

The fact that a thermal relic with a cross section characteristic of the weak interaction gives the right dark matter abundance is called the *WIMP miracle*.

3.3.3 Recombination

An important event in the history of the early universe is the formation of the first atoms. At temperatures above about 1 eV, the universe still consisted of a plasma of free electrons and nuclei. Photons were tightly coupled to the electrons via Compton scattering, which in turn strongly interacted with protons via Coulomb scattering. There was very little neutral hydrogen. When the temperature became low enough, the electrons and nuclei combined to form neutral atoms (*recombination*[20]), and the density of free electrons fell sharply. The photon mean free path grew rapidly and became longer than the horizon distance. The photons *decoupled* from the matter and the universe became transparent. Today, these photons are the *cosmic microwave background*.

Saha Equilibrium

Let us start at $T > 1$ eV, when baryons and photons were still in equilibrium through electromagnetic reactions such as

$$e^- + p^+ \quad \leftrightarrow \quad \mathrm{H} + \gamma \ . \tag{3.3.100}$$

[20] Don't ask me why this is called *re*combination; this is the first time electrons and nuclei combined.

Since $T < m_i$, $i = \{e, p, \mathrm{H}\}$, we have the following equilibrium abundances

$$n_i^{\mathrm{eq}} = g_i \left(\frac{m_i T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_i - m_i}{T}\right) \ , \tag{3.3.101}$$

where $\mu_p + \mu_e = \mu_{\mathrm{H}}$ (recall that $\mu_\gamma = 0$). To remove the dependence on the chemical potentials, we consider the following ratio

$$\left(\frac{n_{\mathrm{H}}}{n_e n_p}\right)_{\mathrm{eq}} = \frac{g_{\mathrm{H}}}{g_e g_p} \left(\frac{m_{\mathrm{H}}}{m_e m_p}\frac{2\pi}{T}\right)^{3/2} e^{(m_p + m_e - m_{\mathrm{H}})/T} \ . \tag{3.3.102}$$

In the prefactor, we can use $m_{\mathrm{H}} \approx m_p$, but in the exponential the small difference between m_{H} and $m_p + m_e$ is crucial: it is the binding energy of hydrogen

$$B_{\mathrm{H}} \equiv m_p + m_e - m_{\mathrm{H}} = 13.6 \text{ eV} \ . \tag{3.3.103}$$

The number of internal degrees of freedom are $g_p = g_e = 2$ and $g_{\mathrm{H}} = 4$.[21] Since, as far as we know, the universe isn't electrically charged, we have $n_e = n_p$. Eq. (3.3.102) therefore becomes

$$\left(\frac{n_{\mathrm{H}}}{n_e^2}\right)_{\mathrm{eq}} = \left(\frac{2\pi}{m_e T}\right)^{3/2} e^{B_{\mathrm{H}}/T} \ . \tag{3.3.104}$$

We wish to follow the *free electron fraction* defined as the ratio

$$X_e \equiv \frac{n_e}{n_b} \ , \tag{3.3.105}$$

where n_b is the baryon density. We may write the baryon density as

$$n_b = \eta\, n_\gamma = \eta \times \frac{2\zeta(3)}{\pi^2}\, T^3 \ , \tag{3.3.106}$$

where $\eta = 5.5 \times 10^{-10}(\Omega_b h^2/0.020)$ is the *baryon-to-photon ratio*. To simplify the discussion, let us ignore all nuclei other than protons (over 90% (by number) of the nuclei are protons). The total baryon number density can then be approximated as $n_b \approx n_p + n_{\mathrm{H}} = n_e + n_{\mathrm{H}}$ and hence

$$\frac{1 - X_e}{X_e^2} = \frac{n_{\mathrm{H}}}{n_e^2}\, n_b \ . \tag{3.3.107}$$

Substituting (3.3.104) and (3.3.106), we arrive at the so-called *Saha equation*,

$$\boxed{\left(\frac{1 - X_e}{X_e^2}\right)_{\mathrm{eq}} = \frac{2\zeta(3)}{\pi^2}\, \eta \left(\frac{2\pi T}{m_e}\right)^{3/2} e^{B_{\mathrm{H}}/T}} \ . \tag{3.3.108}$$

Fig. 3.8 shows the redshift evolution of the free electron fraction as predicted both by the Saha approximation (3.3.108) and by a more exact numerical treatment (see below). The Saha approximation correctly identifies the onset of recombination, but it is clearly insufficient if the aim is to determine the relic density of electrons after freeze-out.

[21] The spins of the electron and proton in a hydrogen atom can be aligned or anti-aligned, giving one singlet state and one triplet state, so $g_{\mathrm{H}} = 1 + 3 = 4$.

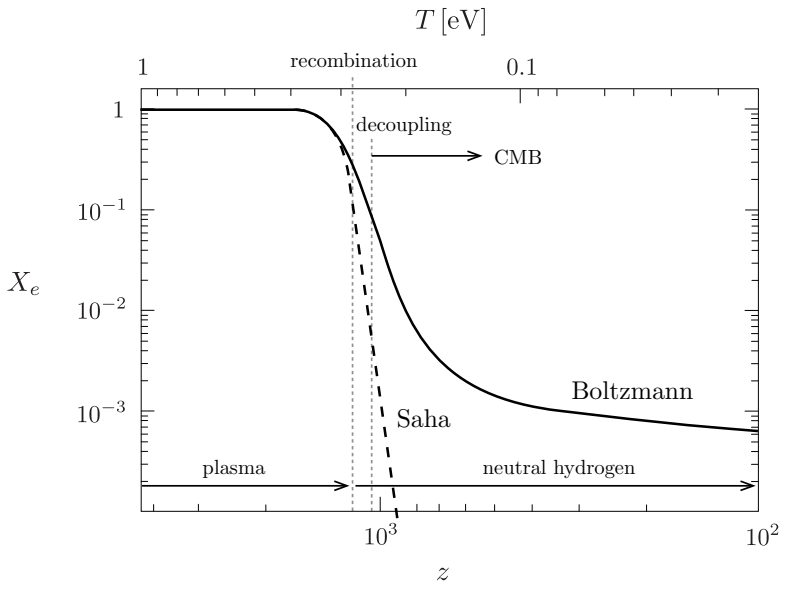


Figure 3.8: *Free electron fraction as a function of redshift.*

Hydrogen Recombination

Let us define the recombination temperature T_{rec} as the temperature where[22] $X_e = 10^{-1}$ in (3.3.108), i.e. when 90% of the electrons have combined with protons to form hydrogen. We find

$$T_{rec} \approx 0.3\,\text{eV} \simeq 3600\,\text{K} \ . \tag{3.3.109}$$

The reason that $T_{rec} \ll B_{\rm H} = 13.6$ eV is that there are very many photons for each hydrogen atom, $\eta \sim 10^{-9} \ll 1$. Even when $T < B_{\rm H}$, the high-energy tail of the photon distribution contains photons with energy $E > B_{\rm H}$ so that they can ionize a hydrogen atom.

Exercise.—Confirm the estimate in (3.3.109).

Using $T_{rec} = T_0(1 + z_{rec})$, with $T_0 = 2.7\,\text{K}$, gives the redshift of recombination,

$$z_{rec} \approx 1320 \ . \tag{3.3.110}$$

Since matter-radiation equality is at $z_{eq} \simeq 3500$, we conclude that recombination occurred in the matter-dominated era. Using $a(t) = (t/t_0)^{2/3}$, we obtain an estimate for the time of recombination

$$t_{rec} = \frac{t_0}{(1 + z_{rec})^{3/2}} \sim 290\,000\,\text{yrs} \ . \tag{3.3.111}$$

Photon Decoupling

Photons are most strongly coupled to the primordial plasma through their interactions with electrons

$$e^- + \gamma \quad \leftrightarrow \quad e^- + \gamma \ , \tag{3.3.112}$$

[22] There is nothing deep about the choice $X_e(T_{rec}) = 10^{-1}$. It is as arbitrary as it looks.

with an interaction rate given by

$$\Gamma_\gamma \approx n_e \sigma_T \ , \tag{3.3.113}$$

where $\sigma_T \approx 2 \times 10^{-3}\,\mathrm{MeV}^{-2}$ is the Thomson cross section. Since $\Gamma_\gamma \propto n_e$, the interaction rate decreases as the density of free electrons drops. Photons and electrons decouple roughly when the interaction rate becomes smaller than the expansion rate,

$$\Gamma_\gamma(T_{dec}) \sim H(T_{dec}) \ . \tag{3.3.114}$$

Writing

$$\Gamma_\gamma(T_{dec}) = n_b X_e(T_{dec})\, \sigma_T = \frac{2\zeta(3)}{\pi^2}\, \eta\, \sigma_T\, X_e(T_{dec}) T_{dec}^3 \ , \tag{3.3.115}$$

$$H(T_{dec}) = H_0 \sqrt{\Omega_m} \left(\frac{T_{dec}}{T_0} \right)^{3/2} \ . \tag{3.3.116}$$

we get

$$X_e(T_{dec}) T_{dec}^{3/2} \sim \frac{\pi^2}{2\zeta(3)} \frac{H_0 \sqrt{\Omega_m}}{\eta \sigma_T T_0^{3/2}} \ . \tag{3.3.117}$$

Using the Saha equation for $X_e(T_{dec})$, we find

$$T_{dec} \sim 0.27\,\mathrm{eV} \ . \tag{3.3.118}$$

Notice that although T_{dec} isn't far from T_{rec}, the ionization fraction decreases significantly between recombination and decoupling, $X_e(T_{rec}) \simeq 0.1 \to X_e(T_{dec}) \simeq 0.01$. This shows that a large degree of neutrality is necessary for the universe to become transparent to photon propagation.

Exercise.—Using (3.3.108), confirm the estimate in (3.3.118).

The redshift and time of decoupling are

$$z_{dec} \ \sim \ 1100 \ , \tag{3.3.119}$$

$$t_{dec} \ \sim \ 380\,000\,\mathrm{yrs} \ . \tag{3.3.120}$$

After decoupling the photons stream freely. Observations of the cosmic microwave background today allow us to probe the conditions at last-scattering.

Electron Freeze-Out*

In fig. 3.8, we see that a residual ionisation fraction of electrons freezes out when the interactions in (3.3.100) become inefficient. To follow the free electron fraction after freeze-out, we need to solve the Boltzmann equation, just as we did for the dark matter freeze-out.

We apply our non-equilibrium master equation (3.3.85) to the reaction (3.3.100). To a reasonably good approximation the neutral hydrogen tracks its equilibrium abundance throughout, $n_{\mathrm{H}} \approx n_{\mathrm{H}}^{\mathrm{eq}}$. The Boltzmann equation for the electron density can then be written as

$$\frac{1}{a^3} \frac{d(n_e a^3)}{dt} = -\langle \sigma v \rangle \left[n_e^2 - (n_e^{\mathrm{eq}})^2 \right] \ . \tag{3.3.121}$$

Actually computing the thermally averaged recombination cross section $\langle\sigma v\rangle$ from first principles is quite involved, but a reasonable approximation turns out to be

$$\langle\sigma v\rangle \;\simeq\; \sigma_T \left(\frac{B_{\rm H}}{T}\right)^{1/2} \;. \tag{3.3.122}$$

Writing $n_e = n_b X_e$ and using that $n_b a^3 = const.$, we find

$$\boxed{\frac{dX_e}{dx} = -\frac{\lambda}{x^2}\left[X_e^2 - (X_e^{\rm eq})^2\right]}\;, \tag{3.3.123}$$

where $x \equiv B_{\rm H}/T$. We have used the fact that the universe is matter-dominated at recombination and defined

$$\lambda \equiv \left[\frac{n_b\langle\sigma v\rangle}{xH}\right]_{x=1} = 3.9\times 10^3 \left(\frac{\Omega_b h}{0.03}\right) \;. \tag{3.3.124}$$

Exercise.—Derive eq. (3.3.123).

Notice that eq. (3.3.123) is identical to eq. (3.3.91)—the Riccati equation for dark matter freeze-out. We can therefore immediately write down the electron freeze-out abundance, cf. eq. (3.3.95),

$$X_e^\infty \simeq \frac{x_f}{\lambda} \;=\; 0.9\times 10^{-3}\left(\frac{x_f}{x_{rec}}\right)\left(\frac{0.03}{\Omega_b h}\right) \;. \tag{3.3.125}$$

Assuming that freeze-out occurs close to the time of recombination, $x_{rec} \approx 45$, we capture the relic electron abundance pretty well (see fig. 3.8).

Exercise.—Using $\Gamma_e(T_f) \sim H(T_f)$, show that the freeze-out temperature satisfies

$$X_e(T_f)T_f = \frac{\pi^2}{2\zeta(3)}\frac{H_0\sqrt{\Omega_m}}{\eta\sigma_T T_0^{3/2} B_{\rm H}^{1/2}} \;. \tag{3.3.126}$$

Use the Saha equation to show that $T_f \sim 0.25$ eV and hence $x_f \sim 54$.

3.3.4 Big Bang Nucleosynthesis

Let us return to $T \sim 1$ MeV. Photons, electron and positrons are in equilibrium. Neutrinos are about to decouple. Baryons are non-relativistic and therefore much fewer in number than the relativistic species. Nevertheless, we now want to study what happened to these trace amounts of baryonic matter. The total number of nucleons stays constant due to baryon number conservation. This baryon number can be in the form of protons and neutrons or heavier nuclei. Weak nuclear reactions may convert neutrons and protons into each other and strong nuclear reactions may build nuclei from them. In this section, I want to show you how the light elements hydrogen, helium and lithium were synthesised in the Big Bang. I won't give a complete account of all of the complicated details of Big Bang Nucleosynthesis (BBN). Instead, the goal of this section will be more modest: I want to give you a theoretical understanding of a single number: the ratio of the density of helium to hydrogen,

$$\frac{n_{\rm He}}{n_{\rm H}} \sim \frac{1}{16} \;. \tag{3.3.127}$$

Fig. 3.9 summarizes the four steps that will lead us from protons and neutrons to helium.

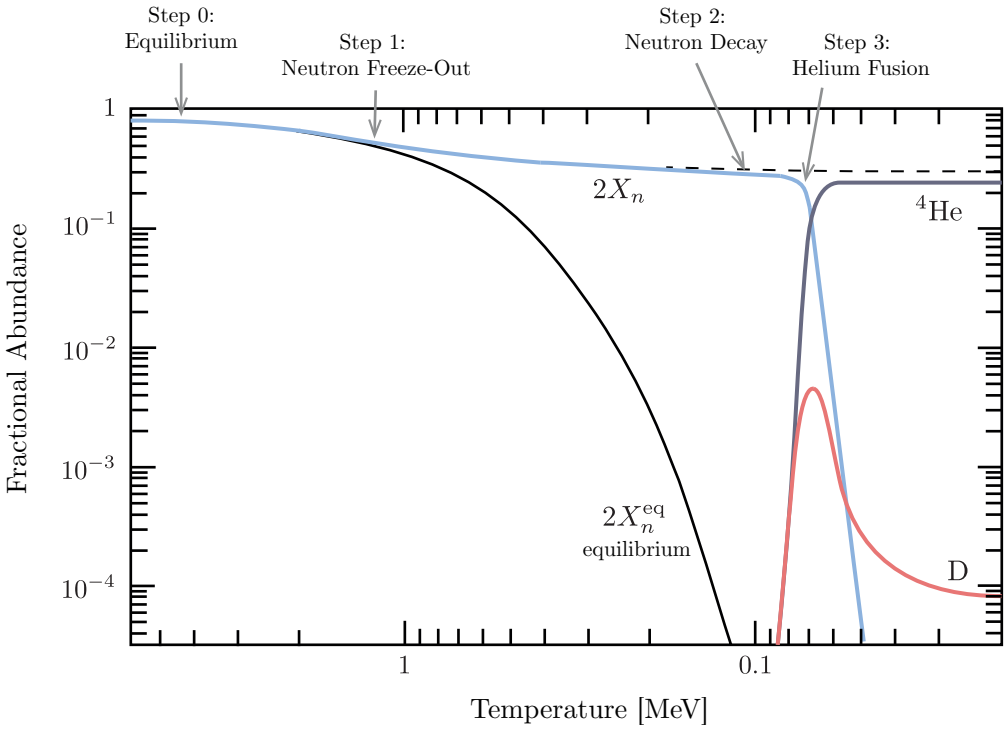


Figure 3.9: *Numerical results for helium production in the early universe.*

Step 0: Equilibrium Abundances

In principle, BBN is a very complicated process involving many coupled Boltzmann equations to track all the nuclear abundances. In practice, however, two simplifications will make our life a lot easier:

1. *No elements heavier than helium.*

   Essentially no elements heavier than helium are produced at appreciable levels. So the only nuclei that we need to track are hydrogen and helium, and their isotopes: deuterium, tritium, and ^{3}He.

2. *Only neutrons and protons above 0.1 MeV.*

   Above $T \approx 0.1$ MeV only free protons and neutrons exist, while other light nuclei haven't been formed yet. Therefore, we can first solve for the neutron/proton ratio and then use this abundance as input for the synthesis of deuterium, helium, etc.

Let us demonstrate that we can indeed restrict our attention to neutrons and protons above 0.1 MeV. In order to do this, we compare the equilibrium abundances of the different nuclei:

- First, we determine the relative abundances of neutrons and protons. In the early universe, neutrons and protons are coupled by weak interactions, e.g. β-decay and inverse β-decay

$$\begin{aligned} n + \nu_e &\leftrightarrow p^+ + e^- \ , \\ n + e^+ &\leftrightarrow p^+ + \bar{\nu}_e \ . \end{aligned} \tag{3.3.128}$$

  Let us assume that the chemical potentials of electrons and neutrinos are negligibly small,

so that $\mu_n = \mu_p$. Using (3.3.101) for $n_i^{\rm eq}$, we then have

$$\left(\frac{n_n}{n_p}\right)_{\rm eq} = \left(\frac{m_n}{m_p}\right)^{3/2} e^{-(m_n-m_p)/T} \ . \tag{3.3.129}$$

The small difference between the proton and neutron mass can be ignored in the first factor, but crucially has to be kept in the exponential. Hence, we find

$$\left(\frac{n_n}{n_p}\right)_{\rm eq} = e^{-\mathcal{Q}/T} \ , \tag{3.3.130}$$

where $\mathcal{Q} \equiv m_n - m_p = 1.30$ MeV. For $T \gg 1$ MeV, there are therefore as many neutrons as protons. However, for $T < 1$ MeV, the neutron fraction gets smaller. If the weak interactions would operate efficiently enough to maintain equilibrium indefinitely, then the neutron abundance would drop to zero. Luckily, in the real world the weak interactions are not so efficient.

- Next, we consider *deuterium* (an isotope of hydrogen with one proton and one neutron). This is produced in the following reaction

$$n + p^+ \ \leftrightarrow \ {\rm D} + \gamma \ . \tag{3.3.131}$$

Since $\mu_\gamma = 0$, we have $\mu_n + \mu_p = \mu_{\rm D}$. To remove the dependence on the chemical potentials we consider

$$\left(\frac{n_{\rm D}}{n_n n_p}\right)_{\rm eq} = \frac{3}{4}\left(\frac{m_{\rm D}}{m_n m_p}\frac{2\pi}{T}\right)^{3/2} e^{-(m_{\rm D}-m_n-m_p)/T} \ , \tag{3.3.132}$$

where, as before, we have used (3.3.101) for $n_i^{\rm eq}$ (with $g_{\rm D} = 3$ and $g_p = g_n = 2$). In the prefactor, $m_{\rm D}$ can be set equal to $2m_n \approx 2m_p \approx 1.9$ GeV, but in the exponential the small difference between $m_n + m_p$ and $m_{\rm D}$ is crucial: it is the binding energy of deuterium

$$B_{\rm D} \equiv m_n + m_p - m_{\rm D} = 2.22 \text{ MeV} \ . \tag{3.3.133}$$

Therefore, as long as chemical equilibrium holds the deuterium-to-proton ratio is

$$\left(\frac{n_{\rm D}}{n_p}\right)_{\rm eq} = \frac{3}{4} n_n^{\rm eq}\left(\frac{4\pi}{m_p T}\right)^{3/2} e^{B_{\rm D}/T} \ . \tag{3.3.134}$$

To get an order of magnitude estimate, we approximate the neutron density by the baryon density and write this in terms of the photon temperature and the baryon-to-photon ratio,

$$n_n \sim n_b = \eta\, n_\gamma = \eta \times \frac{2\zeta(3)}{\pi^2}\, T^3 \ . \tag{3.3.135}$$

Eq. (3.3.134) then becomes

$$\left(\frac{n_{\rm D}}{n_p}\right)_{\rm eq} \approx \eta \left(\frac{T}{m_p}\right)^{3/2} e^{B_{\rm D}/T} \ . \tag{3.3.136}$$

The smallness of the baryon-to-photon ratio η inhibits the production of deuterium until the temperature drops well beneath the binding energy $B_{\rm D}$. The temperature has to drop enough so that $e^{B_{\rm D}/T}$ can compete with $\eta \sim 10^{-9}$. The same applies to all other nuclei. At temperatures above 0.1 MeV, then, virtually all baryons are in the form of neutrons and protons. Around this time, deuterium and helium are produced, but the reaction rates are by now too low to produce any heavier elements.

Step 1: Neutron Freeze-Out

The primordial ratio of neutrons to protons is of particular importance to the outcome of BBN, since essentially all the neutrons become incorporated into ^{4}He. As we have seen, weak interactions keep neutrons and protons in equilibrium until $T \sim$ MeV. After that, we must solve the Boltzmann equation (3.3.85) to track the neutron abundance. Since this is a bit involved, I won't describe it in detail (but see the box below). Instead, we will estimate the answer a bit less rigorously.

It is convenient to define the neutron fraction as

$$X_n \equiv \frac{n_n}{n_n + n_p} \ . \tag{3.3.137}$$

From the equilibrium ratio of neutrons to protons (3.3.130), we then get

$$X_n^{\rm eq}(T) = \frac{e^{-\mathcal{Q}/T}}{1 + e^{-\mathcal{Q}/T}} \ . \tag{3.3.138}$$

Neutrons follows this equilibrium abundance until neutrinos decouple at[23] $T_f \sim T_{dec} \sim 0.8$ MeV (see §3.2.4). At this moment, weak interaction processes such as (3.3.128) effectively shut off. The equilibrium abundance at that time is

$$X_n^{\rm eq}(0.8\,{\rm MeV}) = 0.17 \ . \tag{3.3.139}$$

We will take this as a rough estimate for the final freeze-out abundance,

$$X_n^\infty \sim X_n^{\rm eq}(0.8\,{\rm MeV}) \sim \frac{1}{6} \ . \tag{3.3.140}$$

We have converted the result to a fraction to indicate that this is only an order of magnitude estimate.

Exact treatment*.—OK, since you asked, I will show you some details of the more exact treatment. To be clear, this box is *definitely* **not** examinable!

Using the Boltzmann equation (3.3.85), with 1 = neutron, 3 = proton, and 2, 4 = leptons (with $n_\ell = n_\ell^{\rm eq}$), we find

$$\frac{1}{a^3}\frac{d(n_n a^3)}{dt} = -\Gamma_n \left[n_n - \left(\frac{n_n}{n_p}\right)_{\rm eq} n_p \right] \ , \tag{3.3.141}$$

where we have defined the rate for neutron/proton conversion as $\Gamma_n \equiv n_\ell \langle \sigma v \rangle$. Substituting (3.3.137) and (3.3.138), we find

$$\frac{dX_n}{dt} = -\Gamma_n \left[X_n - (1 - X_n) e^{-\mathcal{Q}/T} \right] \ . \tag{3.3.142}$$

Instead of trying to solve this for X_n as a function of time, we introduce a new evolution variable

$$x \equiv \frac{\mathcal{Q}}{T} \ . \tag{3.3.143}$$

We write the l.h.s. of (3.3.142) as

$$\frac{dX_n}{dt} = \frac{dx}{dt}\frac{dX_n}{dx} = -\frac{x}{T}\frac{dT}{dt}\frac{dX_n}{dx} = xH\frac{dX_n}{dx} \ , \tag{3.3.144}$$

[23] If is fortunate that $T_f \sim \mathcal{Q}$. This seems to be a coincidence: $\mathcal{Q}$ is determined by the strong and electromagnetic interactions, while the value of T_f is fixed by the weak interaction. Imagine a world in which $T_f \ll \mathcal{Q}$!

where in the last equality we used that $T \propto a^{-1}$. During BBN, we have

$$H = \sqrt{\frac{\rho}{3M_{\rm pl}^2}} = \underbrace{\frac{\pi}{3}\sqrt{\frac{g_\star}{10}}\frac{\mathcal{Q}^2}{M_{\rm pl}}}_{\equiv H_1 \approx 1.13\, s^{-1}} \frac{1}{x^2} \,, \qquad \text{with} \quad g_\star = 10.75 \,. \tag{3.3.145}$$

Eq. (3.3.142) then becomes

$$\frac{dX_n}{dx} = \frac{\Gamma_n}{H_1}\, x \left[e^{-x} - X_n(1+e^{-x})\right] \,. \tag{3.3.146}$$

Finally, we need an expression for the neutron-proton conversion rate, Γ_n. You can find a sketch of the required QFT calculation in Dodelson's book. Here, I just cite the answer

$$\Gamma_n(x) = \frac{255}{\tau_n} \cdot \frac{12 + 6x + x^2}{x^5} \,, \tag{3.3.147}$$

where $\tau_n = 886.7 \pm 0.8$ sec is the neutron lifetime. One can see that the conversion time Γ_n^{-1} is comparable to the age of the universe at a temperature of ~ 1 MeV. At later times, $T \propto t^{-1/2}$ and $\Gamma_n \propto T^3 \propto t^{-3/2}$, so the neutron-proton conversion time $\Gamma_n^{-1} \propto t^{3/2}$ becomes longer than the age of the universe. Therefore we get *freeze-out*, i.e. the reaction rates become slow and the neutron/proton ratio approaches a constant. Indeed, solving eq. (3.3.146) numerically, we find (see fig. 3.9)

$$X_n^\infty \equiv X_n(x=\infty) = 0.15 \,. \tag{3.3.148}$$

Step 2: Neutron Decay

At temperatures below 0.2 MeV (or $t \gtrsim 100$ sec) the finite lifetime of the neutron becomes important. To include neutron decay in our computation we simply multiply the freeze-out abundance (3.3.148) by an exponential decay factor

$$\boxed{X_n(t) = X_n^\infty e^{-t/\tau_n} = \frac{1}{6}\, e^{-t/\tau_n}} \,, \tag{3.3.149}$$

where $\tau_n = 886.7 \pm 0.8$ sec.

Step 3: Helium Fusion

At this point, the universe is mostly protons and neutron. Helium cannot form directly because the density is too low and the time available is too short for reactions involving three or more incoming nuclei to occur at any appreciable rate. The heavier nuclei therefore have to be built sequentially from lighter nuclei in two-particle reactions. The first nucleus to form is therefore deuterium,

$$n + p^+ \;\leftrightarrow\; \mathrm{D} + \gamma \,. \tag{3.3.150}$$

Only when deuterium is available can helium be formed,

$$\mathrm{D} + p^+ \;\leftrightarrow\; {}^3\mathrm{He} + \gamma \,, \tag{3.3.151}$$

$$\mathrm{D} + {}^3\mathrm{He} \;\leftrightarrow\; {}^4\mathrm{He} + p^+ \,. \tag{3.3.152}$$

Since deuterium is formed directly from neutrons and protons it can follow its equilibrium abundance as long as enough free neutrons are available. However, since the deuterium binding

energy is rather small, the deuterium abundance becomes large rather late (at $T < 100$ keV). So although heavier nuclei have larger binding energies and hence would have larger equilibrium abundances, they cannot be formed until sufficient deuterium has become available. This is the *deuterium bottleneck*. Only when there is enough deuterium, can helium be produced. To get a rough estimate for the time of nucleosynthesis, we determine the temperature $T_{\rm nuc}$ when the deuterium fraction in equilibrium would be of order one, i.e. $(n_{\rm D}/n_p)_{\rm eq} \sim 1$. Using (3.3.136), I find

$$T_{\rm nuc} \sim 0.06\,{\rm MeV}\ , \tag{3.3.153}$$

which via (3.2.68) with $g_\star = 3.38$ translates into

$$t_{\rm nuc} = 120\,{\rm sec}\left(\frac{0.1\,{\rm MeV}}{T_{\rm nuc}}\right)^2 \sim 330\ {\rm sec}. \tag{3.3.154}$$

Comment.—From fig. 3.9, we see that a better estimate would be $n_{\rm D}^{\rm eq}(T_{\rm nuc}) \simeq 10^{-3} n_p^{\rm eq}(T_{\rm nuc})$. This gives $T_{\rm nuc} \simeq 0.07$ MeV and $t_{\rm nuc} \simeq 250$ sec. Notice that $t_{\rm nuc} \ll \tau_n$, so eq. (3.3.149) won't be very sensitive to the estimate for $t_{\rm nuc}$.

Substituting $t_{\rm nuc} \sim 330$ sec into (3.3.149), we find

$$\boxed{X_n(t_{\rm nuc}) \sim \frac{1}{8}}\ . \tag{3.3.155}$$

Since the binding energy of helium is larger than that of deuterium, the Boltzmann factor $e^{B/T}$ favours helium over deuterium. Indeed, in fig. 3.9 we see that helium is produced almost immediately after deuterium. Virtually all remaining neutrons at $t \sim t_{\rm nuc}$ then are processed into ^{4}He. Since two neutrons go into one nucleus of ^{4}He, the final ^{4}He abundance is equal to half of the neutron abundance at $t_{\rm nuc}$, i.e. $n_{\rm He} = \frac{1}{2} n_n(t_{\rm nuc})$, or

$$\frac{n_{\rm He}}{n_{\rm H}} = \frac{n_{\rm He}}{n_p} \simeq \frac{\frac{1}{2}X_n(t_{\rm nuc})}{1 - X_n(t_{\rm nuc})} \sim \frac{1}{2}X_n(t_{\rm nuc}) \sim \boxed{\frac{1}{16}}\ , \tag{3.3.156}$$

as we wished to show. Sometimes, the result is expressed as the mass fraction of helium,

$$\boxed{\frac{4n_{\rm He}}{n_{\rm H}} \sim \frac{1}{4}}\ . \tag{3.3.157}$$

This prediction is consistent with the observed helium in the universe (see fig. 3.10).

BBN as a Probe of BSM Physics

We have arrived at a number for the final helium mass fraction, but we should remember that this number depends on several input parameters:

- $g_\star$: the number of relativistic degrees of freedom determines the Hubble parameter during the radiation era, $H \propto g_\star^{1/2}$, and hence affects the freeze-out temperature

$$G_F^2 T_f^5 \sim \sqrt{G_N g_\star}\, T_f^2 \quad \to \quad T_f \propto g_\star^{1/6}\ . \tag{3.3.158}$$

  Increasing $g_\star$ increases T_f, which increases the n/p ratio at freeze-out and hence increases the final helium abundance.

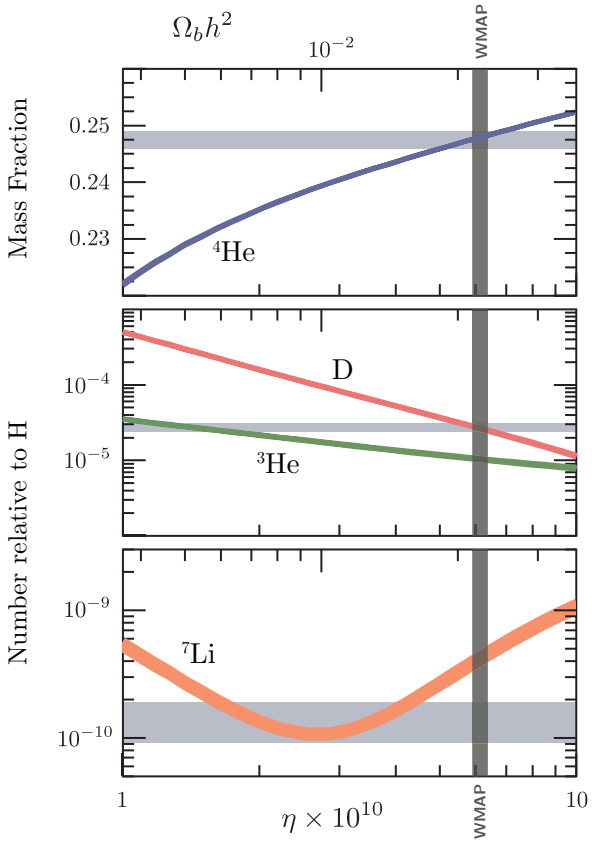


Figure 3.10: *Theoretical predictions (colored bands) and observational constraints (grey bands).*

- τ_n: a large neutron lifetime would reduce the amount of neutron decay after freeze-out and therefore would increase the final helium abundance.
- $\mathcal{Q}$: a larger mass difference between neutrons and protons would decrease the n/p ratio at freeze-out and therefore would decrease the final helium abundance.
- η: the amount of helium increases with increasing η as nucleosythesis starts earlier for larger baryon density.
- G_N: increasing the strength of gravity would increase the freeze-out temperature, $T_f \propto G_N^{1/6}$, and hence would increase the final helium abundance.
- G_F: increasing the weak force would decrease the freeze-out temperature, $T_f \propto G_F^{-2/3}$, and hence would decrease the final helium abundance.

Changing the input, e.g. by new physics beyond the Standard Model (BSM) in the early universe, would change the predictions of BBN. In this way BBN is a probe of fundamental physics.

Light Element Synthesis*

To determine the abundances of other light elements, the *coupled* Boltzmann equations have to be solved numerically (see fig. 3.11 for the result of such a computation). Fig. 3.10 shows that theoretical predictions for the light element abundances as a function of η (or Ω_b). The fact that we find reasonably good quantitative agreement with observations is one of the great triumphs of the Big Bang model.

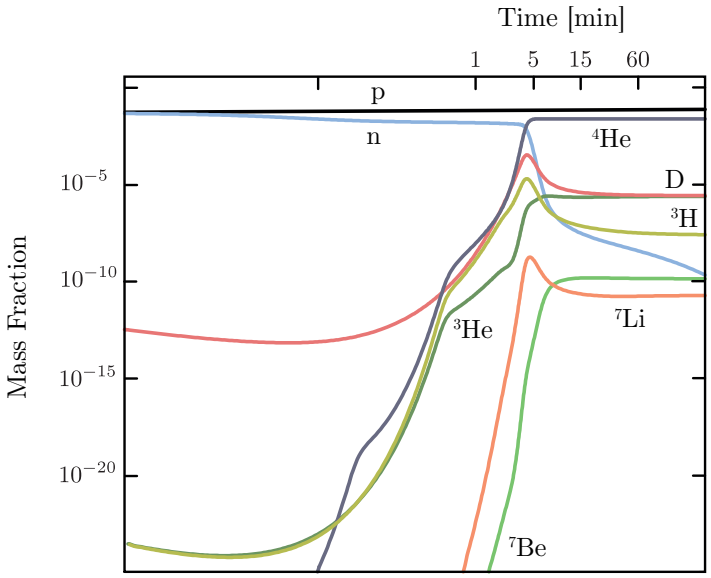


Figure 3.11: *Numerical results for the evolution of light element abundances.*

The shape of the curves in fig. 3.11 can easily be understood: The abundance of ^{4}He increases with increasing η as nucleosythesis starts earlier for larger baryon density. D and ^{3}He are burnt by fusion, thus their abundances decrease as η increases. Finally, ^{7}Li is destroyed by protons at low η with an efficiency that increases with η. On the other hand, its precursor ^{7}Be is produced more efficiently as η increases. This explains the valley in the curve for ^{7}Li.

Part II

The Inhomogeneous Universe

4 Cosmological Perturbation Theory

So far, we have treated the universe as perfectly homogeneous. To understand the formation and evolution of large-scale structures, we have to introduce inhomogeneities. As long as these perturbations remain relatively small, we can treat them in perturbation theory. In particular, we can expand the Einstein equations order-by-order in perturbations to the metric and the stress tensor. This makes the complicated system of coupled PDEs manageable.

4.1 Newtonian Perturbation Theory

Newtonian gravity is an adequate description of general relativity on scales well inside the Hubble radius and for non-relativistic matter (e.g. cold dark matter and baryons after decoupling). We will start with Newtonian perturbation theory because it is more intuitive than the full treatment in GR.

4.1.1 Perturbed Fluid Equations

Consider a non-relativistic fluid with *mass* density ρ, pressure $P \ll \rho$ and velocity $\boldsymbol{u}$. Denote the position vector of a fluid element by $\boldsymbol{r}$ and time by t. The equations of motion are given by basic fluid dynamics.[1] Mass conservation implies the *continuity equation*

$$\partial_t \rho = -\boldsymbol{\nabla}_{\boldsymbol{r}} \cdot (\rho \boldsymbol{u}) \ , \tag{4.1.1}$$

while momentum conservation leads to the *Euler equation*

$$(\partial_t + \boldsymbol{u} \cdot \boldsymbol{\nabla}_{\boldsymbol{r}})\, \boldsymbol{u} = -\frac{\boldsymbol{\nabla}_{\boldsymbol{r}} P}{\rho} - \boldsymbol{\nabla}_{\boldsymbol{r}} \Phi \ . \tag{4.1.2}$$

The last equation is simply "$F = ma$" for a fluid element. The gravitational potential Φ is determined by the *Poisson equation*

$$\nabla_{\boldsymbol{r}}^2 \Phi = 4\pi G \rho \ . \tag{4.1.3}$$

*Convective derivative.**—Notice that the acceleration in (4.1.2) is not given by $\partial_t \boldsymbol{u}$ (which measures how the velocity changes at a given position), but by the "convective time derivative" $D_t \boldsymbol{u} \equiv (\partial_t + \boldsymbol{u} \cdot \boldsymbol{\nabla})\, \boldsymbol{u}$ which follows the fluid element as it moves. Let me remind you how this comes about.

Consider a fixed volume in space. The total mass in the volume can only change if there is a flux of momentum through the surface. Locally, this is what the continuity equation describes: $\partial_t \rho + \nabla_j (\rho u_j) = 0$. Similarly, in the absence of any forces, the total momentum in the volume

[1] See Landau and Lifshitz, *Fluid Mechanics*.

can only change if there is a flux through the surface: $\partial_t(\rho u_i) + \nabla_j(\rho u_i u_j) = 0$. Expanding the derivatives, we get

$$\begin{aligned}\partial_t(\rho u_i) + \nabla_j(\rho u_i u_j) &= \rho\left[\partial_t + u_j\nabla_j\right]u_i + u_i\underbrace{\left[\partial_t\rho + \nabla_j(\rho u_j)\right]}_{=0} \\ &= \rho\left[\partial_t + u_j\nabla_j\right]u^i \ .\end{aligned}$$

In the absence of forces it is therefore the convective derivative of the velocity, $D_t\boldsymbol{u}$, that vanishes, not $\partial_t\boldsymbol{u}$. Adding forces gives the Euler equation.

We wish to see what these equation imply for the evolution of small perturbations around a homogeneous background. We therefore decompose all quantities into background values (denoted by an overbar) and perturbations—e.g. $\rho(t,\boldsymbol{r}) = \bar\rho(t) + \delta\rho(t,\boldsymbol{r})$, and similarly for the pressure, the velocity and the gravitational potential. Assuming that the fluctuations are small, we can linearise eqs. (4.1.1) and (4.1.2), i.e. we can drop products of fluctuations.

Static space without gravity

Let us first consider static space and ignore gravity ($\Phi \equiv 0$). It is easy to see that a solution for the background is $\bar\rho = const.$, $\bar P = const.$ and $\bar{\boldsymbol{u}} = 0$. The linearised evolution equations for the fluctuations are

$$\partial_t\delta\rho = -\boldsymbol{\nabla}_{\boldsymbol{r}}\cdot(\bar\rho\boldsymbol{u}) \ , \tag{4.1.4}$$

$$\bar\rho\,\partial_t\boldsymbol{u} = -\boldsymbol{\nabla}_{\boldsymbol{r}}\delta P \ . \tag{4.1.5}$$

Combining ∂_t(4.1.4) and $\boldsymbol{\nabla}_{\boldsymbol{r}}\cdot$(4.1.5), one finds

$$\partial_t^2\delta\rho - \nabla_{\boldsymbol{r}}^2\delta P = 0 \ . \tag{4.1.6}$$

For adiabatic fluctuations (see below), the pressure fluctuations are proportional to the density fluctuations, $\delta P = c_s^2\delta\rho$, where c_s is called the *speed of sound*. Eq. (4.1.6) then takes the form of a wave equation

$$\left(\partial_t^2 - c_s^2\nabla^2\right)\delta\rho = 0 \ . \tag{4.1.7}$$

This is solved by a plane wave, $\delta\rho = A\exp[i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})]$, where $\omega = c_s k$, with $k \equiv |\boldsymbol{k}|$. We see that in a static spacetime *fluctuations oscillate with constant amplitude* if we ignore gravity.

Fourier space.—The more formal way to solve PDEs like (4.1.7) is to expand $\delta\rho$ in terms of its Fourier components

$$\delta\rho(t,\boldsymbol{r}) = \int\frac{\mathrm{d}^3k}{(2\pi)^3}\,e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\delta\rho_{\boldsymbol{k}}(t) \ . \tag{4.1.8}$$

The PDE (4.1.7) turns into an ODE for each Fourier mode

$$\left(\partial_t^2 + c_s^2k^2\right)\delta\rho_{\boldsymbol{k}} = 0 \ , \tag{4.1.9}$$

which has the solution

$$\delta\rho_{\boldsymbol{k}} = A_{\boldsymbol{k}}e^{i\omega_k t} + B_{\boldsymbol{k}}e^{-i\omega_k t} \ , \qquad \omega_k \equiv c_s k \ . \tag{4.1.10}$$

Static space with gravity

Now we turn on gravity. Eq. (4.1.7) then gets a source term

$$\left(\partial_t^2 - c_s^2 \nabla_{\boldsymbol{r}}^2\right) \delta\rho = 4\pi G \bar{\rho} \delta\rho \ , \tag{4.1.11}$$

where we have used the perturbed Poisson equation, $\nabla^2 \delta\Phi = 4\pi G \delta\rho$. This is still solved by $\delta\rho = A \exp[i(\omega t - \boldsymbol{k} \cdot \boldsymbol{r})]$, but now with

$$\omega^2 = c_s^2 k^2 - 4\pi G \bar{\rho} \ . \tag{4.1.12}$$

We see that there is a critical wavenumber for which the frequency of oscillations is zero:

$$k_{\rm J} \equiv \frac{\sqrt{4\pi G \bar{\rho}}}{c_s} \ . \tag{4.1.13}$$

For small scales (i.e. large wavenumber), $k > k_{\rm J}$, the pressure dominates and we find the same oscillations as before. However, on large scales, $k < k_{\rm J}$, gravity dominates, the frequency ω becomes imaginary and the *fluctuations grow exponentially.* The crossover happens at the *Jeans' length*

$$\lambda_{\rm J} = \frac{2\pi}{k_{\rm J}} = c_s \sqrt{\frac{\pi}{G\bar{\rho}}} \ . \tag{4.1.14}$$

Expanding space

In an expanding space, we have the usual relationship between physical coordinates $\boldsymbol{r}$ and comoving coordinates $\boldsymbol{x}$,

$$\boldsymbol{r}(t) = a(t) \boldsymbol{x} \ . \tag{4.1.15}$$

The velocity field is then given by

$$\boldsymbol{u}(t) = \dot{\boldsymbol{r}} = H \boldsymbol{r} + \boldsymbol{v} \ , \tag{4.1.16}$$

where $H\boldsymbol{r}$ is the Hubble flow and $\boldsymbol{v} = a\dot{\boldsymbol{x}}$ is the proper velocity. In a static spacetime, the time and space derivates defined from t and $\boldsymbol{r}$ were independent. In an expanding spacetime this is not the case anymore. It is then convenient to use space derivatives defined with respect to the comoving coordinates $\boldsymbol{x}$, which we denote by $\boldsymbol{\nabla}_{\boldsymbol{x}}$. Using (4.1.15), we have

$$\boldsymbol{\nabla}_{\boldsymbol{r}} = a^{-1} \boldsymbol{\nabla}_{\boldsymbol{x}} \ . \tag{4.1.17}$$

The relationship between time derivatives at fixed $\boldsymbol{r}$ and at fixed $\boldsymbol{x}$ is

$$\begin{aligned} \left(\frac{\partial}{\partial t}\right)_{\boldsymbol{r}} &= \left(\frac{\partial}{\partial t}\right)_{\boldsymbol{x}} + \left(\frac{\partial \boldsymbol{x}}{\partial t}\right)_{\boldsymbol{r}} \cdot \boldsymbol{\nabla}_{\boldsymbol{x}} = \left(\frac{\partial}{\partial t}\right)_{\boldsymbol{x}} + \left(\frac{\partial a^{-1}(t) \boldsymbol{r}}{\partial t}\right)_{\boldsymbol{r}} \cdot \boldsymbol{\nabla}_{\boldsymbol{x}} \\ &= \left(\frac{\partial}{\partial t}\right)_{\boldsymbol{x}} - H \boldsymbol{x} \cdot \boldsymbol{\nabla}_{\boldsymbol{x}} \ . \end{aligned} \tag{4.1.18}$$

From now on, we will drop the subscripts $\boldsymbol{x}$.

With this in mind, let us look at the fluid equations in an expanding universe:

- *Continuity equation*

  Substituting (4.1.17) and (4.1.18) for $\boldsymbol{\nabla}_{\boldsymbol{r}}$ and ∂_t in the continuity equation (4.1.1), we get

  $$\left[\frac{\partial}{\partial t} - H\boldsymbol{x}\cdot\boldsymbol{\nabla}\right]\left[\bar{\rho}(1+\delta)\right] + \frac{1}{a}\boldsymbol{\nabla}\cdot\left[\bar{\rho}(1+\delta)(Ha\boldsymbol{x}+\boldsymbol{v})\right] = 0\ , \tag{4.1.19}$$

  Here, I have introduced the *fractional density perturbation*

  $$\delta \equiv \frac{\delta\rho}{\bar{\rho}}\ . \tag{4.1.20}$$

  Sometimes δ is called the *density contrast*.

  Let us analyse this order-by-order in perturbation theory:

  – At zeroth order in fluctuations (i.e. dropping the perturbations δ and $\boldsymbol{v}$), we have

  $$\frac{\partial\bar{\rho}}{\partial t} + 3H\bar{\rho} = 0\ , \tag{4.1.21}$$

  where I have used $\boldsymbol{\nabla}_{\boldsymbol{x}}\cdot\boldsymbol{x} = 3$. We recognise this as the continuity equation for the homogeneous *mass* density, $\bar{\rho}\propto a^{-3}$.

  – At first order in fluctuations (i.e. dropping products of δ and $\boldsymbol{v}$), we get

  $$\left[\frac{\partial}{\partial t} - H\boldsymbol{x}\cdot\boldsymbol{\nabla}\right]\left[\bar{\rho}\delta\right] + \frac{1}{a}\boldsymbol{\nabla}\cdot\left[\bar{\rho}Ha\boldsymbol{x}\,\delta + \bar{\rho}\boldsymbol{v}\right] = 0\ , \tag{4.1.22}$$

  which we can write as

  $$\left[\frac{\partial\bar{\rho}}{\partial t} + 3H\bar{\rho}\right]\delta + \bar{\rho}\,\frac{\partial\delta}{\partial t} + \frac{\bar{\rho}}{a}\,\boldsymbol{\nabla}\cdot\boldsymbol{v} = 0\ . \tag{4.1.23}$$

  The first term vanishes by (4.1.21), so we find

  $$\boxed{\dot{\delta} \;=\; -\frac{1}{a}\boldsymbol{\nabla}\cdot\boldsymbol{v}}\ , \tag{4.1.24}$$

  where we have used an overdot to denote the derivative with respect to time.

- *Euler equation*

  Similar manipulations of the Euler equation (4.1.2) lead to

  $$\boxed{\dot{\boldsymbol{v}} + H\boldsymbol{v} \;=\; -\frac{1}{a\bar{\rho}}\boldsymbol{\nabla}\delta P - \frac{1}{a}\boldsymbol{\nabla}\delta\Phi}\ . \tag{4.1.25}$$

  In the absence of pressure and gravitational perturbations, this equation simply says that $\boldsymbol{v}\propto a^{-1}$, which is something we already discovered in Chapter 1.

- *Poisson equation*

  It takes hardly any work to show that the Poisson equation (4.1.3) becomes

  $$\boxed{\boldsymbol{\nabla}^2\delta\Phi \;=\; 4\pi G a^2\bar{\rho}\delta}\ . \tag{4.1.26}$$

Exercise.—Derive eq. (4.1.25).

4.1.2 Jeans' Instability

Combining ∂_t(4.1.24) with $\boldsymbol{\nabla}\cdot$(4.1.25) and (4.1.26), we find

$$\boxed{\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2}\nabla^2\delta = 4\pi G\bar{\rho}\delta} \ . \tag{4.1.27}$$

This implies the same Jeans' length as in (4.1.14), but unlike the case of a static spacetime, it now depends on time via $\bar{\rho}(t)$ and $c_s(t)$. Compared to (4.1.11), the equation of motion in the expanding spacetime includes a friction term, $2H\dot{\delta}$. This has two effects: Below the Jeans' length, the fluctuations oscillate with decreasing amplitude. Above the Jeans' length, the *fluctuations experience power-law growth*, rather than the exponential growth we found for static space.

4.1.3 Dark Matter inside Hubble

The Newtonian framework describes the evolution of matter fluctuations. We can apply it to the evolution dark matter on sub-Hubble scales. (We will ignore small effects due to baryons.)

- During the *matter-dominated era*, eq. (4.1.27) reads

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G\bar{\rho}_m\delta_m = 0 \ , \tag{4.1.28}$$

  where we have dropped the pressure term, since $c_s = 0$ for linearised CDM fluctuations. (Non-linear effect produce a finite, but small, sound speed.) Since $a \propto t^{2/3}$, we have $H = 2/3t$ and hence

$$\ddot{\delta}_m + \frac{4}{3t}\dot{\delta}_m - \frac{2}{3t^2}\delta_m = 0 \ , \tag{4.1.29}$$

  where we have used $4\pi G\bar{\rho}_m = \frac{3}{2}H^2$. Trying $\delta_m \propto t^p$ gives the following two solutions:

$$\delta_m \propto \begin{cases} t^{-1} \ \propto \ a^{-3/2} \\ t^{2/3} \ \propto \ a \end{cases} \ . \tag{4.1.30}$$

  Hence, the *growing mode* of dark matter fluctuations grows like the scale factor during the MD era. This is a famous result that is worth remembering.

- During the *radiation-dominated era*, eq. (4.1.27) gets modified to

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G\sum_I \bar{\rho}_I\delta_I = 0 \ , \tag{4.1.31}$$

  where the sum is over matter and radiation. (It is the *total* density fluctuation $\delta\rho = \delta\rho_m + \delta\rho_r$ which sources $\delta\Phi$!) Radiation fluctuations on scales smaller than the Hubble radius oscillate as sound waves (supported by large radiation pressure) and their time-averaged density contrast vanishes. To prove this rigorously requires relativistic perturbation theory (see below). It follows that the CDM is essentially the only clustered component during the acoustic oscillations of the radiation, and so

$$\ddot{\delta}_m + \frac{1}{t}\dot{\delta}_m - 4\pi G\bar{\rho}_m\delta_m \approx 0 \ . \tag{4.1.32}$$

Since δ_m evolves only on cosmological timescales (it has no pressure support for it to do otherwise), we have

$$\ddot{\delta}_m \sim H^2 \delta_m \sim \frac{8\pi G}{3} \bar{\rho}_r \delta_m \gg 4\pi G \bar{\rho}_m \delta_m \ , \tag{4.1.33}$$

where we have used that $\bar{\rho}_r \gg \bar{\rho}_m$. We can therefore ignore the last term in (4.1.32) compared to the others. We then find

$$\delta_m \propto \begin{cases} const. \\ \ln t \ \propto \ \ln a \end{cases} . \tag{4.1.34}$$

We see that the rapid expansion due to the effectively unclustered radiation reduces the growth of δ_m to only logarithmic. This is another fact worth remembering: we need to wait until the universe becomes matter dominated in order for the dark matter density fluctuations to grow significantly.

- During the Λ-*dominated era*, eq. (4.1.27) reads

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G \sum_i \bar{\rho}_I \delta_I = 0 \ , \tag{4.1.35}$$

where $I = m, \Lambda$. As far as we can tell, dark energy doesn't cluster (almost by definition), so we can write

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G \bar{\rho}_m \delta_m = 0 \ , \tag{4.1.36}$$

Notice that this is *not* the same as (4.1.28), because H is different. Indeed, in the Λ-dominated regime $H^2 \approx const. \gg 4\pi G \bar{\rho}_m$. Dropping the last term in (4.1.36), we get

$$\ddot{\delta}_m + 2H\dot{\delta}_m \approx 0 \ , \tag{4.1.37}$$

which has the following solutions

$$\delta_m \propto \begin{cases} const. \\ e^{-2Ht} \ \propto \ a^{-2} \end{cases} . \tag{4.1.38}$$

We see that the matter fluctuations stop growing once dark energy comes to dominate.

4.2 Relativistic Perturbation Theory

The Newtonian treatment of cosmological perturbations is inadequate on scales larger than the Hubble radius, and for relativistic fluids (like photons and neutrinos). The correct description requires a full general-relativistic treatment which we will now develop.

4.2.1 Perturbed Spacetime

The basic idea is to consider small perturbations $\delta g_{\mu\nu}$ around the FRW metric $\bar{g}_{\mu\nu}$,

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu} \ . \tag{4.2.39}$$

Through the Einstein equations, the metric perturbations will be coupled to perturbations in the matter distribution.

Perturbations of the Metric

To avoid unnecessary technical distractions, we will only present the case of a flat FRW background spacetime

$$\mathrm{d}s^2 = a^2(\tau)\Big[\mathrm{d}\tau^2 - \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\Big] \ . \tag{4.2.40}$$

The perturbed metric can then be written as

$$\mathrm{d}s^2 = a^2(\tau)\Big[(1+2A)\mathrm{d}\tau^2 - 2B_i\,\mathrm{d}x^i\mathrm{d}\tau - (\delta_{ij}+h_{ij})\mathrm{d}x^i\mathrm{d}x^j\Big] \ , \tag{4.2.41}$$

where A, B_i and h_{ij} are functions of space and time. We shall adopt the useful convention that Latin indices on spatial vectors and tensors are raised and lowered with δ_{ij}, e.g. $h^i{}_i = \delta^{ij}h_{ij}$.

Scalar, Vectors and Tensors

It will be extremely useful to perform a scalar-vector-tensor (SVT) decomposition of the perturbations. For 3-vectors, this should be familiar. It simply means that we can split any 3-vector into the gradient of a scalar and a divergenceless vector

$$B_i = \underbrace{\partial_i B}_{\text{scalar}} + \underbrace{\hat{B}_i}_{\text{vector}} \ , \tag{4.2.42}$$

with $\partial^i\hat{B}_i = 0$. Similarly, any rank-2 symmetric tensor can be written

$$h_{ij} = \underbrace{2C\delta_{ij} + 2\partial_{\langle i}\partial_{j\rangle}E}_{\text{scalar}} + \underbrace{2\partial_{(i}\hat{E}_{j)}}_{\text{vector}} + \underbrace{2\hat{E}_{ij}}_{\text{tensor}} \ , \tag{4.2.43}$$

where

$$\partial_{\langle i}\partial_{j\rangle}E \equiv \Big(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\Big)E \ , \tag{4.2.44}$$

$$\partial_{(i}\hat{E}_{j)} \equiv \frac{1}{2}\Big(\partial_i\hat{E}_j + \partial_j\hat{E}_i\Big) \ . \tag{4.2.45}$$

As before, the hatted quantities are divergenceless, i.e. $\partial^i\hat{E}_i = 0$ and $\partial^i\hat{E}_{ij} = 0$. The tensor perturbation is traceless, $\hat{E}^i{}_i = 0$. The 10 degrees of freedom of the metric have thus been decomposed into $4+4+2$ SVT degrees of freedom:

- *scalars:* A, B, C, E
- *vectors:* $\hat{B}_i$, $\hat{E}_i$
- *tensors:* $\hat{E}_{ij}$

What makes the SVT-decomposition so powerful is the fact that the Einstein equations for scalars, vectors and tensors don't mix at linear order and can therefore be treated separately. In these lectures, we will mostly be interested in scalar fluctuations and the associated density perturbations. Vector perturbations aren't produced by inflation and even if they were, they would decay quickly with the expansion of the universe. Tensor perturbations are an important prediction of inflation and we will discuss them briefly in Chapter 6.

The Gauge Problem

Before we continue, we have to address an important subtlety. The metric perturbations in (4.2.41) aren't uniquely defined, but depend on our choice of coordinates or the *gauge choice*. In particular, when we wrote down the perturbed metric, we implicitly chose a specific time slicing of the spacetime and defined specfic spatial coordinates on these time slices. Making a different choice of coordinates, can change the values of the perturbation variables. It may even introduce fictitious perturbations. These are fake perturbations that can arise by an inconvenient choice of coordinates even if the background is perfectly homogeneous.

For example, consider the homogeneous FRW spacetime (4.2.40) and make the following change of the spatial coordinates, $x^i \mapsto \tilde{x}^i = x^i + \xi^i(\tau, \boldsymbol{x})$. We assume that ξ^i is small, so that it can also be treated as a perturbation. Using $\mathrm{d}x^i = \mathrm{d}\tilde{x}^i - \partial_\tau \xi^i \mathrm{d}\tau - \partial_k \xi^i \mathrm{d}\tilde{x}^k$, eq. (4.2.40) becomes

$$\mathrm{d}s^2 = a^2(\tau)\left[\mathrm{d}\tau^2 - 2\xi_i' \mathrm{d}\tilde{x}^i \mathrm{d}\tau - \left(\delta_{ij} + 2\partial_{(i}\xi_{j)}\right)\mathrm{d}\tilde{x}^i \mathrm{d}\tilde{x}^j\right] \ , \tag{4.2.46}$$

where we have dropped terms that are quadratic in ξ^i and defined $\xi_i' \equiv \partial_\tau \xi_i$. We apparently have introduced the metric perturbations $B_i = \xi_i'$ and $\hat{E}_i = \xi_i$. But these are just fictitious *gauge modes* that can be removed by going back to the old coordinates.

Similar, we can change our time slicing, $\tau \mapsto \tau + \xi^0(\tau, \boldsymbol{x})$. The homogeneous density of the universe then gets perturbed, $\rho(\tau) \mapsto \rho(\tau + \xi^0(\tau, \boldsymbol{x})) = \bar{\rho}(\tau) + \bar{\rho}'\xi^0$. So even in an unperturbed universe, a change of the time coordinate can introduce a fictitious density perturbation

$$\delta\rho = \bar{\rho}'\xi^0 \ . \tag{4.2.47}$$

Similarly, we can remove a real perturbation in the energy density by choosing the hypersurface of constant time to coincide with the hypersurface of constant energy density. Then $\delta\rho = 0$ although there are real inhomogeneities.

These examples illustrate that we need a more physical way to identify true perturbations. One way to do this is to define perturbations in such a way that they don't change under a change of coordinates.

Gauge Transformations

Consider the coordinate transformation

$$X^\mu \ \mapsto \ \tilde{X}^\mu \equiv X^\mu + \xi^\mu(\tau, \boldsymbol{x}) \ , \qquad \text{where} \qquad \xi^0 \equiv T \ , \quad \xi^i \equiv L^i = \partial^i L + \hat{L}^i \ . \tag{4.2.48}$$

We have split the spatial shift L^i into a scalar, L, and a divergenceless vector, $\hat{L}^i$. We wish to know how the metric transforms under this change of coordinates. The trick is to exploit the invariance of the spacetime interval,

$$\mathrm{d}s^2 = g_{\mu\nu}(X)\mathrm{d}X^\mu \mathrm{d}X^\nu = \tilde{g}_{\alpha\beta}(\tilde{X})\mathrm{d}\tilde{X}^\alpha \mathrm{d}\tilde{X}^\beta \ , \tag{4.2.49}$$

where I have used a different set of dummy indices on both sides to make the next few lines clearer. Writing $\mathrm{d}\tilde{X}^\alpha = (\partial\tilde{X}^\alpha/\partial X^\mu)\mathrm{d}X^\mu$ (and similarly for $\mathrm{d}X^\beta$), we find

$$g_{\mu\nu}(X) = \frac{\partial\tilde{X}^\alpha}{\partial X^\mu}\frac{\partial\tilde{X}^\beta}{\partial X^\nu}\,\tilde{g}_{\alpha\beta}(\tilde{X}) \ . \tag{4.2.50}$$

This relates the metric in the old coordinates, $g_{\mu\nu}$, to the metric in the new coordinates, $\tilde{g}_{\alpha\beta}$.

Let us see what (4.2.50) implies for the transformation of the metric perturbations in (4.2.41). I will work out the 00-component as an example and leave the rest as an exercise. Consider $\mu = \nu = 0$ in (4.2.50):

$$g_{00}(X) = \frac{\partial \tilde{X}^\alpha}{\partial \tau} \frac{\partial \tilde{X}^\beta}{\partial \tau} \tilde{g}_{\alpha\beta}(\tilde{X}) \ . \tag{4.2.51}$$

The only term that contributes to the l.h.s. is the one with $\alpha = \beta = 0$. Consider for example $\alpha = 0$ and $\beta = i$. The off-diagonal component of the metric $\tilde{g}_{0i}$ is proportional to $\tilde{B}_i$, so it is a first-order perturbation. But $\partial \tilde{X}^i / \partial \tau$ is proportional to the first-order variable ξ^i, so the product is second order and can be neglected. A similar argument holds for $\alpha = i$ and $\beta = j$. Eq. (4.2.51) therefore reduces to

$$g_{00}(X) = \left(\frac{\partial \tilde{\tau}}{\partial \tau} \right)^2 \tilde{g}_{00}(\tilde{X}) \ . \tag{4.2.52}$$

Substituting (4.2.48) and (4.2.41), we get

$$\begin{aligned} a^2(\tau)\big(1 + 2A\big) &= \big(1 + T'\big)^2 a^2(\tau + T)\big(1 + 2\tilde{A}\big) \\ &= \big(1 + 2T' + \cdots\big)\big(a(\tau) + a'T + \cdots\big)^2\big(1 + 2\tilde{A}\big) \\ &= a^2(\tau)\big(1 + 2\mathcal{H}T + 2T' + 2\tilde{A} + \cdots\big) \ , \end{aligned} \tag{4.2.53}$$

where $\mathcal{H} \equiv a'/a$ is the Hubble parameter in conformal time. Hence, we find that at first order, the metric perturbation A transforms as

$$A \mapsto \tilde{A} = A - T' - \mathcal{H}T \ . \tag{4.2.54}$$

I leave it to you to repeat the argument for the other metric components and show that

$$B_i \mapsto \tilde{B}_i = B_i + \partial_i T - L_i' \ , \tag{4.2.55}$$

$$h_{ij} \mapsto \tilde{h}_{ij} = h_{ij} - 2\partial_{(i} L_{j)} - 2\mathcal{H}T\delta_{ij} \ . \tag{4.2.56}$$

Exercise.—Derive eqs. (4.2.55) and (4.2.56).

In terms of the SVT-decomposition, we get

$$A \mapsto A - T' - \mathcal{H}T \ , \tag{4.2.57}$$

$$B \mapsto B + T - L' \ , \qquad \hat{B}_i \mapsto \hat{B}_i - \hat{L}_i' \ , \tag{4.2.58}$$

$$C \mapsto C - \mathcal{H}T - \frac{1}{3}\nabla^2 L \ , \tag{4.2.59}$$

$$E \mapsto E - L \ , \qquad \hat{E}_i \mapsto \hat{E}_i - \hat{L}_i \ , \qquad \hat{E}_{ij} \mapsto \hat{E}_{ij} \ . \tag{4.2.60}$$

Gauge-Invariant Perturbations

One way to avoid the gauge problems is to define special combinations of metric perturbations that do not transform under a change of coordinates. These are the *Bardeen variables*:

$$\Psi \equiv A + \mathcal{H}(B - E') + (B - E')' \ , \qquad \hat{\Phi}_i \equiv \hat{E}_i' - \hat{B}_i \ , \qquad \hat{E}_{ij} \ , \tag{4.2.61}$$

$$\Phi \equiv -C - \mathcal{H}(B - E') + \frac{1}{3}\nabla^2 E \ . \tag{4.2.62}$$

Exercise.—Show that Ψ, Φ and $\hat{\Phi}_i$ don't change under a coordinate transformation.

These gauge-invariant variables can be considered as the 'real' spacetime perturbations since they cannot be removed by a gauge transformation.

Gauge Fixing

An alternative (but related) solution to the gauge problem is to *fix the gauge* and keep track of *all* perturbations (metric and matter). For example, we can use the freedom in the gauge functions T and L in (4.2.48) to set two of the four scalar metric perturbations to zero:

- *Newtonian gauge.*—The choice
$$B = E = 0 \ , \tag{4.2.63}$$
gives the metric
$$\mathrm{d}s^2 = a^2(\tau)\left[(1+2\Psi)\mathrm{d}\tau^2 - (1-2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right] \ . \tag{4.2.64}$$
Here, we have renamed the remaining two metric perturbations, $A \equiv \Psi$ and $C \equiv -\Phi$, in order to make contact with the Bardeen potentials in (4.2.61) and (4.2.62). For perturbations that decay at spatial infinity, the Newtonian gauge is unique (i.e. the gauge is fixed completely).[2] In this gauge, the physics appears rather simple since the hypersurfaces of constant time are orthogonal to the worldlines of observers at rest in the coordinates (since $B = 0$) and the induced geometry of the constant-time hypersurfaces is isotropic (since $E = 0$). In the absence of anisotropic stress, $\Psi = \Phi$. Note the similarity of the metric to the usual weak-field limit of GR about Minkowski space; we shall see that Ψ plays the role of the gravitational potential. Newtonian gauge will be our preferred gauge for studying the formation of large-scale structures (Chapter 5) and CMB anisotropies (Chapter **??**).

- *Spatially-flat qauge.*—A convenient gauge for computing inflationary perturbations is
$$C = E = 0 \ . \tag{4.2.65}$$
In this gauge, we will be able to focus most directly on the fluctuations in the inflaton field $\delta\phi$ (see Chapter 6) .

4.2.2 Perturbed Matter

In Chapter 1, we showed that the matter in a homogeneous and isotropic universe has to take the form of a perfect fluid
$$\bar{T}^{\mu}{}_{\nu} = (\bar{\rho} + \bar{P})\bar{U}^{\mu}\bar{U}_{\nu} - \bar{P}\,\delta^{\mu}_{\nu} \ , \tag{4.2.66}$$
where $\bar{U}_{\mu} = a\delta^0_{\mu}$, $\bar{U}^{\mu} = a^{-1}\delta^{\mu}_0$ for a comoving observer. Now, we consider small perturbations of the stress-energy tensor
$$T^{\mu}{}_{\nu} = \bar{T}^{\mu}{}_{\nu} + \delta T^{\mu}{}_{\nu} \ . \tag{4.2.67}$$

[2] More generally, a gauge transformation that corresponds to a small, time-dependent but spatially constant boost – i.e. $L^i(\tau)$ and a compensating time translation with $\partial_i T = L_i(\tau)$ to keep the constant-time hypersurfaces orthogonal – will preserve $E_{ij} = 0$ and $B_i = 0$ and hence the form of the metric in eq. (4.4.168). However, such a transformation would not preserve the decay of the perturbations at infinity.

Perturbations of the Stress-Energy Tensor

In a perturbed universe, the energy density ρ, the pressure P and the four-velocity U^μ can be functions of position. Moreover, the stress-energy tensor can now have a contribution from *anisotropic stress*, $\Pi^\mu{}_\nu$. The perturbation of the stress-energy tensor is

$$\delta T^\mu{}_\nu = (\delta\rho + \delta P)\bar{U}^\mu\bar{U}_\nu + (\bar{\rho} + \bar{P})(\delta U^\mu \bar{U}_\nu + \bar{U}^\mu \delta U_\nu) - \delta P\,\delta^\mu_\nu - \Pi^\mu{}_\nu \ . \tag{4.2.68}$$

The spatial part of the anisotropic stress tensor can be chosen to be traceless, $\Pi^i{}_i = 0$, since its trace can always be absorbed into a redefinition of the isotropic pressure, P. The anisotropic stress tensor can also be chosen to be orthogonal to U^μ, i.e. $U^\mu\Pi_{\mu\nu} = 0$. Without loss of generality, we can then set $\Pi^0{}_0 = \Pi^0{}_i = 0$. In practice, the anisotropic stress will always be negligible in these lectures. We will keep it for now, but at some point we will drop it.

Perturbations in the four-velocity can induce non-vanishing *energy flux*, $T^0{}_j$, and *momentum density*, $T^i{}_0$. To find these, let us compute the perturbed four-velocity in the perturbed metric (4.2.41). Since $g_{\mu\nu}U^\mu U^\nu = 1$ and $\bar{g}_{\mu\nu}\bar{U}^\mu\bar{U}^\nu = 1$, we have, at linear order,

$$\delta g_{\mu\nu}\bar{U}^\mu\bar{U}^\nu + 2\bar{U}_\mu\delta U^\mu = 0 \ . \tag{4.2.69}$$

Using $\bar{U}^\mu = a^{-1}\delta^0_\mu$ and $\delta g_{00} = 2a^2A$, we find $\delta U^0 = -Aa^{-1}$. We then write $\delta U^i \equiv v^i/a$, where $v^i \equiv dx^i/d\tau$ is the *coordinate velocity*, so that

$$U^\mu = a^{-1}[1 - A, v^i] \ . \tag{4.2.70}$$

From this, we derive

$$U_0 = g_{00}U^0 + \overbrace{g_{0i}U^i}^{\mathcal{O}(2)} = a^2(1+2A)a^{-1}(1-A) = a(1+A) \ , \tag{4.2.71}$$

$$U_i = g_{i0}U^0 + g_{ij}U^j = -a^2B_ia^{-1} - a^2\delta_{ij}a^{-1}v^j = -a(B_i + v_i) \ , \tag{4.2.72}$$

i.e.

$$U_\mu = a[1 + A, -(v_i + B_i)] \ . \tag{4.2.73}$$

Using (4.2.70) and (4.2.73) in (4.2.68), we find

$$\delta T^0{}_0 = \delta\rho \ , \tag{4.2.74}$$

$$\delta T^i{}_0 = (\bar{\rho} + \bar{P})v^i \ , \tag{4.2.75}$$

$$\delta T^0{}_j = -(\bar{\rho} + \bar{P})(v_j + B_j) \ , \tag{4.2.76}$$

$$\delta T^i{}_j = -\delta P\delta^i_j - \Pi^i{}_j \ . \tag{4.2.77}$$

We will use q^i for the *momentum density* $(\bar{\rho} + \bar{P})v^i$. If there are several contributions to the stress-energy tensor (e.g. photons, baryons, dark matter, etc.), they are added: $T_{\mu\nu} = \sum_I T^I_{\mu\nu}$. This implies

$$\delta\rho = \sum_I \delta\rho_I \ , \quad \delta P = \sum_I \delta P_I \ , \quad q^i = \sum_I q^i_I \ , \quad \Pi^{ij} = \sum_I \Pi^{ij}_I \ . \tag{4.2.78}$$

We see that the perturbations in the density, pressure and anisotropic stress simply add. The velocities do *not* add, but the momentum densities do.

Finally, we note that the SVT decomposition can also be applied to the perturbations of the stress-energy tensor: $\delta\rho$ and δP have scalar parts only, q_i has scalar and vector parts,

$$q_i = \partial_i q + \hat{q}_i \ , \tag{4.2.79}$$

and Π_{ij} has scalar, vector and tensor parts,

$$\Pi_{ij} = \partial_{\langle i}\partial_{j\rangle}\Pi + \partial_{(i}\hat{\Pi}_{j)} + \hat{\Pi}_{ij} \ . \tag{4.2.80}$$

Gauge Transformations

Under the coordinate transformation (4.2.48), the stress-energy tensor transform as

$$T^{\mu}{}_{\nu}(X) = \frac{\partial X^{\mu}}{\partial \tilde{X}^{\alpha}}\frac{\partial \tilde{X}^{\beta}}{\partial X^{\nu}}\,\tilde{T}^{\alpha}{}_{\beta}(\tilde{X}) \ . \tag{4.2.81}$$

Evaluating this for the different components, we find

$$\delta\rho \mapsto \delta\rho - T\bar{\rho}' \ , \tag{4.2.82}$$

$$\delta P \mapsto \delta P - T\bar{P}' \ , \tag{4.2.83}$$

$$q_i \mapsto q_i + (\bar{\rho} + \bar{P})L_i' \ , \tag{4.2.84}$$

$$v_i \mapsto v_i + L_i' \ , \tag{4.2.85}$$

$$\Pi_{ij} \mapsto \Pi_{ij} \ . \tag{4.2.86}$$

Exercise.—Confirm eqs. (4.2.82)–(4.2.86). [Hint: First, convince yourself that the inverse of a matrix of the form $\mathbf{1} + \boldsymbol{\varepsilon}$, were $\mathbf{1}$ is the identity and $\boldsymbol{\varepsilon}$ is a small perturbation, is $\mathbf{1} - \boldsymbol{\varepsilon}$ to first order in $\boldsymbol{\varepsilon}$.]

Gauge-Invariant Perturbations

There are various gauge-invariant quantities that can be formed from metric and matter variables. One useful combination is

$$\bar{\rho}\Delta \equiv \delta\rho + \bar{\rho}'(v + B) \ , \tag{4.2.87}$$

where $v_i = \partial_i v$. The quantity Δ is called the *comoving-gauge density perturbation.*

Exercise.—Show that Δ is gauge-invariant.

Gauge Fixing

Above we used our gauge freedom to set two of the metric perturbations to zero. Alternatively, we can define the gauge in the matter sector:

- *Uniform density gauge.*—We can use the freedom in the time-slicing to set the total density perturbation to zero

$$\delta\rho = 0 \ . \tag{4.2.88}$$

- *Comoving gauge.*—Similarly, we can ask for the scalar momentum density to vanish,

$$q = 0 \ . \tag{4.2.89}$$

  Fluctuations in comoving gauge are most naturally connected to the inflationary initial conditions. This will be explained in §4.3.1 and Chapter 6.

There are different versions of uniform density and comoving gauge depending on which of the metric fluctuations is set to zero. In these lectures, we will choose $B = 0$.

Adiabatic Fluctuations

Simple inflation models predict initial fluctuations that are *adiabatic* (see Chapter 6). Adiabatic perturbations have the property that the local state of matter (determined, for example, by the energy density ρ and the pressure P) at some spacetime point $(\tau, \boldsymbol{x})$ of the perturbed universe is the same as in the *background* universe at some slightly different time $\tau + \delta\tau(\boldsymbol{x})$. (Notice that the time shift varies with location $\boldsymbol{x}$!) We can thus view adiabatic perturbations as some parts of the universe being "ahead" and others "behind" in the evolution. If the universe is filled with multiple fluids, adiabatic perturbations correspond to perturbations induced by a *common, local shift in time* of all background quantities; e.g. adiabatic density perturbations are defined as

$$\delta\rho_I(\tau, \boldsymbol{x}) \equiv \bar{\rho}_I(\tau + \delta\tau(\boldsymbol{x})) - \bar{\rho}_I(\tau) = \bar{\rho}_I' \, \delta\tau(\boldsymbol{x}) \ , \tag{4.2.90}$$

where $\delta\tau$ is the same for all species I. This implies

$$\delta\tau = \frac{\delta\rho_I}{\bar{\rho}_I'} = \frac{\delta\rho_J}{\bar{\rho}_J'} \qquad \text{for all species } I \text{ and } J \ . \tag{4.2.91}$$

Using[3] $\bar{\rho}_I' = -3\mathcal{H}(1 + w_I)\bar{\rho}_I$, we can write this as

$$\frac{\delta_I}{1 + w_I} = \frac{\delta_J}{1 + w_J} \qquad \text{for all species } I \text{ and } J \ , \tag{4.2.92}$$

where we have defined the *fractional density contrast*

$$\delta_I \equiv \frac{\delta\rho_I}{\bar{\rho}_I} \ . \tag{4.2.93}$$

Thus, for adiabatic perturbations, all matter components ($w_m \approx 0$) have the same fractional perturbation, while all radiation perturbations ($w_r = \frac{1}{3}$) obey

$$\delta_r = \frac{4}{3}\delta_m \ . \tag{4.2.94}$$

It follows that for adiabatic fluctuations, the total density perturbation,

$$\delta\rho_{\rm tot} = \bar{\rho}_{\rm tot}\delta_{\rm tot} = \sum_I \bar{\rho}_I \delta_I \ , \tag{4.2.95}$$

is dominated by the species that is dominant in the background since all the δ_I are comparable. We will have more to say about adiabatic initial conditions in §4.3.

[3] If there is no energy transfer between the fluid components at the background level, the energy continuity equation is satisfied by them separately.

Isocurvature Fluctuations

The complement of adiabatic perturbations are *isocurvature perturbations*. While adiabatic perturbations correspond to a change in the total energy density, isocurvature perturbations only correspond to perturbations between the different components. Eq. (4.2.92) suggests the following definition of isocurvature fluctuations

$$S_{IJ} \equiv \frac{\delta_I}{1+w_I} - \frac{\delta_J}{1+w_J} \ . \tag{4.2.96}$$

Single-field inflation predicts that the primordial perturbations are purely adiabatic, i.e. $S_{IJ} = 0$, for all species I and J. Moreover, all present observational data is consistent with this expectation. We therefore won't consider isocurvature fluctuations further in these lectures.

4.2.3 Linearised Evolution Equations

Our next task is to derive the perturbed Einstein equations, $\delta G_{\mu\nu} = 8\pi G \delta T_{\mu\nu}$, from the perturbed metric and the perturbed stress-energy tensor. We will work in Newtonian gauge with

$$g_{\mu\nu} = a^2 \begin{pmatrix} 1+2\Psi & 0 \\ 0 & -(1-2\Phi)\delta_{ij} \end{pmatrix} \ . \tag{4.2.97}$$

In these lectures, we will never encounter situations where anisotropic stress plays a significant role. From now on, we will therefore set anisotropic stress to zero, $\Pi_{ij} = 0$. As we will see, this enforces $\Phi = \Psi$.

Perturbed Connection Coefficients

To derive the field equations, we first require the perturbed connection coefficients. Recall that

$$\Gamma^\mu_{\nu\rho} = \frac{1}{2} g^{\mu\lambda} \left(\partial_\nu g_{\lambda\rho} + \partial_\rho g_{\lambda\nu} - \partial_\lambda g_{\nu\rho} \right) \ . \tag{4.2.98}$$

Since the metric (4.2.97) is diagonal, it is simple to invert

$$g^{\mu\nu} = \frac{1}{a^2} \begin{pmatrix} 1-2\Psi & 0 \\ 0 & -(1+2\Phi)\delta^{ij} \end{pmatrix} \ . \tag{4.2.99}$$

Substituting (4.2.97) and (4.2.99) into (4.2.98), gives

$$\Gamma^0_{00} = \mathcal{H} + \Psi' \ , \tag{4.2.100}$$

$$\Gamma^0_{0i} = \partial_i \Psi \ , \tag{4.2.101}$$

$$\Gamma^i_{00} = \delta^{ij} \partial_j \Psi \ , \tag{4.2.102}$$

$$\Gamma^0_{ij} = \mathcal{H}\delta_{ij} - \left[\Phi' + 2\mathcal{H}(\Phi + \Psi) \right] \delta_{ij} \ , \tag{4.2.103}$$

$$\Gamma^i_{j0} = \mathcal{H}\delta^i_j - \Phi' \delta^i_j \ , \tag{4.2.104}$$

$$\Gamma^i_{jk} = -2\delta^i_{(j}\partial_{k)}\Phi + \delta_{jk}\delta^{il}\partial_l \Phi \ . \tag{4.2.105}$$

I will work out Γ^0_{00} as an example and leave the remaining terms as an exercise.

Example.—From the definition of the Christoffel symbol we have

$$\begin{aligned}\Gamma^0_{00} &= \frac{1}{2}g^{00}(2\partial_0 g_{00} - \partial_0 g_{00}) \\ &= \frac{1}{2}g^{00}\partial_0 g_{00} \ .\end{aligned} \tag{4.2.106}$$

Substituting the metric components, we find

$$\begin{aligned}\Gamma^0_{00} &= \frac{1}{2a^2}(1-2\Psi)\partial_0[a^2(1+2\Psi)] \\ &= \mathcal{H} + \Psi' \ ,\end{aligned} \tag{4.2.107}$$

at linear order in Ψ.

Exercise.—Derive eqs. (4.2.101)–(4.2.105).

Perturbed Stress-Energy Conservation

Equipped with the perturbed connection, we can immediately derive the perturbed conservation equations from

$$\begin{aligned}\nabla_\mu T^\mu{}_\nu &= 0 \\ &= \partial_\mu T^\mu{}_\nu + \Gamma^\mu_{\mu\alpha}T^\alpha{}_\nu - \Gamma^\alpha_{\mu\nu}T^\mu{}_\alpha \ .\end{aligned} \tag{4.2.108}$$

Continuity Equation

Consider first the $\nu = 0$ component

$$\partial_0 T^0{}_0 + \partial_i T^i{}_0 + \Gamma^\mu_{\mu 0}T^0{}_0 + \underbrace{\Gamma^\mu_{\mu i}T^i{}_0}_{\mathcal{O}(2)} - \Gamma^0_{00}T^0{}_0 - \underbrace{\Gamma^0_{i0}T^i{}_0}_{\mathcal{O}(2)} - \underbrace{\Gamma^i_{00}T^0{}_i}_{\mathcal{O}(2)} - \Gamma^i_{j0}T^j{}_i = 0 \ . \tag{4.2.109}$$

Substituting the perturbed stress-energy tensor and the connection coefficients gives

$$\begin{aligned}&\partial_0(\bar\rho + \delta\rho) + \partial_i q^i + (\mathcal{H} + \Psi' + 3\mathcal{H} - 3\Phi')(\bar\rho + \delta\rho) \\ &\quad - (\mathcal{H} + \Psi')(\bar\rho + \delta\rho) - (\mathcal{H} - \Phi')\delta^i_j\big[-(\bar P + \delta P)\delta^j_i\big] = 0 \ ,\end{aligned} \tag{4.2.110}$$

and hence

$$\bar\rho' + \delta\rho' + \partial_i q^i + 3\mathcal{H}(\bar\rho + \delta\rho) - 3\bar\rho\Phi' + 3\mathcal{H}(\bar P + \delta P) - 3\bar P\Phi' = 0 \ . \tag{4.2.111}$$

Writing the zeroth-order and first-order parts separately, we get

$$\bar\rho' = -3\mathcal{H}(\bar\rho + \bar P) \ , \tag{4.2.112}$$

$$\delta\rho' = -3\mathcal{H}(\delta\rho + \delta P) + 3\Phi'(\bar\rho + \bar P) - \boldsymbol{\nabla}\cdot\boldsymbol{q} \ . \tag{4.2.113}$$

The zeroth-order part (4.2.112) simply is the conservation of energy in the homogeneous background. Eq. (4.2.113) describes the evolution of the density perturbation. The first term on the right-hand side is just the dilution due to the background expansion (as in the background

equation), the $\boldsymbol{\nabla}\cdot\boldsymbol{q}$ term accounts for the local fluid flow due to peculiar velocity, and the Φ' term is a purely relativistic effect corresponding to the density changes caused by perturbations to the local expansion rate [$(1-\Phi)a$ is the "local scale factor" in the spatial part of the metric in Newtonian gauge].

It is convenient to write the equation in terms of the fractional overdensity and the 3-velocity,

$$\delta \equiv \frac{\delta\rho}{\bar\rho} \qquad \text{and} \qquad \boldsymbol{v} = \frac{\boldsymbol{q}}{\bar\rho + \bar P} \ . \tag{4.2.114}$$

Eq. (4.2.113) then becomes

$$\boxed{\delta' + \left(1+\frac{\bar P}{\bar\rho}\right)\left(\boldsymbol{\nabla}\cdot\boldsymbol{v} - 3\Phi'\right) + 3\mathcal{H}\left(\frac{\delta P}{\delta\rho} - \frac{\bar P}{\bar\rho}\right)\delta = 0} \ . \tag{4.2.115}$$

This is the relativistic version of the *continuity equation*. In the limit $P \ll \rho$, we recover the Newtonian continuity equation in conformal time, $\delta' + \boldsymbol{\nabla}\cdot\boldsymbol{v} - 3\Phi' = 0$, but with a general-relativistic correction due to the perturbation to the rate of exansion of space. This correction is small on sub-horizon scales ($k \gg \mathcal{H}$) — we will prove this rigorously in Chapter 5.

Euler Equation

Next, consider the $\nu = i$ component of eq. (4.2.108),

$$\partial_\mu T^\mu{}_i + \Gamma^\mu_{\mu\rho} T^\rho{}_i - \Gamma^\rho{}_{\mu i} T^\mu{}_\rho = 0 \ , \tag{4.2.116}$$

and hence

$$\partial_0 T^0{}_i + \partial_j T^j{}_i + \Gamma^\mu_{\mu 0} T^0{}_i + \Gamma^\mu_{\mu j} T^j{}_i - \Gamma^0_{0i} T^0{}_0 - \Gamma^0_{ji} T^j{}_0 - \Gamma^j_{0i} T^0{}_j - \Gamma^j_{ki} T^k{}_j = 0 \ . \tag{4.2.117}$$

Using eqs. (4.2.74)–(4.2.77), with $T^0{}_i = -q_i$ in Newtonian gauge, eq. (4.2.117) becomes

$$\begin{aligned} -q_i' + \partial_j\left[-(\bar P + \delta P)\delta^j_i\right] - 4\mathcal{H}q_i - (\partial_j\Psi - 3\partial_j\Phi)\,\bar P\,\delta^j_i - \partial_i\Psi\bar\rho \\ -\mathcal{H}\delta_{ji}q^j + \mathcal{H}\delta^j_i q_j + \underbrace{\left(-2\delta^j_{(i}\partial_{k)}\Phi + \delta_{ki}\delta^{jl}\partial_l\Phi\right)\bar P\,\delta^k_j}_{-3\partial_i\Phi\,\bar P} \ = \ 0 \ , \end{aligned} \tag{4.2.118}$$

or

$$-q_i' - \partial_i\delta P - 4\mathcal{H}q_i - (\bar\rho + \bar P)\partial_i\Psi = 0 \, . \tag{4.2.119}$$

Using eqs. (4.2.112) and (4.2.114), we get

$$\boxed{\boldsymbol{v}' + \mathcal{H}\boldsymbol{v} - 3\mathcal{H}\frac{\bar P'}{\bar\rho'}\boldsymbol{v} = -\frac{\boldsymbol{\nabla}\delta P}{\bar\rho + \bar P} - \boldsymbol{\nabla}\Psi} \ . \tag{4.2.120}$$

This is the relativistic version of the *Euler equation* for a viscous fluid. Pressure gradients ($\boldsymbol{\nabla}\delta P$) and gravitational infall ($\boldsymbol{\nabla}\Psi$) drive $\boldsymbol{v}'$. The equation captures the redshifting of peculiar velocities ($\mathcal{H}\boldsymbol{v}$) and includes a small correction for relativistic fluids ($\bar P'/\bar\rho'$). Adiabatic fluctuations satisfy $\bar P'/\bar\rho' = c_s^2$. Non-relativistic matter fluctuations have a very small sound speed, so the relativistic correction in the Euler equation (4.2.120) is much smaller than the redshifting

term. The limit $P \ll \rho$ then reproduces the Euler equation (4.1.25) of the linearised Newtonian treatment.

Eqs. (4.2.115) and (4.2.120) apply for the total matter and velocity, and *also separately* for any non-interacting components so that the individual stress-energy tensors are separately conserved. Once an equation of state of the matter (and other constitutive relations) are specified, we just need the gravitational potentials Ψ and Φ to close the system of equations. Equations for Ψ and Φ follow from the perturbed Einstein equations.

Perturbed Einstein Equations

Let us now compute the linearised Einstein equation in Newtonian gauge. We require the perturbation to the Einstein tensor, $G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$, so we first need to calculate the perturbed Ricci tensor $R_{\mu\nu}$ and scalar R.

Ricci tensor.—We recall that the Ricci tensor can be expressed in terms of the connection as

$$R_{\mu\nu} = \partial_\lambda \Gamma^\lambda_{\mu\nu} - \partial_\nu \Gamma^\lambda_{\mu\lambda} + \Gamma^\lambda_{\lambda\rho}\Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\rho} \ . \tag{4.2.121}$$

Substituting the perturbed connection coefficients (4.2.100)–(4.2.105), we find

$$R_{00} = -3\mathcal{H}' + \nabla^2\Psi + 3\mathcal{H}(\Phi' + \Psi') + 3\Phi'' \ , \tag{4.2.122}$$

$$R_{0i} = 2\partial_i\Phi' + 2\mathcal{H}\partial_i\Psi \ , \tag{4.2.123}$$

$$R_{ij} = \left[\mathcal{H}' + 2\mathcal{H}^2 - \Phi'' + \nabla^2\Phi - 2(\mathcal{H}' + 2\mathcal{H}^2)(\Phi + \Psi) - \mathcal{H}\Psi' - 5\mathcal{H}\Phi'\right]\delta_{ij} + \partial_i\partial_j(\Phi - \Psi) \ . \tag{4.2.124}$$

I will derive R_{00} here and leave the others as an exercise.

Example.—The 00 component of the Ricci tensor is

$$R_{00} = \partial_\rho\Gamma^\rho_{00} - \partial_0\Gamma^\rho_{0\rho} + \Gamma^\alpha_{00}\Gamma^\rho_{\alpha\rho} - \Gamma^\alpha_{0\rho}\Gamma^\rho_{0\alpha} \ . \tag{4.2.125}$$

When we sum over ρ, the terms with $\rho = 0$ cancel so we need only consider summing over $\rho = 1, 2, 3$, i.e.

$$\begin{aligned} R_{00} &= \partial_i\Gamma^i_{00} - \partial_0\Gamma^i_{0i} + \Gamma^\alpha_{00}\Gamma^i_{\alpha i} - \Gamma^\alpha_{0i}\Gamma^i_{0\alpha} \\ &= \partial_i\Gamma^i_{00} - \partial_0\Gamma^i_{0i} + \Gamma^0_{00}\Gamma^i_{0i} + \underbrace{\Gamma^j_{00}\Gamma^i_{ji}}_{\mathcal{O}(2)} - \underbrace{\Gamma^0_{0i}\Gamma^i_{00}}_{\mathcal{O}(2)} - \Gamma^j_{0i}\Gamma^i_{0j} \\ &= \nabla^2\Psi - 3\partial_0(\mathcal{H} - \Phi') + 3(\mathcal{H} + \Psi')(\mathcal{H} - \Phi') - (\mathcal{H} - \Phi')^2\delta^j_i\delta^i_j \\ &= -3\mathcal{H}' + \nabla^2\Psi + 3\mathcal{H}(\Phi' + \Psi') + 3\Phi'' \ . \end{aligned} \tag{4.2.126}$$

Exercise.—Derive eqs. (4.2.123) and (4.2.124).

Ricci scalar.—It is now relatively straightforward to compute the Ricci scalar

$$R = g^{00}R_{00} + 2\underbrace{g^{0i}R_{0i}}_{0} + g^{ij}R_{ij} \ . \tag{4.2.127}$$

It follows that

$$\begin{aligned} a^2R &= (1-2\Psi)R_{00} - (1+2\Phi)\delta^{ij}R_{ij} \\ &= (1-2\Psi)\left[-3\mathcal{H}' + \nabla^2\Psi + 3\mathcal{H}(\Phi'+\Psi') + 3\Phi''\right] \\ &\quad - 3(1+2\Phi)\left[\mathcal{H}' + 2\mathcal{H}^2 - \Phi'' + \nabla^2\Phi - 2(\mathcal{H}'+2\mathcal{H}^2)(\Phi+\Psi) - \mathcal{H}\Psi' - 5\mathcal{H}\Phi'\right] \\ &\quad - (1+2\Phi)\nabla^2(\Phi-\Psi)\ . \end{aligned} \tag{4.2.128}$$

Dropping non-linear terms, we find

$$a^2R = -6(\mathcal{H}'+\mathcal{H}^2) + 2\nabla^2\Psi - 4\nabla^2\Phi + 12(\mathcal{H}'+\mathcal{H}^2)\Psi + 6\Phi'' + 6\mathcal{H}(\Psi'+3\Phi')\ . \tag{4.2.129}$$

Einstein tensor.—Computing the Einstein tensor is now just a matter of collecting our previous results. The 00 component is

$$\begin{aligned} G_{00} &= R_{00} - \frac{1}{2}g_{00}R \\ &= -3\mathcal{H}' + \nabla^2\Psi + 3\mathcal{H}(\Phi'+\Psi') + 3\Phi'' + 3(1+2\Psi)(\mathcal{H}'+\mathcal{H}^2) \\ &\quad - \frac{1}{2}\left[2\nabla^2\Psi - 4\nabla^2\Phi + 12(\mathcal{H}'+\mathcal{H}^2)\Psi + 6\Phi'' + 6\mathcal{H}(\Psi'+3\Phi')\right]\ . \end{aligned} \tag{4.2.130}$$

Most of the terms cancel leaving the simple result

$$G_{00} = 3\mathcal{H}^2 + 2\nabla^2\Phi - 6\mathcal{H}\Phi'\ . \tag{4.2.131}$$

The $0i$ component of the Einstein tensor is simply R_{0i} since $g_{0i} = 0$ in Newtonian gauge:

$$G_{0i} = 2\,\partial_i(\Phi' + \mathcal{H}\Psi)\ . \tag{4.2.132}$$

The remaining components are

$$\begin{aligned} G_{ij} &= R_{ij} - \frac{1}{2}g_{ij}R \\ &= \left[\mathcal{H}' + 2\mathcal{H}^2 - \Phi'' + \nabla^2\Phi - 2(\mathcal{H}'+2\mathcal{H}^2)(\Phi+\Psi) - \mathcal{H}\Psi' - 5\mathcal{H}\Phi'\right]\delta_{ij} + \partial_i\partial_j(\Phi-\Psi) \\ &\quad - 3(1-2\Phi)(\mathcal{H}'+\mathcal{H}^2)\delta_{ij} \\ &\quad + \frac{1}{2}\left[2\nabla^2\Psi - 4\nabla^2\Phi + 12(\mathcal{H}'+\mathcal{H}^2)\Psi + 6\Phi'' + 6\mathcal{H}(\Psi'+3\Phi')\right]\delta_{ij}\ . \end{aligned} \tag{4.2.133}$$

This neatens up (only a little!) to give

$$\begin{aligned} G_{ij} &= -(2\mathcal{H}'+\mathcal{H}^2)\delta_{ij} + \left[\nabla^2(\Psi-\Phi) + 2\Phi'' + 2(2\mathcal{H}'+\mathcal{H}^2)(\Phi+\Psi) + 2\mathcal{H}\Psi' + 4\mathcal{H}\Phi'\right]\delta_{ij} \\ &\quad + \partial_i\partial_j(\Phi-\Psi)\ . \end{aligned} \tag{4.2.134}$$

Einstein Equations

Substituting the perturbed Einstein tensor, metric and stress-energy tensor into the Einstein equation gives the equations of motion for the metric perturbations and the zeroth-order Friedmann equations:

- Let us start with the trace-free part of the ij equation, $G_{ij} = 8\pi G T_{ij}$. Since we have dropped anisotropic stress there is no source on the right-hand side. From eq. (4.2.134), we get

$$\boxed{\partial_{\langle i}\partial_{j\rangle}(\Phi-\Psi) = 0}\ . \tag{4.2.135}$$

Had we kept anisotropic stress, the right-hand side would be $-8\pi G a^2 \Pi_{ij}$. In the absence of anisotropic stress[4] (and assuming appropriate decay at infinity), we get[5]

$$\Phi = \Psi \ . \tag{4.2.136}$$

There is then only one gauge-invariant degree of freedom in the metric. In the following, we will write all equations in terms of Φ.

- Next, we consider the 00 equation, $G_{00} = 8\pi G T_{00}$. Using eq. (4.2.131), we get

$$\begin{aligned} 3\mathcal{H}^2 + 2\nabla^2\Phi - 6\mathcal{H}\Phi' &= 8\pi G\, g_{0\mu} T^{\mu}{}_0 \\ &= 8\pi G \left(g_{00} T^0{}_0 + g_{0i} T^i{}_0 \right) \\ &= 8\pi G a^2 (1+2\Phi)(\bar{\rho} + \delta\rho) \\ &= 8\pi G a^2 \bar{\rho} (1 + 2\Phi + \delta) \ . \end{aligned} \tag{4.2.137}$$

The zeroth-order part gives

$$\mathcal{H}^2 = \frac{8\pi G}{3} a^2 \bar{\rho} \ , \tag{4.2.138}$$

which is just the Friedmann equation. The first-order part of eq. (4.2.137) gives

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho}\delta + 8\pi G a^2 \bar{\rho}\Phi + 3\mathcal{H}\Phi' \ . \tag{4.2.139}$$

which, on using eq. (4.2.138), reduces to

$$\boxed{\nabla^2\Phi = 4\pi G a^2 \bar{\rho}\delta + 3\mathcal{H}(\Phi' + \mathcal{H}\Phi)} \ . \tag{4.2.140}$$

- Moving on to $0i$ equation, $G_{0i} = 8\pi G T_{0i}$, with

$$T_{0i} = g_{0\mu} T^{\mu}{}_i = g_{00} T^0{}_i = \bar{g}_{00} T^0{}_i = -a^2 q_i \ . \tag{4.2.141}$$

It follows that

$$\partial_i(\Phi' + \mathcal{H}\Phi) = -4\pi G a^2 q_i \ . \tag{4.2.142}$$

If we write $q_i = (\bar{\rho} + \bar{P})\partial_i v$ and assume the perturbations decay at infinity, we can integrate eq. (4.2.142) to get

$$\boxed{\Phi' + \mathcal{H}\Phi = -4\pi G a^2 (\bar{\rho} + \bar{P}) v} \ . \tag{4.2.143}$$

- Substituting eq. (4.2.143) into the 00 Einstein equation (4.2.140) gives

$$\boxed{\nabla^2\Phi = 4\pi G a^2 \bar{\rho}\Delta} \ , \qquad \text{where} \quad \bar{\rho}\Delta \equiv \bar{\rho}\delta - 3\mathcal{H}(\bar{\rho} + \bar{P}) v \ . \tag{4.2.144}$$

[4] In reality, neutrinos develop anisotropic stress after neutrino decoupling (i.e. they do not behave like a perfect fluid). Therefore, Φ and Ψ actually differ from each other by about 10% in the time between neutrino decoupling and matter-radiation equality. After the universe becomes matter-dominated, the neutrinos become unimportant, and Φ and Ψ rapidly approach each other. The same thing happens to photons after photon decoupling, but the universe is then already matter-dominated, so they do not cause a significant $\Phi - \Psi$ difference.

[5] In Fourier space, eq. (4.2.135) becomes

$$\left(k_i k_j - \tfrac{1}{3}\delta_{ij} k^2\right)(\Phi - \Psi) = 0 \ .$$

For finite k, we therefore must have $\Phi = \Psi$. For $k = 0$, $\Phi - \Psi = const.$ would be a solution. However, the constant must be zero, since the mean of the perturbations vanishes.

This is of the form of a *Poisson equation*, but with source density given by the gauge-invariant variable Δ of eq. (4.2.87) since $B = 0$ in the Newtonian gauge. Let us introduce *comoving hypersurfaces* as those that are orthogonal to the worldlines of a set of observers comoving with the total matter (i.e. they see $q^i = 0$) and are the constant-time hypersurfaces in the *comoving gauge* for which $q^i = 0$ and $B_i = 0$. It follows that Δ is the fractional overdensity in the comoving gauge and we see from eq. (4.2.144) that this is the source term for the gravitational potential Φ.

- Finally, we consider the trace-part of the ij equation, i.e. $G^i{}_i = 8\pi G T^i{}_i$. We compute the left-hand side from eq. (4.2.134) (with $\Phi = \Psi$),

$$\begin{aligned} G^i{}_i &= g^{i\mu} G_{\mu_i} \\ &= g^{ik} G_{ki} \\ &= -a^{-2}(1+2\Phi)\delta^{ik}\left[-(2\mathcal{H}' + \mathcal{H}^2)\delta_{ki} + \left(2\Phi'' + 6\mathcal{H}\Phi' + 4(2\mathcal{H}' + \mathcal{H}^2)\Phi\right)\delta_{ki}\right] \\ &= -3a^{-2}\left[-(2\mathcal{H}' + \mathcal{H}^2) + 2\left(\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi\right)\right] \ . \end{aligned} \tag{4.2.145}$$

We combine this with $T^i{}_i = -3(\bar{P} + \delta P)$. At zeroth order, we find

$$2\mathcal{H}' + \mathcal{H}^2 = -8\pi G a^2 \bar{P} \ , \tag{4.2.146}$$

which is just the second Friedmann equation. At first order, we get

$$\boxed{\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = 4\pi G a^2 \delta P} \ . \tag{4.2.147}$$

Of course, the Einstein equations and the energy and momentum conservation equations form a redundant (but consistent!) set of equations because of the Bianchi identity. We can use whichever subsets are most convenient for the particular problem at hand.

4.3 Conserved Curvature Perturbation

There is an important quantity that is *conserved* on super-Hubble scales for adiabatic fluctuations irrespective of the equation of state of the matter: the *comoving curvature perturbation*. As we will see below, the comoving curvature perturbation provides the essential link between the fluctuations that we observe in the late-time universe (Chapter 5) and the primordial seed fluctuations created by inflation (Chapter 6).

4.3.1 Comoving Curvature Perturbation

In some arbitrary gauge, let us work out the *intrinsic curvature* of surfaces of constant time. The *induced metric*, γ_{ij}, on these surfaces is just the spatial part of eq. (4.2.41), i.e.

$$\gamma_{ij} \equiv a^2\left[(1+2C)\delta_{ij} + 2E_{ij}\right] \ . \tag{4.3.148}$$

where $E_{ij} \equiv \partial_{\langle i}\partial_{j\rangle} E$ for scalar perturbations. In a tedious, but straightforward computation, we derive the three-dimensional Ricci scalar associated with γ_{ij},

$$a^2 R_{(3)} = -4\nabla^2\left(C - \frac{1}{3}\nabla^2 E\right) \ . \tag{4.3.149}$$

In the following insert I show all the steps.

Derivation.—The connection corresponding to γ_{ij} is

$$ {}^{(3)}\Gamma^i_{jk} = \frac{1}{2}\gamma^{il}\left(\partial_j\gamma_{kl} + \partial_k\gamma_{jl} - \partial_l\gamma_{jk}\right) \ , \tag{4.3.150} $$

where γ^{ij} is the inverse of the induced metric,

$$ \gamma^{ij} = a^{-2}\left[(1-2C)\delta^{ij} - 2E^{ij}\right] = a^{-2}\delta^{ij} + \mathcal{O}(1) \ . \tag{4.3.151} $$

In order to compute the connection to first order, we actually only need the inverse metric to zeroth order, since the spatial derivatives of the γ_{ij} are all first order in the perturbations. We have

$$ \begin{aligned} {}^{(3)}\Gamma^i_{jk} &= \delta^{il}\partial_j\left(C\delta_{kl} + E_{kl}\right) + \delta^{il}\partial_k\left(C\delta_{jl} + E_{jl}\right) - \delta^{il}\partial_l\left(C\delta_{jk} + E_{jk}\right) \\ &= 2\delta^i_{(j}\partial_{k)}C - \delta^{il}\delta_{jk}\partial_l C + 2\partial_{(j}E_{k)}{}^i - \delta^{il}\partial_l E_{jk} \ . \end{aligned} \tag{4.3.152} $$

The intrinsic curvature is the associated Ricci scalar, given by

$$ R_{(3)} = \gamma^{ik}\partial_l{}^{(3)}\Gamma^l_{ik} - \gamma^{ik}\partial_k{}^{(3)}\Gamma^l_{il} + \gamma^{ik\,(3)}\Gamma^l_{ik}{}^{(3)}\Gamma^m_{lm} - \gamma^{ik\,(3)}\Gamma^{m\,(3)}_{il}\Gamma^l_{km} \ . \tag{4.3.153} $$

To first order, this reduces to

$$ a^2\,R_{(3)} = \delta^{ik}\partial_l{}^{(3)}\Gamma^l_{ik} - \delta^{ik}\partial_k{}^{(3)}\Gamma^l_{il} \ . \tag{4.3.154} $$

This involves two contractions of the connection. The first is

$$ \begin{aligned} \delta^{ik\,(3)}\Gamma^l_{ik} &= \delta^{ik}\Big(2\delta^l_{(i}\partial_{k)}C - \delta^{jl}\delta_{ik}\partial_j C\Big) + \delta^{ik}\Big(2\partial_{(i}E_{k)}{}^l - \delta^{jl}\partial_j E_{ik}\Big) \\ &= 2\delta^{kl}\partial_k C - 3\delta^{jl}\partial_j C + 2\partial_i E^{il} - \delta^{jl}\partial_j\underbrace{\left(\delta^{ik}E_{ik}\right)}_{0} \\ &= -\delta^{kl}\partial_k C + 2\partial_k E^{kl} \ . \end{aligned} \tag{4.3.155} $$

The second is

$$ \begin{aligned} {}^{(3)}\Gamma^l_{il} &= \delta^l_l\partial_i C + \delta^l_i\partial_l C - \partial_i C + \partial_l E_i{}^l + \partial_i E_l{}^l - \partial_l E_i{}^l \\ &= 3\,\partial_i C \ . \end{aligned} \tag{4.3.156} $$

Eq. (4.3.154) therefore becomes

$$ \begin{aligned} a^2\,R_{(3)} &= \partial_l\left(-\delta^{kl}\partial_k C + 2\partial_k E^{kl}\right) - 3\delta^{ik}\partial_k\partial_i C \\ &= -\nabla^2 C + 2\partial_i\partial_j E^{ij} - 3\nabla^2 C \\ &= -4\nabla^2 C + 2\partial_i\partial_j E^{ij} \ . \end{aligned} \tag{4.3.157} $$

Note that this vanishes for vector and tensor perturbations (as do all perturbed scalars) since then $C = 0$ and $\partial_i\partial_j E^{ij} = 0$. For scalar perturbations, $E_{ij} = \partial_{\langle i}\partial_{j\rangle}E$ so

$$ \begin{aligned} \partial_i\partial_j E^{ij} &= \delta^{il}\delta^{jm}\partial_i\partial_j\left(\partial_l\partial_m E - \frac{1}{3}\delta_{lm}\nabla^2 E\right) \\ &= \nabla^2\nabla^2 E - \frac{1}{3}\nabla^2\nabla^2 E \\ &= \frac{2}{3}\nabla^4 E \ . \end{aligned} \tag{4.3.158} $$

Finally, we get eq. (4.3.149).

We define the *curvature perturbation* as $C - \frac{1}{3}\nabla^2 E$. The *comoving curvature perturbation* $\mathcal{R}$

is the curvature perturbation evaluated in the comoving gauge ($B_i = 0 = q^i$). It will prove convenient to have a gauge-invariant expression for $\mathcal{R}$, so that we can evaluate it from the perturbations in any gauge (for example, in Newtonian gauge). Since B and v vanish in the comoving gauge, we can always add linear combinations of these to $C - \frac{1}{3}\nabla^2 E$ to form a gauge-invariant combination that equals $\mathcal{R}$. Using eqs. (4.2.58)–(4.2.60) and (4.2.85), we see that the correct gauge-invariant expression for the comoving curvature perturbation is

$$\boxed{\mathcal{R} = C - \frac{1}{3}\nabla^2 E + \mathcal{H}(B + v)} \,. \tag{4.3.159}$$

Exercise.—Show that $\mathcal{R}$ is gauge-invariant.

4.3.2 A Conservation Law

We now want to prove that the comoving curvature perturbation $\mathcal{R}$ is indeed conserved on large scales and for adiabatic perturbations. We shall do so by working in the *Newtonian gauge*, in which case

$$\mathcal{R} = -\Phi + \mathcal{H}v \,, \tag{4.3.160}$$

since $B = E = 0$ and $C \equiv -\Phi$. We can use the $0i$ Einstein equation (4.2.143) to eliminate the peculiar velocity in favour of the gravitational potential and its time derivative:

$$\mathcal{R} = -\Phi - \frac{\mathcal{H}(\Phi' + \mathcal{H}\Phi)}{4\pi G a^2(\bar{\rho} + \bar{P})} \,. \tag{4.3.161}$$

Taking a time derivative of (4.3.161) and using the evolution equations of the previous section, we find

$$\boxed{-4\pi G a^2(\bar{\rho} + \bar{P})\,\mathcal{R}' \;=\; 4\pi G a^2 \mathcal{H}\,\delta P_{\rm nad} + \mathcal{H}\frac{\bar{P}'}{\bar{\rho}'}\nabla^2\Phi} \,, \tag{4.3.162}$$

where we have defined the *non-adiabatic pressure perturbation*

$$\delta P_{\rm nad} \equiv \delta P - \frac{\bar{P}'}{\bar{\rho}'}\,\delta\rho \,. \tag{4.3.163}$$

*Derivation.**—We differentiate eq. (4.3.161) to find

$$\begin{aligned} -4\pi G a^2(\bar{\rho} + \bar{P})\,\mathcal{R}' \;=\; & 4\pi G a^2(\bar{\rho} + \bar{P})\Phi' + \mathcal{H}'(\Phi' + \mathcal{H}\Phi) + \mathcal{H}(\Phi'' + \mathcal{H}'\Phi + \mathcal{H}\Phi') \\ & + \mathcal{H}^2(\Phi' + \mathcal{H}\Phi) + 3\mathcal{H}^2\frac{\bar{P}'}{\bar{\rho}'}(\Phi' + \mathcal{H}\Phi) \,, \end{aligned} \tag{4.3.164}$$

where we used $\bar{\rho}' = -3\mathcal{H}(\bar{\rho} + \bar{P})$. This needs to be cleaned up a bit. In the first term on the right, we use the Friedmann equation to write $4\pi G a^2(\bar{\rho} + \bar{P})$ as $\mathcal{H}^2 - \mathcal{H}'$. In the last term, we use the Poisson equation (4.2.140) to write $3\mathcal{H}(\Phi' + \mathcal{H}\Phi)$ as $(\nabla^2\Phi - 4\pi G a^2\bar{\rho}\delta)$. We then find

$$\begin{aligned} -4\pi G a^2(\bar{\rho} + \bar{P})\,\mathcal{R}' \;=\; & (\mathcal{H}^2 - \mathcal{H}')\Phi' + \mathcal{H}'(\Phi' + \mathcal{H}\Phi) + \mathcal{H}(\Phi'' + \mathcal{H}'\Phi + \mathcal{H}\Phi') \\ & + \mathcal{H}^2(\Phi' + \mathcal{H}\Phi) + \mathcal{H}\frac{\bar{P}'}{\bar{\rho}'}\left(\nabla^2\Phi - 4\pi G a^2\bar{\rho}\delta\right) \,. \end{aligned} \tag{4.3.165}$$

Adding and subtracting $4\pi G a^2 \mathcal{H}\delta P$ on the right-hand side and simplifying gives

$$-4\pi G a^2(\bar{\rho}+\bar{P})\,\mathcal{R}' = \mathcal{H}\left[\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi - 4\pi G a^2 \delta P\right] + 4\pi G a^2 \mathcal{H}\delta P_{\rm nad} + \mathcal{H}\frac{\bar{P}'}{\bar{\rho}'}\nabla^2\Phi \ , \tag{4.3.166}$$

where $\delta P_{\rm nad}$ was defined in (4.3.163). The first term on the right-hand side vanishes by eq. (4.2.147), so we obtain eq. (4.3.162).

Exercise.—Show that $\delta P_{\rm nad}$ is gauge-invariant.

The non-adiabatic pressure $\delta P_{\rm nad}$ vanishes for a barotropic equation of state, $P = P(\rho)$ (and, more generally, for adiabatic fluctuations in a mixture of barotropic fluids). In that case, the right-hand side of eq. (4.3.162) scales as $\mathcal{H}k^2\Phi \sim \mathcal{H}k^2\mathcal{R}$, so that

$$\frac{d\ln\mathcal{R}}{d\ln a} \sim \left(\frac{k}{\mathcal{H}}\right)^2 \ . \tag{4.3.167}$$

Hence, we find that $\mathcal{R}$ doesn't evolve on super-Hubble scales, $k \ll \mathcal{H}$. This means that the value of $\mathcal{R}$ that we will compute at horizon crossing during inflation (Chapter 6) survives unaltered until later times.

4.4 Summary

We have derived the linearised evolution equations for scalar perturbations in Newtonian gauge, where the metric has the following form

$$\mathrm{d}s^2 = a^2(\tau)\left[(1+2\Psi)\mathrm{d}\tau^2 - (1-2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right] \ . \tag{4.4.168}$$

In these lectures, we won't encounter situations where anisotropic stress plays a significant role, so we will always be able to set $\Psi = \Phi$.

- The Einstein equations then are

$$\nabla^2\Phi - 3\mathcal{H}(\Phi' + \mathcal{H}\Phi) = 4\pi G a^2\,\delta\rho \ , \tag{4.4.169}$$

$$\Phi' + \mathcal{H}\Phi = -4\pi G a^2(\bar{\rho}+\bar{P})v \ , \tag{4.4.170}$$

$$\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = 4\pi G a^2\,\delta P \ . \tag{4.4.171}$$

  The source terms on the right-hand side should be interpreted as the sum over all relevant matter components (e.g. photons, dark matter, baryons, etc.). The Poisson equation takes a particularly simple form if we introduce the comoving gauge density contrast

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho}\,\Delta \ . \tag{4.4.172}$$

- From the conservation of the stress-tensor, we derived the relativistic generalisations of the continuity equation and the Euler equation

$$\delta' + 3\mathcal{H}\left(\frac{\delta P}{\delta\rho} - \frac{\bar{P}}{\bar{\rho}}\right)\delta = -\left(1+\frac{\bar{P}}{\bar{\rho}}\right)\left(\boldsymbol{\nabla}\cdot\boldsymbol{v} - 3\Phi'\right) \ , \tag{4.4.173}$$

$$\boldsymbol{v}' + 3\mathcal{H}\left(\frac{1}{3} - \frac{\bar{P}'}{\bar{\rho}'}\right)\boldsymbol{v} = -\frac{\boldsymbol{\nabla}\delta P}{\bar{\rho}+\bar{P}} - \boldsymbol{\nabla}\Phi \ . \tag{4.4.174}$$

These equations apply for the total matter and velocity, and also separately for any non-interacting components so that the individual stress-energy tensors are separately conserved.

- A very important quantity is the comoving curvature perturbation

$$\mathcal{R} = -\Phi - \frac{\mathcal{H}(\Phi' + \mathcal{H}\Phi)}{4\pi G a^2(\bar{\rho} + \bar{P})} \,. \tag{4.4.175}$$

We have shown that $\mathcal{R}$ doesn't evolve on super-Hubble scales, $k \ll \mathcal{H}$, unless non-adiabatic pressure is significant. This fact is crucial for relating late-time observables, such as the distributions of galaxies (Chapter 5), to the initial conditions from inflation (Chapter 6).

5 Structure Formation

In the previous chapter, we derived the evolution equations for all matter and metric perturbations. In principle, we could now solve these equations. The complex interactions between the different species (see fig. 5.1) means that we get a large number of coupled differential equations. This set of equations is easy to solve numerically and this is what is usually done. However, our goal in this chapter is to obtain some analytical insights into the basic qualitative features of the solutions.

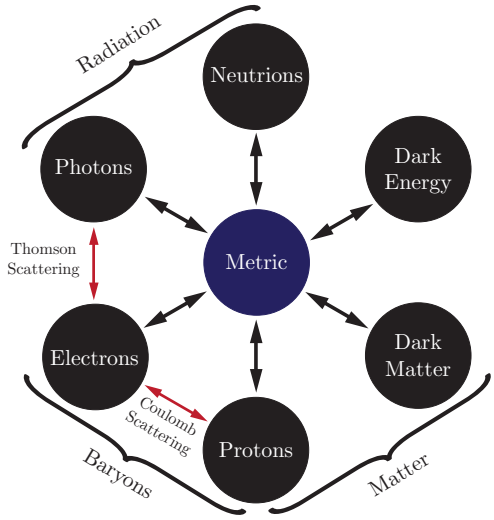


Figure 5.1: *Interactions between the different forms of matter in the universe.*

5.1 Initial Conditions

Any mode of interest for observations today was outside the Hubble radius if we go back sufficiently far into the past. Inflation sets the initial condition for these superhorizon modes. The prediction from inflation (see Ch. 6) is presented most conveniently in terms of a spectrum of fluctuations for the curvature perturbation $\mathcal{R}$. Eq. (4.4.175) relates this to the gravitational potential Φ in Newtonian gauge

$$\mathcal{R} = -\Phi - \frac{2}{3(1+w)}\left(\frac{\Phi'}{\mathcal{H}} + \Phi\right) \ , \tag{5.1.1}$$

where w is the equation of state of the background. For adiabatic perturbations, we have $c_s^2 \approx w$ and a combination of Einstein equations imply a closed form evolution equation for the gravitational potential

$$\boxed{\Phi'' + 3(1+w)\mathcal{H}\Phi' + wk^2\Phi = 0} \ . \tag{5.1.2}$$

Notice that in deriving (5.1.2) we have assumed a constant equation of state. It therefore only applies if a single component dominates the universe. For the more general case, you should consult (4.4.171).

Exercise.—Derive eq. (5.1.2) from the Einstein equations.

5.1.1 Superhorizon Limit

On superhorizon scales, $k \ll \mathcal{H}$, we can drop the last term in (5.1.2). The growing-mode solution then is

$$\Phi = const. \qquad \text{(superhorizon)} \ . \tag{5.1.3}$$

Notice that this superhorizon solution is independent of the equation of state w (as long as $w = const.$). In particular, the gravitational potential is frozen outside the horizon during both the radiation and matter eras.

The Poisson equation (4.4.169) relates the gravitational potential to the *total* Newtonian-gauge density contrast

$$\delta = -\frac{2}{3}\frac{k^2}{\mathcal{H}^2}\,\Phi - \frac{2}{\mathcal{H}}\,\Phi' - 2\Phi \ , \tag{5.1.4}$$

where we have used $\frac{3}{2}\mathcal{H}^2 = 4\pi G a^2 \bar{\rho}$. On superhorizon scales, only the decaying mode contributes to Φ'. The first and second term in (5.1.4) then are of the same order and both are much smaller than the third term. We therefore get

$$\delta \approx -2\Phi = const. \ , \tag{5.1.5}$$

so δ is also frozen on superhorizon scales. For adiabatic initial conditions, we can relate the primordial potential Φ to the fluctuations in both the matter and the radiation:

$$\delta_m = \frac{3}{4}\delta_r \approx -\frac{3}{2}\Phi_{\rm RD} \ , \tag{5.1.6}$$

where we have used that $\delta_r \approx \delta$ for adiabatic perturbations during the radiation era. On superhorizon scales, the density perturbations are therefore simply proportional to the curvature perturbation set up by inflation.

5.1.2 Radiation-to-Matter Transition

We have seen that the gravitational potential is frozen on superhorizon scales as long as the equation of state of the background is constant. However, unlike the curvature perturbation $\mathcal{R}$, the gravitational doesn't stay constant when the equation of state changes. To follow the evolution of Φ through the radiation-to-matter transition, we exploit the conservation of $\mathcal{R}$.

In the superhorizon limit, the comoving curvature perturbation (4.4.175) becomes

$$\boxed{\mathcal{R} = -\frac{5+3w}{3+3w}\,\Phi} \qquad \text{(superhorizon)} \ . \tag{5.1.7}$$

This provides an important link between the source term for the evolution of fluctuations (Φ) and the primordial initial conditions set up by inflation ($\mathcal{R}$). Evaluating (5.1.7) for $w = \frac{1}{3}$ and

$w = 0$ relates the amplitudes of Φ during the radiation era and the matter eta

$$\mathcal{R} = -\frac{3}{2}\Phi_{\rm RD} = -\frac{5}{3}\Phi_{\rm MD} \quad \Rightarrow \quad \Phi_{\rm MD} = \frac{9}{10}\Phi_{\rm RD} \ , \tag{5.1.8}$$

where we have used that $\mathcal{R} = const.$ throughout. We see that the gravitational potential decreases by a factor of 9/10 in the transition from radiation-dominated to matter-dominated.

5.2 Evolution of Fluctuations

We wish to understand what happens to the superhorizon initial conditions, when modes enter the horizon. We will first study the evolution of the gravitational potential (§5.2.1), and then the perturbations in radiation (§5.2.2), matter (§5.2.3) and baryons (§5.2.4).

5.2.1 Gravitational Potential

To determine the evolution of Φ during both the radiation era and the matter era, we simply have to specialise (5.1.2) to $w = \frac{1}{3}$ and $w = 0$, respectively.

Radiation Era

In the radiation era, $w = \frac{1}{3}$, we get

$$\Phi'' + \frac{4}{\tau}\Phi' + \frac{k^2}{3}\Phi = 0 \ . \tag{5.2.9}$$

This equation has the following exact solution

$$\Phi_{\boldsymbol{k}}(\tau) = A_{\boldsymbol{k}}\,\frac{j_1(x)}{x} + B_{\boldsymbol{k}}\,\frac{n_1(x)}{x} \ , \qquad x \equiv \frac{1}{\sqrt{3}}k\tau \ , \tag{5.2.10}$$

where the subscript $\boldsymbol{k}$ indicates that the solution can have different amplitudes for each value of $\boldsymbol{k}$. The size of the initial fluctuations as a function of wavenumber will be a prediction of inflation. The functions $j_1(x)$ and $n_1(x)$ in (5.2.10) are the spherical Bessel and Neumann functions

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} = \frac{x}{3} + \mathcal{O}(x^3) \ , \tag{5.2.11}$$

$$n_1(x) = -\frac{\cos x}{x^2} - \frac{\sin x}{x} = -\frac{1}{x^2} + \mathcal{O}(x^0) \ . \tag{5.2.12}$$

Since $n_1(x)$ blows up for small x (early times), we reject that solution on the basis of initial conditions, i.e. we set $B_{\boldsymbol{k}} \equiv 0$. We match the constant $A_{\boldsymbol{k}}$ to the primordial value of the potential, $\Phi_{\boldsymbol{k}}(0) = -\frac{2}{3}\mathcal{R}_{\boldsymbol{k}}(0)$. Using (5.2.11), we find

$$\Phi_{\boldsymbol{k}}(\tau) = -2\mathcal{R}_{\boldsymbol{k}}(0)\left(\frac{\sin x - x\cos x}{x^3}\right) \qquad \text{(all scales)} \ . \tag{5.2.13}$$

Notice that (5.2.13) is valid on all scales. Outside the (sound) horizon, $x = \frac{1}{\sqrt{3}}k\tau \ll 1$, the solution approaches $\Phi = const.$, while on subhorizon scales, $x \gg 1$, we get

$$\Phi_{\boldsymbol{k}}(\tau) \approx -6\mathcal{R}_{\boldsymbol{k}}(0)\,\frac{\cos\left(\frac{1}{\sqrt{3}}k\tau\right)}{(k\tau)^2} \qquad \text{(subhorizon)} \ . \tag{5.2.14}$$

During the radiation era, subhorizon modes of Φ therefore oscillate with frequency $\frac{1}{\sqrt{3}}k$ and an amplitude that decays as $\tau^{-2} \propto a^{-2}$ (see fig. 5.2). Remember this.

Matter Era

In the matter era, $w = 0$, the evolution of the potential is

$$\Phi'' + \frac{6}{\tau}\Phi' = 0 \ , \tag{5.2.15}$$

whose solution is

$$\Phi \propto \begin{cases} const. \\ \tau^{-5} \propto a^{-5/2} \end{cases} \ . \tag{5.2.16}$$

We conclude that the gravitational potential is frozen on *all scales* during matter domination.

Summary

Fig. 5.2 shows the evolution of the gravitational potential for difference wavelengths. As predicted, the potential is constant when the modes are outside the horizon. Two of the modes enter the horizon during the radiation era. While they are inside the horizon during the radiation era their amplitudes decrease as a^{-2}. The resulting amplitudes in the matter era are therefore strongly suppressed. During the matter era the potential is constant on all scales. The longest wavelength mode in the figure enters the horizon during the matter era, so its amplitude is only suppressed by the factor of $\frac{9}{10}$ coming from the radiation-to-matter transition.

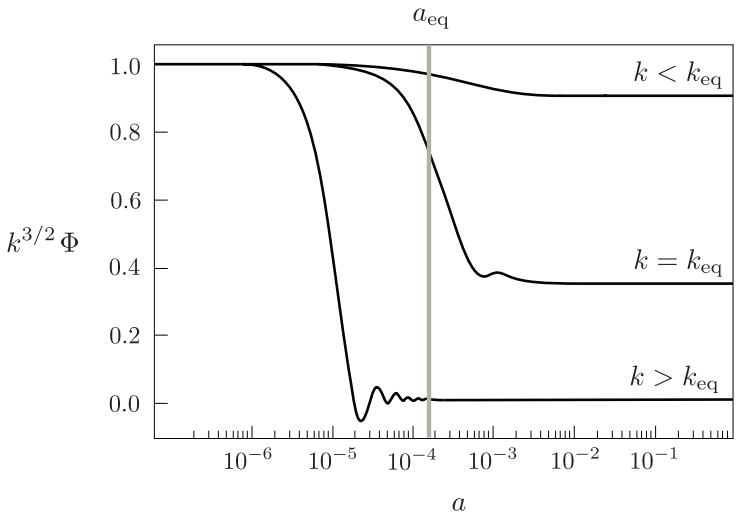


Figure 5.2: *Numerical solutions for the linear evolution of the gravitational potential.*

5.2.2 Radiation

In this section, we wish to determine the evolution of perturbations in the radiation density.

Radiation Era

In the radiation era, perturbations in the radiation density dominate (for adiabatic initial conditions). Given the solution (5.2.13) for Φ during the radiation era, we therefore immediately

obtain a solution for the density contrast of radiation (δ_r or Δ_r) via the Poisson equation

$$\delta_r = -\frac{2}{3}(k\tau)^2\Phi - 2\tau\Phi' - 2\Phi \ , \tag{5.2.17}$$

$$\Delta_r = -\frac{2}{3}(k\tau)^2\Phi \ . \tag{5.2.18}$$

We see that while δ_r is constant outside the horizon, Δ_r grows as $\tau^2 \propto a^2$. Inside the horizon,[1]

$$\delta_r \approx \Delta_r = -\frac{2}{3}(k\tau)^2\Phi = 4\mathcal{R}(0)\cos\left(\frac{1}{\sqrt{3}}k\tau\right) \ , \tag{5.2.19}$$

which is the solution to

$$\boxed{\delta_r'' - \frac{1}{3}\nabla^2\delta_r = 0} \ . \tag{5.2.20}$$

We see that subhorizon fluctuations in the radiation density oscillate with constant amplitude around $\delta_r = 0$.

Matter Era

In the matter era, radiation perturbations are subdominant. Their evolution then has to be determined from the conservation equations. On subhorizon scales, we have

$$\left.\begin{aligned} &\text{(C)} \quad \delta_r' = -\frac{4}{3}\boldsymbol{\nabla}\cdot\boldsymbol{v}_r \\ &\text{(E)} \quad \boldsymbol{v}_r' = -\frac{1}{4}\boldsymbol{\nabla}\delta_r - \boldsymbol{\nabla}\Phi \end{aligned}\right\} \quad \boxed{\delta_r'' - \frac{1}{3}\nabla^2\delta_r = \frac{4}{3}\nabla^2\Phi} = const. \tag{5.2.21}$$

This is the equation of motion of a harmonic oscillator with constant driving force. During the matter era, the subhorizon fluctuations in the radiation density therefore oscillate with constant amplitude around a shifted equilibrium point, $\delta_r = -4\Phi_{\rm MD}(k)$. Here, $\Phi_{\rm MD}(k)$ is the k-dependent amplitude of the gravitational potential in the matter era; cf. fig. 5.2.

Summary

The *acoustic oscillations* in the perturbed radiation density are what gives rise to the peaks in the spectrum of CMB anisotropies (see fig. 6.5 in §6.6.2) This will be analysed in much more detail in the *Advanced Cosmology* course next term.

5.2.3 Dark Matter

In this section, we are interested in the evolution of matter fluctuations from early times (during the radiation era) until late times (when dark energy starts to dominate).

Early Times

At early times, the universe was dominated by a mixture of radiation (r) and pressureless matter (m). For now, we ignore baryons (but see §5.2.4). The conformal Hubble parameter is

$$\mathcal{H}^2 = \frac{\mathcal{H}_0^2\Omega_m^2}{\Omega_r}\left(\frac{1}{y} + \frac{1}{y^2}\right) \ , \qquad y \equiv \frac{a}{a_{\rm eq}} \ . \tag{5.2.22}$$

[1] We see that well inside the horizon, the density perturbations in the comoving and Newtonian gauge coincide. This is indicative of the general result that there are no gauge ambiguities inside the horizon.

We wish to determine how matter fluctuations evolve on subhorizon scales from the radiation era until the matter era. We consider the evolution equations for the matter density contrast and velocity:

$$\left.\begin{array}{ll} \text{(C)} & \delta_m' = -\boldsymbol{\nabla}\cdot\boldsymbol{v}_m \\ \\ \text{(E)} & \boldsymbol{v}_m' = -\mathcal{H}\boldsymbol{v}_m - \boldsymbol{\nabla}\Phi \end{array}\right\} \quad \delta_m'' + \mathcal{H}\delta_m' = \nabla^2\Phi \ . \tag{5.2.23}$$

In general, the potential Φ is sourced by the total density fluctuation. However, we have seen that perturbations in the radiation density oscillate rapidly on small scales. The *time-averaged* gravitational potential is therefore only sourced by the matter fluctuations, and the fluctuations in the radiation can be neglected (see Weinberg, astro-ph/0207375 for further discussion). The evolution of the matter perturbations then satisfies

$$\delta_m'' + \mathcal{H}\delta_m' - 4\pi G a^2 \bar{\rho}_m \delta_m \approx 0 \ , \tag{5.2.24}$$

where $\mathcal{H}$ given by (5.2.22). On Problem Set 3, you will show that this equation can be written as the *Mészáros equation*

$$\boxed{\frac{d^2\delta_m}{dy^2} + \frac{2+3y}{2y(1+y)}\frac{d\delta_m}{dy} - \frac{3}{2y(1+y)}\delta_m = 0} \ . \tag{5.2.25}$$

You will also be asked to show that the solutions to this equation take the form

$$\delta_m \propto \begin{cases} 2+3y \\ \\ (2+3y)\ln\left(\dfrac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right) - 6\sqrt{1+y} \end{cases} \ .$$

In the limit $y \ll 1$ (RD), the growing mode solution is $\delta_m \propto \ln y \propto \ln a$, confirming the logarithmic growth of matter fluctuations in the radiation era. In the limit $y \gg 1$ (MD), we reproduce the expected solution in the matter era: $\delta_m \propto y \propto a$. Table 5.1 summarises the analytical limits for the evolution of the potential Φ and the matter density contrasts δ_m and Δ_m.

		RD		MD	
		Φ	δ_m (Δ_m)	Φ	δ_m (Δ_m)
$k \gg k_{\rm eq}$:	superhorizon	*const.*	*const.* (a^2)	–	–
	subhorizon	a^{-2}	$\ln a$	*const.*	a
$k \ll k_{\rm eq}$:	superhorizon	*const.*	*const.* (a^2)	*const.*	*const.* (a)
	subhorizon	–	–	*const.*	a

Table 5.1: *Analytical limits of the solutions for the potential Φ and the matter density contrasts δ_m and Δ_m.*

Intermediate Times

The solution in the matter era also follows directly from the solution (5.2.16) for the gravitational potential, which determines the comoving density contrast

$$\Delta_m = \frac{\nabla^2\Phi}{4\pi G a^2 \bar{\rho}} \propto \begin{cases} a \\ a^{-3/2} \end{cases} , \tag{5.2.26}$$

just as in the Newtonian treatment [cf. eq. (4.1.30)], but now valid on all scales. Notice that the growing mode of Δ_m grows as a outside the horizon, while δ_m is constant. Inside the horizon, $\delta_m \approx \Delta_m$ and the density contrasts in both gauges evolve as a.

Late Times

At late times, the universe is a mixture of pressureless matter (m) and dark energy (Λ). Since dark energy doesn't have fluctuations, we still have

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho}_m \Delta_m \ . \tag{5.2.27}$$

Pressure fluctuations are negligible, so the Einstein equations give

$$\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = 0 \ . \tag{5.2.28}$$

To get an evolution equation for Δ_m, we use a neat trick. Since $a^2\bar{\rho}_m \propto a^{-1}$, we have $\Phi \propto \Delta_m/a$. Hence, eq. (5.2.28) implies

$$\partial_\tau^2(\Delta_m/a) + 3\mathcal{H}\partial_\tau(\Delta_m/a) + (2\mathcal{H}' + \mathcal{H}^2)(\Delta_m/a) = 0 \ , \tag{5.2.29}$$

which rearranges to

$$\Delta_m'' + \mathcal{H}\Delta_m' + (\mathcal{H}' - \mathcal{H}^2)\Delta_m = 0 \ . \tag{5.2.30}$$

Exercise.—Show that (5.2.30) follows from (5.2.29). Use the Friedmann and conservation equations to show that

$$\mathcal{H}' - \mathcal{H}^2 = -4\pi G a^2(\bar{\rho} + \bar{P}) = -4\pi G a^2 \bar{\rho}_m \ . \tag{5.2.31}$$

Using (5.2.31), eq. (5.2.30) becomes

$$\boxed{\Delta_m'' + \mathcal{H}\Delta_m' - 4\pi G a^2 \bar{\rho}_m \Delta_m = 0} \ . \tag{5.2.32}$$

This is the conformal-time version of the Newtonian equation (4.1.36), but now valid on all scales. So we recover the usual suppression of the growth of structure by Λ, but now on all scales (see also Problem Set 3).

Summary

Fig. 5.3 shows the evolution of the matter density contrast δ_m for the same modes as in fig. 5.2. Fluctuations are frozen until they enter the horizon. Subhorizon matter fluctuations in the radiation era only grow logarithmically, $\delta_m \propto \ln a$. This changes to power-law growth, $\delta_m \propto a$

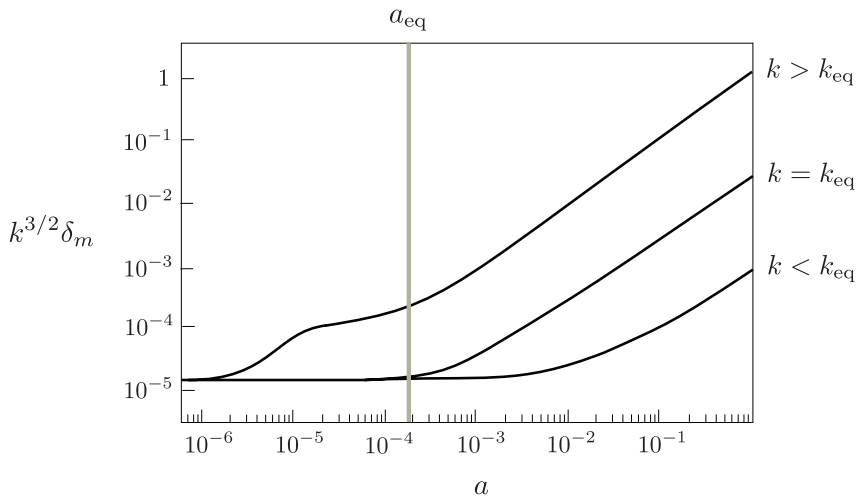


Figure 5.3: *Evolution of the matter density contrast for the same modes as in fig. 5.2.*

when the universe becomes matter dominated. When the universe becomes dominated by dark energy, perturbations stop growing.

The effects we discussed above lead to a post-processing of the primordial perturbations. This evolution is often encoded in the so-called *transfer function*. For example, the value of the matter perturbation at redshift z is related to the primordial perturbation $\mathcal{R}_{\boldsymbol{k}}$ by

$$\Delta_{m,\boldsymbol{k}}(z) = T(k,z)\,\mathcal{R}_{\boldsymbol{k}}\ . \tag{5.2.33}$$

The transfer function $T(k,z)$ depends only on the magnitude k and not on the direction of $\boldsymbol{k}$, because the perturbations are evolving on a homogeneous and isotropic background. The square of the Fourier mode (5.2.33) defines that matter *power spectrum*

$$P_\Delta(k,z) \equiv |\Delta_{m,\boldsymbol{k}}(z)|^2 = T^2(k,z)\,|\mathcal{R}_{\boldsymbol{k}}|^2\ . \tag{5.2.34}$$

Fig. 5.4 shows predicted matter power spectrum for scale-invariant initial conditions, $k^3|\mathcal{R}_{\boldsymbol{k}}|^2 = const.$ (see Chapter 6).

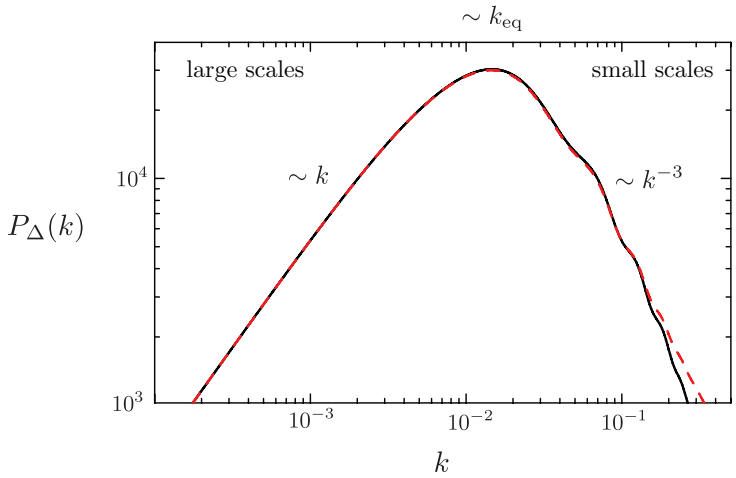


Figure 5.4: *The matter power spectrum $P_\Delta(k)$ at $z = 0$ in linear theory (solid) and with non-linear corrections (dashed). On large scales, $P_\Delta(k)$ grows as k. The power spectrum turns over around $k_{\rm eq} \sim 0.01\,{\rm Mpc}^{-1}$ corresponding to the horizon size at matter-radiation equality. Beyond the peak, the power falls as k^{-3}. Visible are small amplitude baryon acoustic oscillations in the spectrum.*

Exercise.—Explain the asymptotic scalings of the matter power spectrum

$$P_\Delta(k) = \begin{cases} k & k < k_{\rm eq} \\ k^{-3} & k > k_{\rm eq} \end{cases} . \tag{5.2.35}$$

5.2.4 Baryons*

Let us say a few (non-examinable!) words about the evolution of baryons.

Before Decoupling

At early times, $z > z_{dec} \approx 1100$, photons and baryons are coupled strongly to each other via Compton scattering. We can therefore treat the photons and baryons a single fluid, with $\boldsymbol{v}_\gamma = \boldsymbol{v}_b$ and $\delta_\gamma = \frac{4}{3}\delta_b$. The pressure of the photons supports oscillations on small scales (see fig. 5.5). Since the dark matter density contrast δ_c grows like a after matter-radiation equality, it follows that just after decoupling, $\delta_c \gg \delta_b$. Subsequently, the baryons fall into the potential wells sourced mainly by the dark matter and $\delta_b \to \delta_c$ as we shall now show.

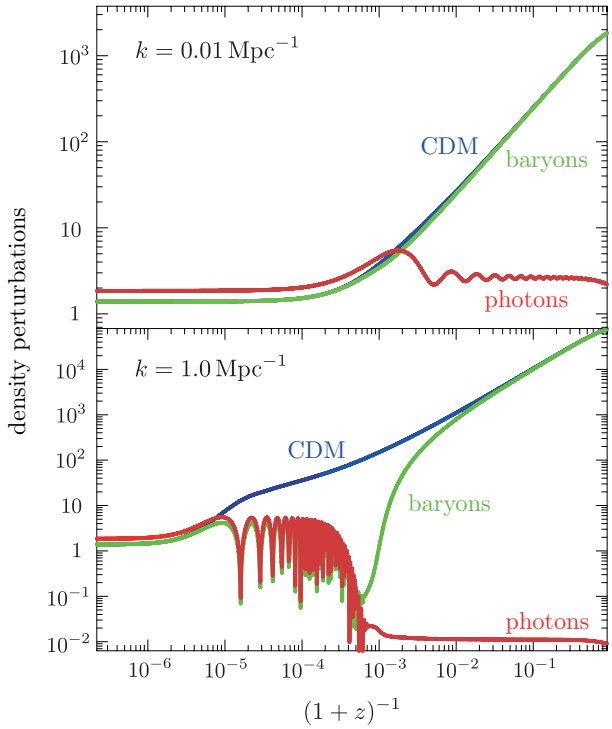


Figure 5.5: *Evolution of photons, baryons and dark matter.*

After Decoupling

After decoupling, the baryons lose the pressure support of the photons and gravitational instability kicks in. Ignoring baryon pressure, the coupled dynamics of the baryon fluid and the dark

matter fluid after decoupling is approximately given by

$$\delta_b'' + \mathcal{H}\delta_b' = 4\pi G a^2(\bar{\rho}_b\delta_b + \bar{\rho}_c\delta_c) \ , \tag{5.2.36}$$

$$\delta_c'' + \mathcal{H}\delta_c' = 4\pi G a^2(\bar{\rho}_b\delta_b + \bar{\rho}_c\delta_c) \ . \tag{5.2.37}$$

The two equations are coupled via the gravitational potential which is sourced by the total density contrast $\bar{\rho}_m\delta_m = \bar{\rho}_b\delta_b + \bar{\rho}_c\delta_c$. We can decouple these equations by defining $D \equiv \delta_b - \delta_c$. Subtracting eqs. (5.2.36) and (5.2.37), we find

$$D'' + \frac{2}{\tau}D' = 0 \quad \Rightarrow \quad D \propto \begin{cases} const. \\ \tau^{-1} \end{cases} \ , \tag{5.2.38}$$

while the evolution of δ_m is governed

$$\delta_m'' + \frac{2}{\tau}\delta_m' - \frac{6}{\tau^2}\delta_m = 0 \quad \Rightarrow \quad \delta_m \propto \begin{cases} \tau^2 \\ \tau^{-3} \end{cases} \ . \tag{5.2.39}$$

Since

$$\frac{\delta_b}{\delta_c} = \frac{\bar{\rho}_m\delta_m + \bar{\rho}_c D}{\bar{\rho}_m\delta_m - \bar{\rho}_b D} \to \frac{\delta_m}{\delta_m} = 1 \ , \tag{5.2.40}$$

we see that δ_b approaches δ_c during matter domination (see fig. 5.2).

The non-zero initial value of δ_b at decoupling, and, more importantly δ_b', leaves a small imprint in the late-time δ_m that oscillates with scale. These *baryon acoustic oscillations* have recently been detected in the clustering of galaxies.

6 Initial Conditions from Inflation

Arguably, the most important consequence of inflation is the fact that it includes a natural mechanism to produce primordial seeds for all of the large-scale structures we see around us. The reason why inflation inevitably produces fluctuations is simple: as we have seen in Chapter 2, the evolution of the inflaton field $\phi(t)$ governs the energy density of the early universe $\rho(t)$ and hence controls the end of inflation. Essentially, the field ϕ plays the role of a local "clock" reading off the amount of inflationary expansion still to occur. By the uncertainty principle, arbitrarily precise timing is not possible in quantum mechanics. Instead, quantum-mechanical clocks necessarily have some variance, so the inflaton will have spatially varying fluctuations $\delta\phi(t, \boldsymbol{x}) \equiv \phi(t, \boldsymbol{x}) - \bar{\phi}(t)$. There will hence be local differences in the time when inflation ends, $\delta t(\boldsymbol{x})$, so that different regions of space inflate by different amounts. These differences in the

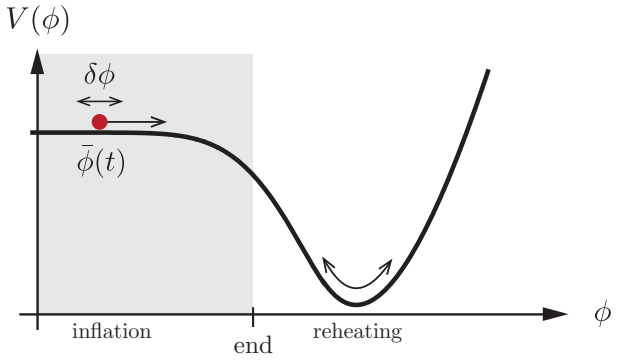


Figure 6.1: *Quantum fluctuations $\delta\phi(t, \boldsymbol{x})$ around the classical background evolution $\bar{\phi}(t)$. Regions acquiring a negative fluctuations $\delta\phi$ remain potential-dominated longer than regions with positive $\delta\phi$. Different parts of the universe therefore undergo slightly different evolutions. After inflation, this induces density fluctuations $\delta\rho(t, \boldsymbol{x})$.*

local expansion histories lead to differences in the local densities after inflation, $\delta\rho(t, \boldsymbol{x})$, and ultimately in the CMB temperature, $\delta T(\boldsymbol{x})$. The main purpose of this chapter is to compute this effect. It is worth remarking that the theory wasn't engineered to produce the CMB fluctuations, but their origin is instead a natural consequence of treating inflation quantum mechanically.

6.1 From Quantum to Classical

Before we get into the details, let me describe the big picture. At early times, all modes of interest were inside the horizon during inflation (see fig. 6.2). On small scales fluctuations in the inflaton field are described by a collection of harmonic oscillators. Quantum fluctuations induce a non-zero variance in the amplitudes of these oscillators

$$\langle |\delta\phi_k|^2 \rangle \equiv \langle 0 | |\delta\phi_k|^2 | 0 \rangle \ . \tag{6.1.1}$$

The inflationary expansion stretches these fluctuations to superhorizon scales. (In comoving coordinates, the fluctuations have constant wavelengths, but the Hubble radius shrinks, creating super-Hubble fluctuations in the process.)

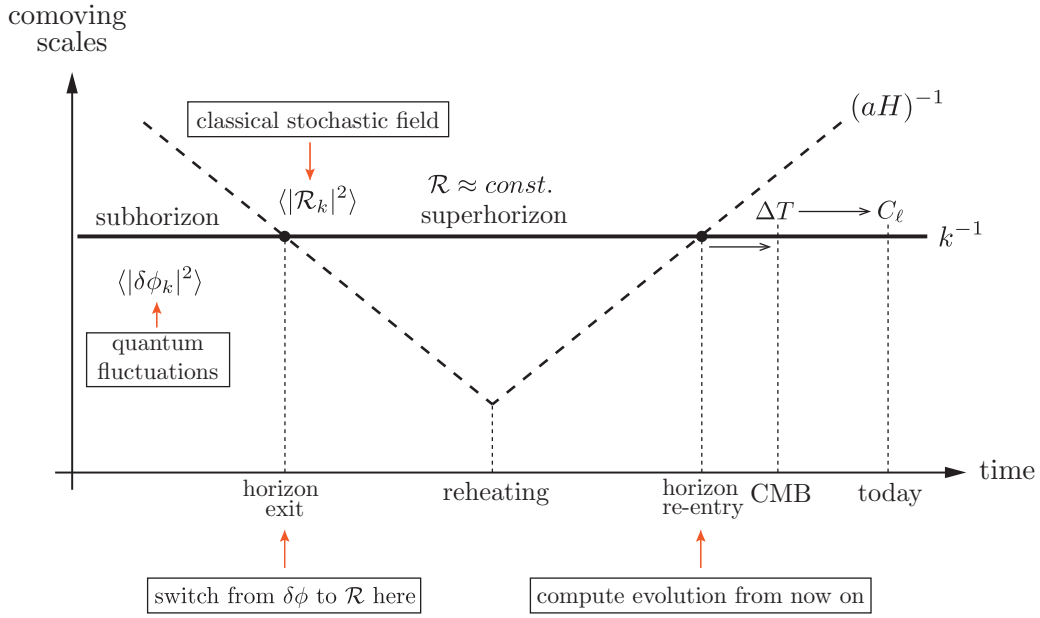


Figure 6.2: *Curvature perturbations during and after inflation: The comoving horizon $(aH)^{-1}$ shrinks during inflation and grows in the subsequent FRW evolution. This implies that comoving scales k^{-1} exit the horizon at early times and re-enter the horizon at late times. While the curvature perturbations $\mathcal{R}$ are outside of the horizon they don't evolve, so our computation for the correlation function $\langle|\mathcal{R}_k|^2\rangle$ at horizon exit during inflation can be related directly to observables at late times.*

At horizon crossing, $k = aH$, it is convenient to switch from inflaton fluctuations $\delta\phi$ to fluctuations in the conserved curvature perturbations $\mathcal{R}$. The relationship between $\mathcal{R}$ and $\delta\phi$ is simplest in *spatially flat gauge* :

$$\boxed{\mathcal{R} = -\frac{\mathcal{H}}{\bar{\phi}'}\,\delta\phi} \ . \tag{6.1.2}$$

$\delta\phi \to \mathcal{R}$.—From the gauge-invariant definition of $\mathcal{R}$, eq. (4.3.159), we get

$$\mathcal{R} = C - \frac{1}{3}\nabla^2 E + \mathcal{H}(B+v) \xrightarrow{\text{spatially flat}} \mathcal{H}(B+v) \ . \tag{6.1.3}$$

We recall that the combination $B+v$ appeared in the off-diagonal component of the perturbed stress tensor, cf. eq. (4.2.76),

$$\delta T^0{}_j = -(\bar{\rho} + \bar{P})\partial_j(B+v) \ . \tag{6.1.4}$$

We compare this to the first-order perturbation of the stress tensor of a scalar field, cf. eq. (2.3.26),

$$\delta T^0{}_j = g^{0\mu}\partial_\mu\phi\partial_j\delta\phi = \bar{g}^{00}\partial_0\bar{\phi}\partial_j\delta\phi = \frac{\bar{\phi}'}{a^2}\partial_j\delta\phi \ , \tag{6.1.5}$$

to get

$$B + v = -\frac{\delta\phi}{\bar{\phi}'} \ . \tag{6.1.6}$$

Substituting (6.1.6) into (6.1.3) we obtain (6.1.2).

The variance of curvature perturbations therefore is

$$\langle|\mathcal{R}_k|^2\rangle = \left(\frac{\mathcal{H}}{\bar{\phi}'}\right)^2 \langle|\delta\phi_k|^2\rangle \ , \tag{6.1.7}$$

where $\delta\phi$ are the inflaton fluctuations in spatially flat gauge.

Outside the horizon, the quantum nature of the field disappears and the quantum expectation value can be identified with the ensemble average of a classical stochastic field. The conservation of $\mathcal{R}$ on superhorizon scales then allows us to relate predictions made at horizon exit (high energies) to the observables after horizon re-entry (low energies). These times are separated by a time interval in which the physics is very uncertain. Not even the equations governing perturbations are well-known. The only reason that we are able to connect late-time observables to inflationary theories is the fact that the wavelengths of the perturbations of interest were outside the horizon during the period from well before the end of inflation until the relatively near present. After horizon re-entry the fluctuations evolve in a computable way.

The rest of this chapter will develop this beautiful story in more detail: In §6.2, we show that inflaton fluctuations in the subhorizon limit can be described as a collection of simple harmonic oscillators. In §6.3, we therefore review the canonical quantisation of a simple harmonic oscillator in quantum mechanics. In particular, we compute the variance of the oscillator amplitude induced by zero-point fluctuations in the ground state. In §6.4, we apply the same techniques to the quantisation of inflaton fluctuations in the inflationary quasi-de Sitter background. In §6.5, we relate this result to the power spectrum of primordial curvature perturbations. We also derive the spectrum of gravitational waves predicted by inflation. Finally, we discuss how late-time observations probe the inflationary initial conditions.

6.2 Classical Oscillators

We first wish to show that the dynamics of inflaton fluctuations on small scales is described by a collection of harmonic oscillators.

6.2.1 Mukhanov-Sasaki Equation

It will be useful to start from the inflaton action (see Problem Set 2)

$$S = \int \mathrm{d}\tau\, \mathrm{d}^3x\, \sqrt{-g}\left[\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right] \ , \tag{6.2.8}$$

where $g \equiv \det(g_{\mu\nu})$. To study the linearised dynamics, we need the action at quadratic order in fluctuations. In spatially flat gauge, the metric perturbations δg_{00} and δg_{0i} are suppressed relative to the inflaton fluctuations by factors of the slow-roll parameter ε. This means that at leading order in the slow-roll expansion, we can ignore the fluctuations in the spacetime geometry and perturb the inflaton field independently. (In a general gauge, we would have to study to coupled dynamics of inflaton and metric perturbations.)

Evaluating (6.2.8) for the *unperturbed* FRW metric, we find

$$S = \int \mathrm{d}\tau \mathrm{d}^3x\, \left[\frac{1}{2}a^2\left((\phi')^2 - (\nabla\phi)^2\right) - a^4V(\phi)\right] \ . \tag{6.2.9}$$

It is convenient to write the perturbed inflaton field as

$$\phi(\tau, \boldsymbol{x}) = \bar{\phi}(\tau) + \frac{f(\tau, \boldsymbol{x})}{a(\tau)} \ . \tag{6.2.10}$$

To get the linearised equation of motion for $f(\tau, \boldsymbol{x})$, we need to expand the action (6.2.9) to second order in the fluctuations:

- Collecting all terms with single powers of the field f, we have

$$S^{(1)} = \int \mathrm{d}\tau \mathrm{d}^3x \left[a\bar{\phi}' f' - a'\bar{\phi}' f - a^3 V_{,\phi} f \right] , \tag{6.2.11}$$

  where $V_{,\phi}$ denotes the derivative of V with respect to ϕ. Integrating the first term by parts (and dropping the boundary term), we find

$$\begin{aligned} S^{(1)} &= -\int \mathrm{d}\tau \mathrm{d}^3x \left[\partial_\tau (a\bar{\phi}') + a'\bar{\phi}' + a^3 V_{,\phi} \right] f \ , \\ &= -\int \mathrm{d}\tau \mathrm{d}^3x \, a \left[\bar{\phi}'' + 2\mathcal{H}\bar{\phi}' + a^2 V_{,\phi} \right] f \ . \end{aligned} \tag{6.2.12}$$

  Requiring that $S^{(1)} = 0$, for all f, gives the Klein-Gordon equation for the background field,

$$\bar{\phi}'' + 2\mathcal{H}\bar{\phi}' + a^2 V_{,\phi} = 0 \ . \tag{6.2.13}$$

- Isolating all terms with two factors of f, we get the quadratic action

$$S^{(2)} = \frac{1}{2} \int \mathrm{d}\tau \mathrm{d}^3x \left[(f')^2 - (\nabla f)^2 - 2\mathcal{H} f f' + \left(\mathcal{H}^2 - a^2 V_{,\phi\phi} \right) f^2 \right] . \tag{6.2.14}$$

  Integrating the $ff' = \frac{1}{2}(f^2)'$ term by parts, gives

$$\begin{aligned} S^{(2)} &= \frac{1}{2} \int \mathrm{d}\tau \mathrm{d}^3x \left[(f')^2 - (\nabla f)^2 + \left(\mathcal{H}' + \mathcal{H}^2 - a^2 V_{,\phi\phi} \right) f^2 \right] , \\ &= \frac{1}{2} \int \mathrm{d}\tau \mathrm{d}^3x \left[(f')^2 - (\nabla f)^2 + \left(\frac{a''}{a} - a^2 V_{,\phi\phi} \right) f^2 \right] . \end{aligned} \tag{6.2.15}$$

  During slow-roll inflation, we have

$$\frac{V_{,\phi\phi}}{H^2} \approx \frac{3M_{\rm pl}^2 V_{,\phi\phi}}{V} = 3\eta_{\rm v} \ll 1 \ . \tag{6.2.16}$$

  Since $a' = a^2 H$, with $H \approx const.$, we also have

$$\frac{a''}{a} \approx 2a'H = 2a^2 H^2 \gg a^2 V_{,\phi\phi} \ . \tag{6.2.17}$$

  Hence, we can drop the $V_{,\phi\phi}$ term in (6.2.15),

$$\boxed{S^{(2)} \approx \int \mathrm{d}\tau \mathrm{d}^3x \ \frac{1}{2} \left[(f')^2 - (\nabla f)^2 + \frac{a''}{a} f^2 \right]} \ . \tag{6.2.18}$$

  Applying the Euler-Lagrange equation to (6.2.18) gives the *Mukhanov-Sasaki equation*

$$f'' - \nabla^2 f - \frac{a''}{a} f = 0 \ , \tag{6.2.19}$$

  or, for each Fourier mode,

$$\boxed{f_{\boldsymbol{k}}'' + \left(k^2 - \frac{a''}{a} \right) f_{\boldsymbol{k}} = 0} \ . \tag{6.2.20}$$

6.2.2 Subhorizon Limit

On subhorizon scales, $k^2 \gg a''/a \approx 2\mathcal{H}^2$, the Mukhanov-Sasaki equation reduces to

$$f_{\boldsymbol{k}}'' + k^2 f_{\boldsymbol{k}} \approx 0 \ . \tag{6.2.21}$$

We see that each Fourier mode satisfies the equation of motion of a *simple harmonic oscillator*, with frequency $\omega_k = k$. Quantum zero-point fluctuations of these oscillators provide the origin of structure in the universe.

6.3 Quantum Oscillators

Our aim is to quantise the field f following the standard methods of quantum field theory. However, before we do this, let us study a slightly simpler problem[1]: the quantum mechanics of a one-dimensional harmonic oscillator. The oscillator has coordinate q, mass $m \equiv 1$ and quadratic potential $V(q) = \frac{1}{2}\omega^2 q^2$. The action therefore is

$$S[q] = \frac{1}{2}\int \mathrm{d}t \Big[\dot{q}^2 - \omega^2 q^2\Big] \ , \tag{6.3.22}$$

and the equation of motion is $\ddot{q} + \omega^2 q = 0$. The conjugate momentum is

$$p = \frac{\partial L}{\partial \dot{q}} = \dot{q} \ . \tag{6.3.23}$$

6.3.1 Canonical Quantisation

Let me now remind you how to quantise the harmonic oscillator: First, we promote the classical variables q, p to quantum operators $\hat{q}$, $\hat{p}$ and impose the canonical commutation relation (CCR)

$$[\hat{q}, \hat{p}] = i \ , \tag{I}$$

in units where $\hbar \equiv 1$. The equation of motion implies that the commutator holds at all times if imposed at some initial time. Note that we are in the Heisenberg picture where operators vary in time while states are time-independent. The operator solution $\hat{q}(t)$ is determined by two initial conditions $\hat{q}(0)$ and $\hat{p}(0) = \partial_t \hat{q}(0)$. Since the evolution equation is linear, the solution is linear in these operators. It is convenient to trade $\hat{q}(0)$ and $\hat{p}(0)$ for a single time-independent non-Hermitian operator $\hat{a}$, in terms of which the solution can be written as

$$\hat{q}(t) = q(t)\,\hat{a} + q^*(t)\,\hat{a}^\dagger \ , \tag{II}$$

where the (complex) mode function $q(t)$ satisfies the classical equation of motion, $\ddot{q}+\omega^2 q = 0$. Of course, $q^*(t)$ is the complex conjugate of $q(t)$ and $\hat{a}^\dagger$ is the Hermitian conjugate of a. Substituting (**II**) into (**I**), we get

$$W[q, q^*] \times [\hat{a}, \hat{a}^\dagger] = 1 \, , \tag{6.3.24}$$

where we have defined the *Wronskian* as

$$W[q_1, q_2^*] \equiv -i\,(q_1 \partial_t q_2^* - (\partial_t q_1) q_2^*) \ . \tag{6.3.25}$$

[1] The reason it looks simpler is that it avoids distractions arising from Fourier labels, etc. The physics is exactly the same.

Without loss of generality, let us assume that the solution q is chosen so that the real number $W[q, q^*]$ is positive. The function q can then be rescaled ($q \to \lambda q$) such that

$$W[q, q^*] \equiv 1 \ , \tag{III}$$

and hence

$$[\hat{a}, \hat{a}^\dagger] = 1 \ . \tag{IV}$$

Eq. (**IV**) is the standard commutation relation for the *raising* and *lowering operators* of the harmonic oscillator. The *vacuum state* $|0\rangle$ is annihilated by the operator $\hat{a}$

$$\hat{a}|0\rangle = 0 \ . \tag{V}$$

Excited states are created by repeated application of creation operators

$$|n\rangle \equiv \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n|0\rangle \, . \tag{6.3.26}$$

These states are eigenstates of the number operator $\hat{N} \equiv \hat{a}^\dagger\hat{a}$ with eigenvalue n, i.e.

$$\hat{N}|n\rangle = n|n\rangle \, . \tag{6.3.27}$$

6.3.2 Choice of Vacuum

At this point, we have only imposed the normalisation $W[q, q^*] = 1$ on the mode functions. A change in $q(t)$ could be accompanied by a change in $\hat{a}$ that keeps the solution $\hat{q}(t)$ unchanged. Via eq. (**V**), each such solution corresponds to a different vacuum state. However, a special choice of $q(t)$ is selected if we require that the vacuum state $|0\rangle$ be the ground state of the Hamiltonian. To see this, consider the Hamiltonian for general $q(t)$,

$$\begin{aligned} \hat{H} &= \frac{1}{2}\hat{p}^2 + \frac{1}{2}\omega^2\hat{q}^2 \\ &= \frac{1}{2}\left[(\dot{q}^2 + \omega^2 q^2)\hat{a}\hat{a} + (\dot{q}^2 + \omega^2 q^2)^* \, \hat{a}^\dagger\hat{a}^\dagger + (|\dot{q}|^2 + \omega^2|q|^2)(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a})\right] \ . \end{aligned} \tag{6.3.28}$$

Using $\hat{a}|0\rangle = 0$ and $[\hat{a}, \hat{a}^\dagger] = 1$, we can determine how the Hamiltonian operator acts on the vacuum state

$$\hat{H}|0\rangle = \frac{1}{2}(\dot{q}^2 + \omega^2 q^2)^* \, \hat{a}^\dagger\hat{a}^\dagger|0\rangle + \frac{1}{2}(|\dot{q}|^2 + \omega^2|q|^2)|0\rangle \ . \tag{6.3.29}$$

We want $|0\rangle$ to be an eigenstate of $\hat{H}$. For this to be the case, the first term in (6.3.29) must vanish, which implies

$$\dot{q} = \pm \, i\omega q \ . \tag{6.3.30}$$

For such a function q, the norm is

$$W[q, q^*] = \mp \, 2\omega|q|^2 \ , \tag{6.3.31}$$

and positivity of the normalisation condition $W[q, q^*] > 0$ selects the minus sign in (6.3.30)

$$\dot{q} = -i\omega q \qquad \Rightarrow \qquad q(t) \propto e^{-i\omega t} \ . \tag{6.3.32}$$

Asking the vacuum state to be the ground state of the Hamiltonian has therefore selected the *positive-frequency solution* $e^{-i\omega t}$ (rather than the negative-frequency solution $e^{+i\omega t}$). Imposing the normalisation $W[q, q^*] = 1$, we get

$$\boxed{q(t) = \frac{1}{\sqrt{2\omega}}\, e^{-i\omega t}} \,. \tag{6.3.33}$$

With this choice of mode function, the Hamiltonian takes the familiar form

$$\hat{H} = \hbar\omega \left(\hat{N} + \frac{1}{2}\right) \,. \tag{6.3.34}$$

We see that the vacuum $|0\rangle$ is the state of minimum energy $\frac{1}{2}\hbar\omega$. If any function other than (6.3.33) is chosen to expand the position operator, then the state annihilated by $\hat{a}$ is *not* the ground state of the oscillator.

6.3.3 Zero-Point Fluctuations

The expectation value of the position operator $\hat{q}$ in the ground state $|0\rangle$ vanishes

$$\begin{aligned} \langle\hat{q}\rangle &\equiv \langle 0|\hat{q}|0\rangle \\ &= \langle 0|q(t)\hat{a} + q^*(t)\hat{a}^\dagger|0\rangle \\ &= 0 \,, \end{aligned} \tag{6.3.35}$$

because $\hat{a}$ annihilates $|0\rangle$ when acting on it from the left, and $\hat{a}^\dagger$ annihilates $\langle 0|$ when acting on it from the right. However, the expectation value of the square of the position operator receives finite zero-point fluctuations

$$\begin{aligned} \langle|\hat{q}|^2\rangle &\equiv \langle 0|\hat{q}^\dagger\hat{q}|0\rangle \\ &= \langle 0|(q^*\hat{a}^\dagger + q\hat{a})(q\hat{a} + q^*\hat{a}^\dagger)|0\rangle \\ &= |q(t)|^2\langle 0|\hat{a}\hat{a}^\dagger|0\rangle \\ &= |q(t)|^2\langle 0|[\hat{a}, \hat{a}^\dagger]|0\rangle \\ &= |q(t)|^2 \,. \end{aligned} \tag{6.3.36}$$

Hence, we find that the variance of the amplitude of the quantum oscillator is given by the square of the mode function

$$\langle|\hat{q}|^2\rangle = |q(t)|^2 = \frac{\hbar}{2\omega} \,. \tag{**VI**}$$

To make the quantum nature of the result manifest, we have reinstated Planck's constant $\hbar$. This is all we need to know about quantum mechanics in order to compute the fluctuation spectrum created by inflation.

6.4 Quantum Fluctuations in de Sitter Space

Let us return to the quadratic action (6.2.18) for the inflaton fluctuation $f = a\delta\phi$. The momentum conjugate to f is

$$\pi \equiv \frac{\partial \mathcal{L}}{\partial f'} = f' \ . \tag{6.4.37}$$

We perform the canonical quantisation just like in the case of the harmonic oscillator.

6.4.1 Canonical Quantisation

We promote the fields $f(\tau, \boldsymbol{x})$ and $\pi(\tau, \boldsymbol{x})$ to quantum operators $\hat{f}(\tau, \boldsymbol{x})$ and $\hat{\pi}(\tau, \boldsymbol{x})$. The operators satisfy the equal time CCR

$$\boxed{[\hat{f}(\tau, \boldsymbol{x}), \hat{\pi}(\tau, \boldsymbol{x}')] = i\delta(\boldsymbol{x} - \boldsymbol{x}')} \ . \tag{I$'$}$$

This is the field theory equivalent of eq. (**I**). The delta function is a signature of *locality*: modes at different points in space are independent and the corresponding operators therefore commute. In Fourier space, we find

$$\begin{aligned}
[\hat{f}_{\boldsymbol{k}}(\tau), \hat{\pi}_{\boldsymbol{k}'}(\tau)] &= \int \frac{\mathrm{d}^3x}{(2\pi)^{3/2}} \int \frac{\mathrm{d}^3x'}{(2\pi)^{3/2}} \underbrace{[\hat{f}(\tau, \boldsymbol{x}), \hat{\pi}(\tau, \boldsymbol{x}')]}_{i\delta(\boldsymbol{x} - \boldsymbol{x}')} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} e^{-i\boldsymbol{k}'\cdot\boldsymbol{x}'} \\
&= i \int \frac{\mathrm{d}^3x}{(2\pi)^3} \, e^{-i(\boldsymbol{k}+\boldsymbol{k}')\cdot\boldsymbol{x}} \\
&= i\delta(\boldsymbol{k} + \boldsymbol{k}') \ ,
\end{aligned} \tag{I$''$}$$

where the delta function implies that modes with different wavelengths commute. Eq. (**I**$''$) is the same as (**I**), but for each independent Fourier mode. The generalisation of the mode expansion (**II**) is

$$\boxed{\hat{f}_{\boldsymbol{k}}(\tau) = f_k(\tau)\, \hat{a}_{\boldsymbol{k}} + f_k^*(\tau)\, a_{\boldsymbol{k}}^\dagger} \ , \tag{II$'$}$$

where $\hat{a}_{\boldsymbol{k}}$ is a time-independent operator, $a_{\boldsymbol{k}}^\dagger$ is its Hermitian conjugate, and $f_k(\tau)$ and its complex conjugate $f_k^*(\tau)$ are two linearly independent solutions of the Mukhanov-Sasaki equation

$$f_k'' + \omega_k^2(\tau) f_k = 0 \ , \qquad \text{where} \quad \omega_k^2(\tau) \equiv k^2 - \frac{a''}{a} \ . \tag{6.4.38}$$

As indicated by dropping the vector notation $\boldsymbol{k}$ on the subscript, the mode functions, $f_k(\tau)$ and $f_k^*(\tau)$, are the same for all Fourier modes with $k \equiv |\boldsymbol{k}|$.[2]

Substituting (**II**$'$) into (**I**$''$), we get

$$W[f_k, f_k^*] \times [\hat{a}_{\boldsymbol{k}}, \hat{a}_{\boldsymbol{k}'}^\dagger] = \delta(\boldsymbol{k} + \boldsymbol{k}') \ , \tag{6.4.39}$$

where $W[f_k, f_k^*]$ is the Wronskian (6.3.25) of the mode functions. As before, cf. (**III**), we can choose to normalize f_k such that

$$W[f_k, f_k^*] \equiv 1 \ . \tag{III$'$}$$

[2]Since the frequency $\omega_k(\tau)$ depends only on $k \equiv |\boldsymbol{k}|$, the evolution does not depend on direction. The constant operators $\hat{a}_{\boldsymbol{k}}$ and $\hat{a}_{\boldsymbol{k}}^\dagger$ define initial conditions which may depend on direction.

Eq. (6.4.39) then becomes

$$\boxed{[\hat{a}_{\boldsymbol{k}}, \hat{a}^\dagger_{\boldsymbol{k}'}] = \delta(\boldsymbol{k} + \boldsymbol{k}')} \,, \tag{IV$'$}$$

which is the same as (**IV**), but for each Fourier mode.

As before, the operators $\hat{a}^\dagger_{\boldsymbol{k}}$ and $\hat{a}_{\boldsymbol{k}}$ may be interpreted as creation and annihilation operators, respectively. As in (**V**), the quantum states in the Hilbert space are constructed by defining the vacuum state $|0\rangle$ via

$$\hat{a}_{\boldsymbol{k}}|0\rangle = 0 \,, \tag{V$'$}$$

and by producing excited states by repeated application of creation operators

$$|m_{\boldsymbol{k}_1}, n_{\boldsymbol{k}_2}, \cdots\rangle = \frac{1}{\sqrt{m!n!\cdots}} \left[(a^\dagger_{\boldsymbol{k}_1})^m (a^\dagger_{\boldsymbol{k}_2})^n \cdots\right] |0\rangle \,. \tag{6.4.40}$$

6.4.2 Choice of Vacuum

As before, we still need to fix the mode function in order to define the vacuum state. Although for general time-dependent backgrounds this procedure can be ambiguous, for inflation there is a preferred choice. To motivate the inflationary vacuum state, let us go back to fig. 6.2. We see that at sufficiently early times (large negative conformal time τ) all modes of cosmological interest were deep inside the horizon, $k/\mathcal{H} \sim |k\tau| \gg 1$. This means that in the remote past all observable modes had time-independent frequencies

$$\omega_k^2 \,=\, k^2 - \frac{a''}{a} \,\approx\, k^2 - \frac{2}{\tau^2} \,\xrightarrow{\tau\to-\infty}\, k^2 \,, \tag{6.4.41}$$

and the Mukhanov-Sasaki equation reduces to

$$f_k'' + k^2 f_k \approx 0 \,. \tag{6.4.42}$$

But this is just the equation for a free field in Minkowkski space, whose two independent solutions are $f_k \propto e^{\pm ik\tau}$. As we have seen above, only the positive frequency mode $f_k \propto e^{-ik\tau}$ corresponds to the 'minimal excitation state', cf. eq. (6.3.33). We will choose this mode to define the inflationary vacuum state. In practice, this means solving the Mukhanov-Sasaki equation with the (Minkowski) initial condition

$$\boxed{\lim_{\tau\to-\infty} f_k(\tau) = \frac{1}{\sqrt{2k}} e^{-ik\tau}} \,. \tag{6.4.43}$$

This defines a preferable set of mode functions and a unique physical vacuum, the *Bunch-Davies vacuum*.

For slow-roll inflation, it will be sufficient to study the Mukhanov-Sasaki equation in de Sitter space[3]

$$f_k'' + \left(k^2 - \frac{2}{\tau^2}\right) f_k = 0 \,. \tag{6.4.44}$$

This has an exact solution

$$f_k(\tau) = \alpha \frac{e^{-ik\tau}}{\sqrt{2k}} \left(1 - \frac{i}{k\tau}\right) + \beta \frac{e^{ik\tau}}{\sqrt{2k}} \left(1 + \frac{i}{k\tau}\right) \,. \tag{6.4.45}$$

[3] See Problem Set 4 for a slightly more accurate treatment.

where α and β are constants that are fixed by the initial conditions. In fact, the initial condition (6.4.43) fixes $\beta = 0$, $\alpha = 1$, and, hence, the mode function is

$$\boxed{f_k(\tau) = \frac{e^{-ik\tau}}{\sqrt{2k}}\left(1 - \frac{i}{k\tau}\right)} \,. \tag{6.4.46}$$

Since the mode function is completely fixed, the future evolution of the mode including its superhorizon dynamics is determined.

6.4.3 Zero-Point Fluctuations

Finally, we can predict the quantum statistics of the operator

$$\hat{f}(\tau, \boldsymbol{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}} \left[f_k(\tau)\,\hat{a}_{\boldsymbol{k}} + f_k^*(\tau)\,a_{\boldsymbol{k}}^\dagger\right] e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \,. \tag{6.4.47}$$

As before, the expectation value of $\hat{f}$ vanishes, i.e. $\langle\hat{f}\rangle = 0$. However, the variance of inflaton fluctuations receive non-zero quantum fluctuations

$$\begin{aligned}
\langle|\hat{f}|^2\rangle &\equiv \langle 0|\hat{f}^\dagger(\tau,\mathbf{0})\hat{f}(\tau,\mathbf{0})|0\rangle \\
&= \int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}} \int \frac{\mathrm{d}^3k'}{(2\pi)^{3/2}}\, \langle 0|\big(f_k^*(\tau)\hat{a}_{\boldsymbol{k}}^\dagger + f_k(\tau)\hat{a}_{\boldsymbol{k}}\big)\big(f_{k'}(\tau)\hat{a}_{\boldsymbol{k}'} + f_{k'}^*(\tau)\hat{a}_{\boldsymbol{k}'}^\dagger\big)|0\rangle \\
&= \int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}} \int \frac{\mathrm{d}^3k'}{(2\pi)^{3/2}}\, f_k(\tau)f_{k'}^*(\tau)\,\langle 0|[\hat{a}_{\boldsymbol{k}}, \hat{a}_{\boldsymbol{k}'}^\dagger]|0\rangle \\
&= \int \frac{\mathrm{d}^3k}{(2\pi)^3}\,|f_k(\tau)|^2 \\
&= \int \mathrm{d}\ln k\;\frac{k^3}{2\pi^2}|f_k(\tau)|^2 \,.
\end{aligned} \tag{6.4.48}$$

We define the (dimensionless) *power spectrum* as

$$\boxed{\Delta_f^2(k,\tau) \equiv \frac{k^3}{2\pi^2}|f_k(\tau)|^2} \,. \tag{VI$'$}$$

As in (**VI**), the square of the classical solution determines the variance of quantum fluctuations. Using (6.4.46), we find

$$\Delta_{\delta\phi}^2(k,\tau) = a^{-2}\Delta_f^2(k,\tau) = \left(\frac{H}{2\pi}\right)^2\left(1 + \left(\frac{k}{aH}\right)^2\right) \xrightarrow{\text{superhorizon}} \left(\frac{H}{2\pi}\right)^2 \,. \tag{6.4.49}$$

We will use the approximation that the power spectrum at horizon crossing is[4]

$$\boxed{\Delta_{\delta\phi}^2(k) \approx \left(\frac{H}{2\pi}\right)^2\Bigg|_{k=aH}} \,. \tag{6.4.50}$$

[4]Computing the power spectrum at a specific instant (horizon crossing, $aH = k$) implicitly extends the result for the pure de Sitter background to a slowly time-evolving quasi-de Sitter space. Different modes exit the horizon as slightly different times when aH has a different value. Evaluating the fluctuations at horizon crossing also has the added benefit that the error we are making by ignoring the metric fluctuations in spatially flat gauge doesn't accumulate over time.

6.4.4 Quantum-to-Classical Transition*

When do the fluctuations become classical? Consider the quantum operator (**II′**) and its conjugate momentum operator

$$\hat{f}(\tau,\boldsymbol{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}} \left[f_k(\tau)\,\hat{a}_{\boldsymbol{k}} + f_k^*(\tau)\,a_{\boldsymbol{k}}^\dagger \right] e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \,, \tag{6.4.51}$$

$$\hat{\pi}(\tau,\boldsymbol{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}} \left[f_k'(\tau)\,\hat{a}_{\boldsymbol{k}} + (f_k^*)'(\tau)\,a_{\boldsymbol{k}}^\dagger \right] e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \,. \tag{6.4.52}$$

In the superhorizon limit, $k\tau \to 0$, we have

$$f_k(\tau) \approx -\frac{1}{\sqrt{2}k^{3/2}}\frac{i}{\tau} \quad \text{and} \quad f_k'(\tau) \approx \frac{1}{\sqrt{2}k^{3/2}}\frac{i}{\tau^2} \,, \tag{6.4.53}$$

and hence

$$\hat{f}(\tau,\boldsymbol{x}) = -\frac{i}{\sqrt{2}\tau}\int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}}\frac{1}{k^{3/2}} \left[\hat{a}_{\boldsymbol{k}} - a_{\boldsymbol{k}}^\dagger \right] e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \,, \tag{6.4.54}$$

$$\hat{\pi}(\tau,\boldsymbol{x}) = \frac{i}{\sqrt{2}\tau^2}\int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}}\frac{1}{k^{3/2}} \left[\hat{a}_{\boldsymbol{k}} - a_{\boldsymbol{k}}^\dagger \right] e^{i\boldsymbol{k}\cdot\boldsymbol{x}} = -\frac{1}{\tau}\hat{f}(\tau,\boldsymbol{x}) \,. \tag{6.4.55}$$

The two operators have become proportional to each other and therefore *commute* on superhorizon scales. This is the signature of classical (rather than quantum) modes. After horizon crossing, the inflaton fluctuation $\delta\phi$ can therefore be viewed as a classical stochastic field and we can identify the quantum expectation value with a classical ensemble average.

6.5 Primordial Perturbations from Inflation

6.5.1 Curvature Perturbations

At horizon crossing, we switch from the inflaton fluctuation $\delta\phi$ to the conserved curvature perturbation $\mathcal{R}$. The power spectra of $\mathcal{R}$ and $\delta\phi$ are related via eq. (6.1.7),

$$\Delta^2_{\mathcal{R}} = \frac{1}{2\varepsilon}\frac{\Delta^2_{\delta\phi}}{M_{\rm pl}^2} \,, \qquad \text{where} \quad \varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M_{\rm pl}^2 H^2} \,. \tag{6.5.56}$$

Substituting (6.4.50), we get

$$\boxed{\Delta^2_{\mathcal{R}}(k) = \frac{1}{8\pi^2}\frac{1}{\varepsilon}\frac{H^2}{M_{\rm pl}^2}\bigg|_{k=aH}} \,. \tag{6.5.57}$$

Exercise.—Show that for slow-roll inflation, eq. (6.5.59) can be written as

$$\boxed{\Delta^2_{\mathcal{R}} = \frac{1}{12\pi^2}\frac{V^3}{M_{\rm pl}^6 (V')^2}} \,. \tag{6.5.58}$$

This expresses the amplitude of curvature perturbations in terms of the shape of the inflaton potential.

Because the right-hand side of (6.5.59) is evaluated at $k = aH$, the power spectrum is purely a function of k. If $\Delta^2_{\mathcal{R}}(k)$ is k-independent, then we call the spectrum *scale-invariant*. However, since H and possibly ε are (slowly-varying) functions of time, we predict that the power spectrum will deviate slightly from the scale-invariant form $\Delta^2_{\mathcal{R}} \sim k^0$. Near a reference scale $k_\star$, the k-dependence of the spectrum takes a power-law form

$$\Delta^2_{\mathcal{R}}(k) \equiv A_s \left(\frac{k}{k_\star}\right)^{n_s - 1} . \tag{6.5.59}$$

The measured amplitude of the scalar spectrum at $k_\star = 0.05\ \mathrm{Mpc}^{-1}$ is

$$A_s = (2.196 \pm 0.060) \times 10^{-9} . \tag{6.5.60}$$

To quantify the deviation from scale-invariance we have introduced the *scalar spectral index*

$$n_s - 1 \equiv \frac{d \ln \Delta^2_{\mathcal{R}}}{d \ln k} , \tag{6.5.61}$$

where the right-hand side is evaluated at $k = k_\star$ and $n_s = 1$ corresponds to perfect scale-invariance. We can split (6.5.61) into two factors

$$\frac{d \ln \Delta^2_{\mathcal{R}}}{d \ln k} = \frac{d \ln \Delta^2_{\mathcal{R}}}{dN} \times \frac{dN}{d \ln k} . \tag{6.5.62}$$

The derivative with respect to e-folds is

$$\frac{d \ln \Delta^2_{\mathcal{R}}}{dN} = 2 \frac{d \ln H}{dN} - \frac{d \ln \varepsilon}{dN} . \tag{6.5.63}$$

The first term is just -2ε and the second term is $-\eta$ (see Chapter 2). The second factor in (6.5.62) is evaluated by recalling the horizon crossing condition $k = aH$, or

$$\ln k = N + \ln H . \tag{6.5.64}$$

Hence, we have

$$\frac{dN}{d \ln k} = \left[\frac{d \ln k}{dN}\right]^{-1} = \left[1 + \frac{d \ln H}{dN}\right]^{-1} \approx 1 + \varepsilon . \tag{6.5.65}$$

To first order in the Hubble slow-roll parameters, we therefore find

$$\boxed{n_s - 1 = -2\varepsilon - \eta} . \tag{6.5.66}$$

The parameter n_s is an interesting probe of the inflationary dynamics. It measures deviations from the perfect de Sitter limit: H, $\dot{H}$, and $\ddot{H}$. Observations have recently detected the small deviation from scale-invariance predicted by inflation

$$n_s = 0.9603 \pm 0.0073 . \tag{6.5.67}$$

Exercise.—For slow-roll inflation, show that

$$\boxed{n_s - 1 = -3M_{\rm pl}^2 \left(\frac{V'}{V}\right)^2 + 2M_{\rm pl}^2 \frac{V''}{V}} . \tag{6.5.68}$$

This relates the value of the spectral index to the shape of the inflaton potential.

6.5.2 Gravitational Waves

Arguably the cleanest prediction of inflation is a spectrum of primordial gravitational waves. These are tensor perturbations to the spatial metric,

$$\mathrm{d}s^2 = a^2(\tau)\left[\mathrm{d}\tau^2 - (\delta_{ij} + 2\hat{E}_{ij})\mathrm{d}x^i\mathrm{d}x^j\right] \ . \tag{6.5.69}$$

We won't go through the details of the quantum production of tensor fluctuations during inflation, but just sketch the logic which is identical to the scalar case (and even simpler).

Substituting (6.5.69) into the Einstein-Hilbert action and expanding to second order gives

$$S = \frac{M_{\rm pl}^2}{2}\int \mathrm{d}^4x\sqrt{-g}\,R \qquad \Rightarrow \qquad S^{(2)} = \frac{M_{\rm pl}^2}{8}\int \mathrm{d}\tau\,\mathrm{d}^3x\,a^2\left[(\hat{E}_{ij}')^2 - (\nabla\hat{E}_{ij})^2\right] \ . \tag{6.5.70}$$

It is convenient to define

$$\frac{M_{\rm pl}}{2}\,a\hat{E}_{ij} \equiv \frac{1}{\sqrt{2}}\begin{pmatrix} f_+ & f_\times & 0 \\ f_\times & -f_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} \ , \tag{6.5.71}$$

so that

$$S^{(2)} = \frac{1}{2}\sum_{I=+,\times}\int \mathrm{d}\tau\mathrm{d}^3x\,\left[(f_I')^2 - (\nabla f_I)^2 + \frac{a''}{a}f_I^2\right] \ . \tag{6.5.72}$$

This is just two copies of the action (6.2.18) for $f = a\delta\phi$, one for each polarization mode of the gravitational wave, $f_{+,\times}$. The power spectrum of tensor modes Δ_t^2 can therefore be inferred directly from our previous result for Δ_f^2,

$$\Delta_t^2 \equiv 2\times\Delta_{\hat{E}}^2 = 2\times\left(\frac{2}{aM_{\rm pl}}\right)^2\times\Delta_f^2 \ . \tag{6.5.73}$$

Using (6.4.50), we get

$$\boxed{\Delta_t^2(k) = \frac{2}{\pi^2}\frac{H^2}{M_{\rm pl}^2}\bigg|_{k=aH}} \ . \tag{6.5.74}$$

This result is the most robust and model-independent prediction of inflation. Notice that the tensor amplitude is a direct measure of the expansion rate H during inflation. This is in contrast to the scalar amplitude which depends on both H and ε.

The scale-dependence of the tensor spectrum is defined in analogy to (6.5.59) as

$$\Delta_t^2(k) \equiv A_t\left(\frac{k}{k_\star}\right)^{n_t} \ , \tag{6.5.75}$$

where n_t is the *tensor spectral index*. Scale-invariance now corresponds to $n_t = 0$. (The different conventions for the scalar and tensor spectral indices are an unfortunate historical accident.) Often the amplitude of tensors is normalised with respect to the measured scalar amplitude (6.5.60), i.e. one defines the *tensor-to-scalar ratio*

$$r \equiv \frac{A_t}{A_s} \ . \tag{6.5.76}$$

Tensors have not been observed yet, so we only have an upper limit on their amplitude, $r \lesssim 0.17$.

Exercise.—Show that

$$r = 16\varepsilon \tag{6.5.77}$$

$$n_t = -2\varepsilon \ . \tag{6.5.78}$$

Notice that this implies the consistency relation $n_t = -r/8$.

Inflationary models can be classified according to their predictions for the parameters n_s and r. Fig. 6.3 shows the predictions of various slow-roll models as well as the latest constraints from measurements of the Planck satellite.

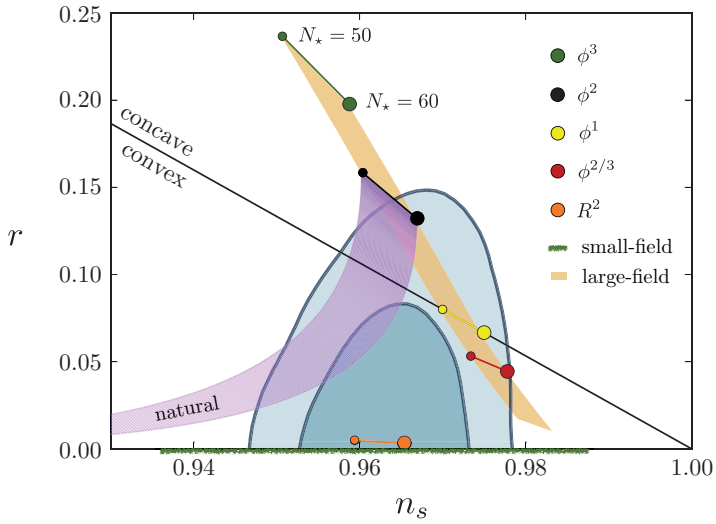


Figure 6.3: *Latest constraints on the scalar spectral index n_s and the tensor amplitude r.*

6.6 Observations

Inflation predicts nearly scale-invariant spectra of superhorizon scalar and tensor fluctuations. Once these modes enter the horizon, they start to evolve according to the processes described in Chapter 5. Since we understand the physics of the subhorizon evolution very well, we can use late-time observations to learn about the initial conditions.

6.6.1 Matter Power Spectrum

In Chapter 5, we showed that subhorizon perturbations evolve differently in the radiation-dominated and matter-dominated epochs. We have seen how this leads to a characteristic shape of the matter power spectrum, cf. fig. 5.4. In fig. 6.4 we compare this prediction to the measured matter power.[5]

[5] With the exception of gravitational lensing, we unfortunately never observe the dark matter directly. Instead galaxy surveys like the Sloan Digital Sky Survey (SDSS) only probe luminous matter. On large scales, the density contrast for galaxies, Δ_g, is simply proportional to density contrast for dark matter: $\Delta_g = b\Delta_m$, where the *bias* parameter b is a constant. On small scales, the relationship isn't as simple.

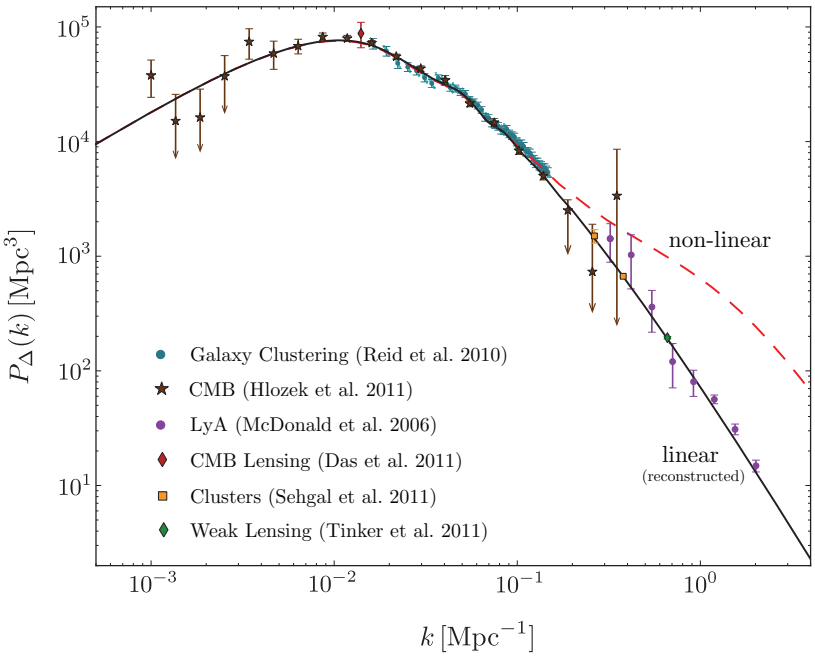


Figure 6.4: *Compilation of the latest measurements of the matter power spectrum.*

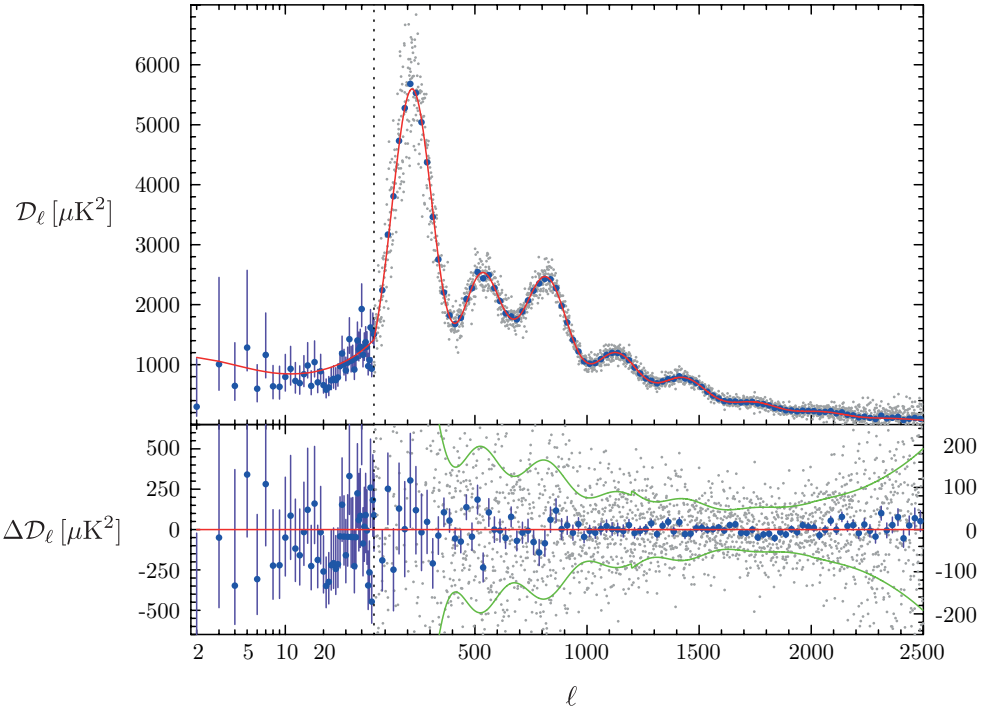


Figure 6.5: *The latest measurements of the CMB angular power spectrum by the Planck satellite.*

6.6.2 CMB Anisotropies

The temperature fluctuations in the cosmic microwave background are sourced predominantly by scalar (density) fluctuations. Acoustic oscillations in the primordial plasma before recombination lead to a characteristic peak structure of the angular power spectrum of the CMB; see fig. 6.5. The precise shape of the spectrum depends both on the initial conditions (through the parameters A_s and n_s) and the cosmological parameters (through parameters like Ω_m, Ω_Λ, Ω_k,

etc.). Measurements of the angular power spectrum therefore reveal information both about the geometry and composition of the universe and its initial conditions.

A major goal of current efforts in observational cosmology is to detect the tensor component of the primordial fluctuations. Its amplitude depends on the energy scale of inflation and it is therefore not predicted (i.e. it varies between models). While this makes the search for primordial tensor modes difficult, it is also what makes it so exciting. Detecting tensors would reveal the energy scale at which inflation occurred, providing an important clue about the physics driving the inflationary expansion.

Most searches for tensors focus on the imprint that tensor modes leave in the *polarisation* of the CMB. Polarisation is generated through the scattering of the anisotropic radiation field off the free electrons just before recombination. The presence of a gravitational wave background creates an anisotropic stretching of the spacetime which induces a special type of polarisation pattern: the so-called *B-mode* pattern (a pattern whose "curl" doesn't vanish). Such a pattern cannot be created by scalar (density) fluctuations and is therefore a unique signature of primordial tensors (gravitational waves). A large number of ground-based, balloon and satellite experiments are currently searching for the B-mode signal predicted by inflation.